Phase Transitions in Language Evolution

Partha Niyogi
The University of Chicago
Hyde Park, Chicago, IL 60637, USA

March 18, 2003

Abstract

Language is transmitted from one generation to the next via learning by individuals. By taking this point of view one is able to link the linguistic behavior of successive generations and therefore study how language evolves over generational time scales. We provide a brief overview of this approach to the study of language evolution, its formalization as a dynamical system, and the analogical connections to the methodological principles of evolutionary biology. We show how the interplay between learning and evolution can be quite subtle and how phase transitions arise in many such models of language evolution. Such phase transitions may provide a suitable theoretical construct with which explanations for rapid language change or evolution may be given. Some illustrative examples are provided.

1 Introduction

Children learn the language of their parents. This cognitive feat is surely grounded in the biology of the human species and much of the linguistic theorizing of the last fifty years has been set against the backdrop of this problem of language acquisition. Now language acquistion is the mechanism by which language is transmitted from the speakers of one generation to those of the next. By taking this point of view, one is able to link the linguistic behavior of successive generations and therefore study how language evolves over generational time. Thus evolutionary and historical phenomena can become the object of study and the last ten years has seen the growth of computational modes of inquiry into these topics.

The primary purpose of this paper is to develop the argument that the framework of dynamical systems is the correct paradigm within which the study of evolving linguistic systems ought to be conducted. Further, as we shall see, the precise nature of the dynamics will depend upon details of linguistic theory, mechanisms of language acquisition, frequency of usage of linguistic expressions, and so on. In most cases of interest, the dynamics will turn out to be nonlinear. The bifurcations (phase transitions) in these systems will turn out to have clear linguistic significance and can be interpreted as the change of the language of a population from one seemingly stable mode to another.

The parallels between biological evolution and linguistic evolution are also worth commenting upon. Evolutionary biologists from the time of Darwin have been motivated in part by trying to explain how biological diversity arises, how it is maintained, and how it evolves over successive generations. There is genotypic variation in any population. The genetic code is transmitted to the next generation through reproduction and Mendelian inheritance. To a first order, the field of population biology then studies the evolution of genetic diversity in populations under the influence of natural selection. In some sense, in this paper, we discuss the emerging field of population linguistics. There is linguistic variation in every population. However, children do not literally inherit the language of their parents. The linguistic code (read grammar) is in fact transmitted from one generation to the next via learning. One may then study how the linguistic variation evolves from one generation to the next under this mode of transmission. In this sense, there are many similarities between the two enterprises. There are, of course, important differences[1]. For example, it is interesting to note that while in the biological case, the child inherits its genes *only* from the parents, the influences on the linguistic development of the child are more diverse. Further, because the mechanism of transmission for language is its acquisition (learning) by children, the theory of learning will play an important role in

[1]These differences make precise analogies difficult. For example, it is unclear whether language evolution is more like population genetics or more like ecology. Arguably it has aspects of both kinds of evolutionary systems and rather than dwelling on analogy too much, we will develop the internal logic of language evolution on its own terms.

the evolutionary models. Finally, there is no clear sense in which natural selection is meaningful or plays a role in language evolution over historical time scales. On the other hand, it must surely play a role in the evolution of new communication systems (across different species) over evolutionary time scales.

A little bit of historical perspective on these parallels and differences is helpful. Since the discovery of the relatedness of the members of the Indo European family of languages by William Jones in the late eighteenth century, historical linguistics dominated the research agenda of much of the nineteenth century. Linguistic family trees were constructed by various methods in an attempt to uncover relatedness and descent of languages. Darwin, living in this century, was by his own admission, greatly influenced by these ideas and several times in *The Descent of Man*, he likens biological diversity to linguistic diversity. Species were like languages. Reproductive compatibility was like communicative compatibility. Like languages, species too could evolve over time and be descended from each other. Both Jones and Darwin were radicals in their own ways. To suggest that Sanskrit (the language of the colonized Indians) was in the same family as Latin (a language with which the imperial masters identified strongly) was against the ingrained notions of those colonial times. To suggest that humans and apes belonged to the same broader family of primates went strongly against the deeply held beliefs of those religious times.

In the twentieth century, both the politics and the science changed. Particularly following the cognitive revolution in linguistics identified most strongly with Chomsky, there was a shift in focus from diachronic to synchronic phenomena as the object of study. Linguistic structure and its acquisition were better understood. In biology, following the genetic revolution brought about by Watson and Crick, the genetic basis of biological variation began to be probed and evolutionary theory quickly incorporated these mechanisms into their models and explanations. Similarly over the last twenty years, the insights from generative grammar and mechanisms of language acquistion are now being used to reexamine the issues and questions of historical linguistics and language evolution.

Finally, it is worthwhile to reflect a little bit on the role of mathematical models in this enterprise. The research program of Haldane, Fisher, and Wright mathematized the discipline of evolutionary biology in part to clarify the seeming tautologies and resulting confusion arising out of various interpretations of Darwinian ideas. In this sense, evolutionary biology has a fairly long mathematical history. In linguistics, particularly over the last fifty years, the notion of a grammar as a computional system underlying language has gained ground. Since grammars are formal objects, the study of their processing, acquisition, and use become amenable to mathematical analysis. In the study of language evolution, as we shall see, one will need to understand the subtle interplay between language learning by individuals and language change in populations. It is difficult to reason effectively and precisely about this interplay through verbal arguments alone. Computational and mathematical models then become essential to make progress and in this paper, we provide a sense of how such models are constructed and the role they might play in future understanding. For a more detailed exposition of these ideas, see Niyogi (2002).

2 The Conceptual Framework of Language Evolution

Let us outline the essential logic of language evolution. The development of the conceptual framework rests on three pillars.

First, that the linguistic systems underlying human communicative behavior are usefully characterized as formal, computational systems. Thus phonemes, features, syllables, words, phrases and so on are formal expressions and follow systematic regularities in the linguistic systems of individual language users. The details of the formal representations of such linguistic systems will depend upon the phenomenological level at hand. Thus, phonological/morphological levels might require different rules and representations from syntactic/semantic levels although interactions surely occur between levels. This assumption that linguistic knowledge may be characterized as a formal system is shared by most modern approaches to linguistic theory though there may be considerable differences in the details of the formal framework used to articulate different theories. See, for example, GB (Chomsky, 1981), HPSG (Pollard and Sag, 1994), LFG (Bresnan, 2001), Optimality Theory (Prince and Smolensky, 1993) and numerous others. We do not need to make any commitment to any particular linguistic theory here.

Second, that at any point in time, there is variation in the linguistic systems of different language users. Let $\mathcal{G}$ be the space of possible linguistic systems that humans might have in normal circumstances. The linguistic system[2] of each individual is an element of $\mathcal{G}$. In a homogeneous linguistic community, all member have similar linguistic systems and so the variation is small. In heterogeneous linguistic communities, the variation might be quite high. In general, one might characterize the linguistic variation in the population by a probability distribution P on the space $\mathcal{G}$. For any grammar $g \in \mathcal{G}$, the quantity $P(g)$ may be interpreted as the proportion of the population using grammatical system g.

Third, language is transmitted from one generation to the next via language acquistion. Thus children do not inherit the language of their parents but rather acquire a language from the primary linguistic data they receive. In general, the source of the primary linguistic data is not just the parents alone but also other members of the child's linguistic environment. Thus the linguistic variation in the adult[3] community would determine the distribution of primary linguistic data (hereafter PLD) the child is exposed to. Language acquisition is a map from primary linguistic data to linguistic systems. We may denote this by $\mathcal{A}$

$$\mathcal{A} : D \to \mathcal{G}$$

where D is the space of possible linguistic experiences (PLD) and $\mathcal{G}$ is the space of possible natural language grammars. Thus for the particular linguistic experience $d \in D$, the child develops the grammar $(\mathcal{A}(d))$. This is the development or growth of language over an individual's learning period[4]
.

[2]We have deliberately let $\mathcal{G}$ be an abstract space that may be treated according to one's linguistic persuasion and application. It could be probabilistic grammars or "convex" combinations of grammars to denote internalized systems that are not consistent with a single grammar in any traditional linguistic sense.

[3]It is worthwhile to remark on an idealization we have made here. We have blocked the population into discrete generations with adults having a mature language and learners trying to acquire a language from exposure to adults. In reality, of course, children learn from each other, older children, as well as adults. What we refer to as variation in the adult community is really better viewed as variation in the generationally heterogeneous linguistic community in which the child is immersed.

[4]Note that D might include many *degenerate* experiences for which $\mathcal{A}(d)$ outputs degenerate grammatical systems. One may reserve a special state $g_{deg} \in \mathcal{G}$ to denote such degenerate systems. If the source of the data is a speaker of a natural language and the child's interaction with such a speaker is normal, then we take it that the linguistic ex-

These three observations taken together allow us to relate how linguistic diversity evolves from one generation to the next. Suppose in generation t we have a linguistic population whose composition is characterized by P_t. Now consider the generation of children growing up in this community. The distribution of data they receive during their learning period will surely be affected by P_t. So let P_D be a probability distribution on D that is induced by P_t. To show this explicit dependence we write P_D as $P_D(d; P_t)$ where for any $d \in D$, the quantity $P_D(d, P_t)$ denotes the likelihood of a typical child of having the particular linguistic experience d during its learning phase. Since language acquisition maps linguistic experience to grammatical systems via the mapping $\mathcal{A}$ we see that $\mathcal{A}(d)$ is a random variable whose distribution characterizes the variable linguistic systems that children might develop. Thus the dependency is as follows:

$$P_t \rightarrow P_D \rightarrow P_{t+1} \tag{1}$$

The influence of P_t on P_D is mediated by social structure, i.e., the pattern of social connectivity among the members of the population. The influence of P_D on P_{t+1} is mediated by language acquisition. The details of $\mathcal{G}$, of exactly how P_t affects P_D, the learning algorithm $\mathcal{A}$, and other variables in this setting will now depend upon the particular linguistic application. Let us instantiate this general logic in a few cases. In each case, we will get a different dynamical system. Often, in such systems, we will see the existence of bifurcations (phase transitions).

---

perience of the child is not degenerate and the child will accordingly develop a natural grammatical system. There is some research and considerable speculation on the role of the extralinguistic environment in language acquisition. We will take the point of view that such extra-linguistic factors allow the child to develop a suitable representation of the linguistic input. In other words, the input to the child is an acoustic stream a in a particular social and physical setting. Extralinguistic considerations may be invoked in converting a into a suitable representation so that the acoustic stimulus contributes towards a linguistically relevant stimulus belonging to D. Now the problem of language acquisition may be considered entirely in the linguistic domain. This is what we consider in this paper.

3 Example 1

We will consider two grammatical variants g_1 and g_2 in competition with each other. Therefore, in this setting, $\mathcal{G} = \{g_1, g_2\}$. Assume that speakers of g_1 produce expressions defining a language L_1 of surface forms. Similarly speakers of g_2 produce expressions defining a language L_2 of surface forms. Speakers of g_1 produce expressions with probability P_1 while speakers of g_2 produce expressions with probability P_2. Naturally, P_1 has support on L_1 while P_2 has support on L_2. The set $L_1 \cap L_2$ consists of *ambiguous* forms that are analyzable under both kinds of grammars. For our purposes, we will also assume that there is a set of *cues* (denote by $C \subset L_2 \setminus L_1$) that will indicate to a potential learner that the speaker is a user of L_2.

Child learners are exposed to primary linguistic data on the basis of which they acquire *either* g_1 or g_2 upon maturation. Let us consider the following *cue*-based learning algorithm that they might follow. g_1 is considered to be the *marked*, or *default* grammatical state of the learner. This is the grammar that is acquired unless enough cues to the contrary occur in the learner's experience. Cues are expressions that belong to $L_2 \setminus L_1$ that provide a clue (cue) to the learner/hearer about the nature of the speaker's underlying grammatical system. In particular, if a cue is heard, it indicates to the learner that the speaker's underlying grammar was g_2.

Let us assume each learner hears a total of k expressions over its learning phase. If there are atleast m cues for g_2 that occur in their learning experience (of k expressions) then they adopt g_2 upon maturation, else they adopt g_1.

This general setup is consistent with particular models of language acquisition that have been suggested in the principles and parameters tradition of linguistic theory. For example, g_1 and g_2 may be two grammars that differ by a linguistic parameter. (see Lightfoot, 1999; Kroch, 2001; Gibson & Wexler, 1994; Fodor, 1998; Bertolo, 2001, and so on for expositions).

We are now in a position to analyze the population dynamics. We will adopt a discrete generational structure. At time t, let $x(t) \in [0, 1]$ denote the proportion of g_1 users in the adult generation. Therefore $1 - x(t)$ is the proportion of g_2 users in that same generation. Thus, we have variation in the adult community. Children born in this community hear expressions from both sorts of grammars. The probability of hearing a cue for grammar

g_2 is given by $px(t)$ where

$$p = \sum_{s \in C} P_2(s).$$

In other words, p is the probability with which a speaker of g_2 will produce a cue for the child to whom the speech is directed. The probability of hearing at least m cues is then easily calculated[5]. This is given by

$$\sum_{i=m}^{k} \binom{k}{i} (px(t))^i (1 - px(t))^{(k-i)}$$

Since this is the probability with which the individual child will attain g_2, one would expect that this would also be the proportion of g_2 users in the next generation. Thus, one obtains the following map.

$$x(t+1) = \sum_{i=m}^{k} \binom{k}{i} (px(t))^i (1 - px(t))^{(k-i)} \tag{2}$$

This map describes the evolution of the linguistic variants in the population from generation to generation. In order to derive the precise nature of the dynamics, we have made the following major simplifying assumptions:

1. Population sizes are infinite.

2. Populations are perfectly mixed. Thus the primary linguistic data that children receive reflect accurately (i.e., without bias) the variation that exists in the adult population.

3. Children acquire exactly one grammatical system based on the linguistic data they receive. Further there is a maturation time or learning period after which their grammatical system crystallizes and does not change thereafter. This maturational time is given by k in our example.

4. Children use a cue-based learning algorithm.

A number of aspects of the dynamics given by eq. 2 are worth highlighting. First, it is noteworthy that the population dynamics is *non-linear*. A number of linguists have informally invoked the imagery of "chaos theory"

---

[5]In other words, we assume a threshold $\tau = \frac{m}{k}$ such that the proportion of cues must be at least τ in order to acquire g_2.

in constructing explanatory paradigms of language change (see in particular, Lightfoot, 1999 or Mufwene, 2002). If such metaphors are to be made precise at all, then mathematical models for the dyamics of language need to be constructed. Now that we have actually derived a non-linear dynamical system one might ask whether chaos really arises. As it turns out, in this case, although bifurcations (phase transitions) do arise the models never enter the chaotic regime.

Clearly the dynamics depends upon p and k. For a fixed k, the bifurcation diagram is shown in fig. 1. Note that p, being a proability value, always lies between 0 and 1. For large p, there are two stable equilibria, $x = 0$ and $x = x_p \approx 1$. This suggests that with the passage of time, one of the two grammars will be eliminated and the population will converge to a linguistically homogeneous one. Which grammar dominates will depend upon the initial conditions but the interesting conclusion is that *variation will be eventually eliminated.*

As p drops below a critical value (p_{crit}) , a *bifurcation* (phase transition) occurs. Suddenly, the system moves from a regime where there are two stable equilibria to one in which there is only one stable equlibrium, viz., $x = 0$. The other stable equilibrium ($x = x_p$) vanishes! Such bifurcations are common in parameterized nonlinear systems and have played a role in understanding rapid change in physical or biological systems. It is therefore interesting to see a concrete example in the case of linguistics.

What do we make of this bifurcation diagram? First, it is noteworthy that a homogeneous population of g_1 users is *always* stable. The analysis here suggests that under the assumptions made a change from g_1 to g_2 is unlikely. The only way in which such a change would come about is if there was a massive influx of g_2 speakers from outside so that the composition changed to the basin of attraction of g_2. Consider on the other hand, the change from g_2 to g_1. This might come about because p drifts from $p > p_{\text{crit}}$ to $p < p_{\text{crit}}$. If p drifts in this manner, the dynamics shifts suddenly to a regime where the only stable mode is $x = 0$ – a population entirely made up of g_1 users. The population composition drifts in this direction. Second, it is worth noting that p_{crit} depends upon k. In particular, it is possible to show that p_{crit} *increases* as the value of k *decreases*. Therefore there are actually *two* different ways in which the bifurcation might occur. The value of p might drift over time and cross the critical threshold for a phase transition. This is equivalent to saying that the frequency with which speakers of a language

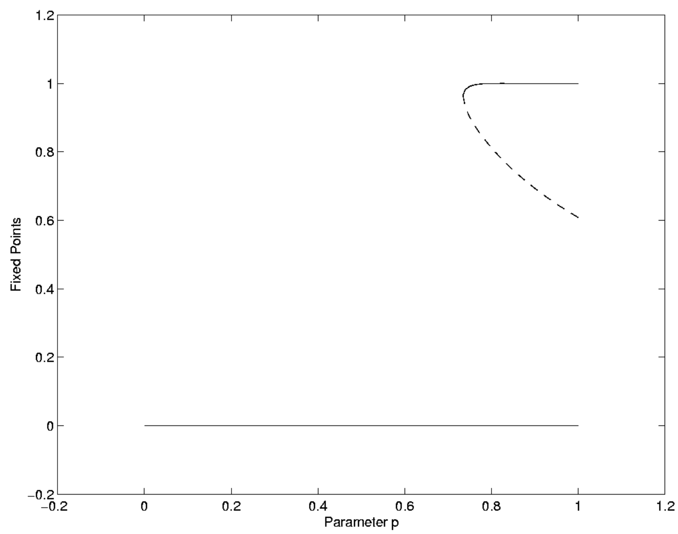

Figure 1: The bifurcation diagram for as the parameter p varies. For this example, $K = 50$ and $\tau = 0.6$ Notice how for small values of p there is only one stable point at $\alpha = 0$. As p increases a new pair of fixed points arises — one of which is unstable (dotted) and the other is stable (solid). For any value of p on the x-axis, the y-axis denotes the values of the fixed points.

provide the cues changes. Alternatively, p might be fixed but the value of k might change. In other words, the total number of examples changes (because the number of linguistic interactions decrease, perhaps). In a later example with word change, we will see how these two modes provide different interpretations of the cause of language change.

3.1 An Example of Lexical Change

Leaving the syntactic domain, one could develop a very similar model for lexical learning and change. Imagine now that a word has two forms w_1 and w_2. Classic examples are alternative pronunciations of the same word such as the American and British pronunications of *tomato*, or alternative pronunciations of *either* (with monophthongized and diphthongized first vowel respectively). The two pronunciations might differ by a phonetic feature and there might be acoustic cues that trigger the acquisition of such a feature.

In many cases of interest, there may be an inherent *asymmetry* in the acquisition of such a feature. For example (see Plauche et al, 1997 for a full discussion), the perceptual confusion between /**k**/ (unvoiced, velar plosive) and /**t**/ (unvoiced alveolar plosive) is asymmetric with /**k**/ being misperceived more often as /**t**/ than the other way around. In Plauche et al, a phonetically grounded discussion is conducted in terms of acoustic cues where it is argued that the acoustic properties of /**k**/ include all the acoustic properties for /**t**/ plus some more distinguishing cues. If these cues are missing the speaker simply hears the intended sound as /**t**/. The other way around never occurs. In particular, following Stevens and Blumstein (1978) it is hypothesized that a mid-frequency burst energy in the region of 3-4 kHz is stronger for velar sounds than it is for alveolar sounds. Consequently, they reason that this energy may be missed more often rather than introduced spuriously by listeners. Therefore velar sounds are more likely to be misperceived as alveolar than the other way around.

Imagine now that the two forms w_1 and w_2 differ exactly by this feature. Now one can look at the distribution of this feature in the linguistic population and how this distribution might change with time. The model is as follows.

Let $x(t)$ be the proportion of individuals (let us refer to these individuals as type 1 individuals) in the population (at time t) who have internalized form w_1 in their lexical inventory. Consequently a proportion $1 - x(t)$ have internalized form w_2 (these are the type 2 individuals). Everytime a type 2 individual uses the word, with probability p they produce enough cues that they are perceived correctly by the listener. Everytime a type 1 individual uses the word, they are always perceived correctly.

Now consider the typical child hearer/learner. This child hears the word used k times in all during its learning period. Since this child is immersed in a heterogeneous linguistic environment, sometimes the word has been uttered by a type 1 individual and sometimes by a type 2 individual. Every type 1 individual is perceived correctly with probability 1. Every type 2 individual is perceived correctly with probability p. With probability $1-p$ form w_2 uttered by a type 2 individual is misperceived as form 1 (w_1). After k instances, the learner acquires form 1 (as its mature underlying representation in its lexical inventory) if it hears form 1 atleast m times in its experience. The evolution of types in the population is given by the exact same equation.

$$x(t+1) = \sum_{i=m}^{k} \binom{k}{i} (px(t))^i (1-px(t))^{(k-i)}$$

Following the previous discussion, we see

1. Form 1 is always stable. A population of type 1 individuals always remains that way. This is unsurprising since type 1 individuals are never (read rarely) misperceived.

2. Form 2 *may or may not* be stable. This brings out the subtlety of the situation. Since there is an inherent asymmetry, one might think that w_2 would always gradually get replaced by w_1. Our analysis here shows that this need not be the case. In fact, if $p > p_{crit}$, we see that a population of largely type 2 individuals can remain stably in the society for all time.

3. We see that a change from w_1 to w_2 is unlikely except with a massive population restructuring due to language contact and migration. On the other hand, a change from w_2 to w_1 could come about because of a phase transition where the system moves from a regime where $p > p_{crit}$ to one where $p < p_{crit}$.

4. In this phonetic example, one might think that the probability p is grounded in our biological perceptual apparatus and this presumably

does not change with time. Consequently, p is probably fixed. Therefore it is more likely that change occurs due to a decrease in the value of k. Note that k is simply how many times the word (in either form) is uttered during the learning period. If the word becomes infrequent (due to stylistic, sociolinguistic or other considerations) then k will become small, the relation between p and p_{crit} might invert and the process of language change might be set in motion.

4 Example 2

Let us briefly consider the application of this point of view to the analysis of syntactic change in French during the period from the fourteenth century to the seventeenth century (A.D.). To keep our discussion concrete we will focus on some particular parametric changes in French syntax over this period. Our analysis draws heavily from the work of Ian Roberts (linguistic work in Roberts, 1993 and computational work in Clark and Roberts, 1993) and a more recent treatment in Yang, 2002.

4.1 Linguistic Background

The discussion that follows is conducted within the principles and parameters tradition (Chomsky, 1981) of linguistic theory. There were two dominant parametric changes that occurred in French syntax over the period under consideration. First, there was loss of subject (pro) drop. In Old French, (like modern Italian), a pronominal subject could be dropped as the following examples show.

Loss of null subjects

1. * Ainsi s'amusaient bien cette nuit. (ModF)
   thus (they) had fun that night

2. Si firent grant joie la nuit. (OF)
   thus (they) made great joy the night

Second, there was loss of verb second phenomena (V2). Old French was a V2 language so that V could raise to C (with the specifier typically filled) and occupy therefore the second position in the linear order of the constituents. This is no longer true as the following examples show.

Loss of V2

1. * Puis entendirect-ils un coup de tonnerre. (ModF)
   then heard-they a clap of thunder

2. Lors oirent ils venir un escoiz de tonoire (OF)
   then heard they come a clap of thunder

Thus the situation is simply summarized as follows. In the beginning there was a relatively stable and homogeneous grammatical system that was +V2 and had null subjects (pro drop). At the end, there was again a relatively stable and homogeneous grammatical system that had lost both V2 and pro drop. In the middle there was variation with multiple grammatical variants co-existing in the population. Thus it is natural for us to analyse the situation within the framework introduced in section 2.

4.2 Computational Analysis

We make the following assumptions.

1. Each speaker is potentially bilingual/multilingual with multiple grammatical systems that provide the basis for linguistic use.

2. Similarly, each child potentially acquires multiple grammatical systems based on its linguistic experience. In particular, in periods when there is linguistic variation in the adult population and the data received is not consistent with a single grammar, the child will accordingly acquire multiple systems.

4.2.1 The Grammatical Setting

For illustrative purposes, we will focus on the competition between two grammatical systems. The two grammars are denoted by g_+ and g_- respectively. The corresponding sets of surface expressions (sentences) are denoted by L_+ and L_- respectively. When using the grammatical system g_+ speakers produce sentences with a probability distribution P_+ (over L_+) and similarly when using g_- speakers produce sentences with a probability distribution P_- (over L_-).

For example, if g_+ were a head-first grammar without verb second movement and no prodrop (like modern French), then L_+ consists of elements like (a) **SVO** (subject-verb-object; like the modern English *Mary sees the children* or the modern French *Marie voit les enfants* (b) **XSVO** (like the English *After dinner, John read the newspaper.*) and so on. In general, in our analysis, various choices may be made for g_+ and g_- and the evolutionary consequences may then be examined.

In contrast to the previous section, recall that speaker/learners here are potentially bilingual. Thus, each speaker has a *grammatical mix factor* $\lambda \in [0, 1]$ that characterizes how often the speaker uses g_+ as opposed to g_-. For example, a speaker with mix factor $\lambda = 0$ (resp. $\lambda = 1$) uses exclusively g_- (resp. g_+) and corresponds to a monolingual speaker with a single underlying grammatical system. Similarly, a speaker with mix factor $\lambda = \frac{1}{2}$ uses g_+ half the time and g_- half the time when producing sentences. In general, a speaker with mix factor λ therefore produces sentences with a probability distribution given by $\lambda P_+ + (1-\lambda)P_-$. Note that this distribution is over the set $L_+ \cup L_-$. Thus there may be *internal* variation within each speaker and the expressions produced by such a speaker are not consistent with a single unique grammar. Studies by Kroch (see Kroch, 2001 for overview) suggest that this state of affairs is often the case especially in the context of language contact and change. Thus in our framework $\mathcal{G} = \{h | h = \lambda g_1 + (1 - \lambda) g_2\}$ where $\mathcal{G}$ is a space of formal convex combinations denoting multiple grammatical systems.

There is also *external* variation in the adult population. Thus different individuals have potentially different λ values and one can therefore imagine the distribution of λ values in the adult population. A summary statistic for the average linguistic behavior of the population as a whole may be provided by the mean value of λ which we denote by $E[\lambda]$. We will be interested in the evolution of this quantity over generational time.

4.2.2 Learning and Evolution:

Children hear expressions produced by speakers in their linguistic community. On the basis of these expressions, they acquire a grammatical system. The acquisition of a grammatical system in our current context ultimately reduces to estimating a λ value (for future use) on the basis of the linguistic experience (data). Let us assume that children use only triggers in estimating the correct value of λ. In other words, they ignore all ambiguous expressions (those that belong to $L_+ \cap L_-$ and therefore may be interpreted as generated by either grammar).

Then a reasonable estimate is provided by

$$\hat{\lambda} = \frac{k_1}{k_1 + k_2}$$

Let us assume that the child learner uses such an estimate in language acquisition. The child receives a draw of k example sentences during its linguistic experience. Of these k_1 are triggers for g_+, i.e., the sentences belong to $L_+ \setminus L_-$ and k_2 are triggers for g_-, i.e., the sentences belong to $L_- \setminus L_+$. The rest ($k_3 = k - k_1 - k_2$ sentences) are ambiguous and are ignored by the learner in developing its grammatical system[6].

Note that the precise values of k_1, k_2 and k_3 will vary across individual children depending upon their particular linguistic experiences, i.e, their primary linguistic data. Thus there will be variation in the generation of children as they mature to adulthood. It now makes sense to ask what the average value of $\hat{\lambda}$ will be in the next generation. It is possible to show that this is given by

$$E[\hat{\lambda}] = \frac{ax}{ax + b(1-x)}$$

where

1. $x = E[\lambda]$ is the average value of λ in the parental generation.
2. $a = \sum_{s \in L_+ \setminus L_-} P_+(s)$ is the probability with which a speaker when using g_+ produces a trigger for that grammatical system.
3. $b = \sum_{s \in L_- \setminus L_+} P_-(s)$ is the probability with which a speaker when using g_- produces a trigger for that grammatical system.

[6] Alternative models that enforce asymmetry may also be developed. For example, it is possible that example sentences belonging to $L_+ \cap L_-$ may not be ignored but rather interpreted with a bias according to a preferred or marked grammatical system. In contexts where there may be a first/second language bias on the part of the learner, such asymmetries may arise naturally.

Thus we see that given the average value of λ in the parental generation, we are able to deduce the average value that it will take in the generation of children. We get the following dynamics for the evolution of the average value of λ over generational time. λ_{t+1} denotes the λ variable for the $(t+1)$th generation and λ_t denotes the same for the (t)th generation.

$$E[\lambda_{t+1}] = \frac{aE[\lambda_t]}{aE[\lambda_t] + b(1 - E[\lambda_t])}$$

4.3 Bifurcations and Syntactic Change

The equation of change is therefore given by

$$x_{t+1} = \frac{ax_t}{ax_t + b(1 - x_t)}$$

where $x_t = E[\lambda_t]$. If one analyzes the above equation, we see that the dynamics of the population depends upon the parameters a and b. In particular, it is possible to show that

1. If $a > b$ then $x = 1$ is the only stable point. From all initial conditions, the population will converge over evolutionary time to a homogeneous population of g_+ users.

2. If $a < b$ then $x = 0$ is the only stable point. From all initial conditions, the population converges to g_-.

3. If $a = b$ then $x_{t+1} = x_t$ for all t. There is no change.

Most interestingly, from our point of view, once again we have a bifurcation in the dynamical system in terms of which one may interpret the facts of language change. Thus, on this account, one would suggest that a homogeneous stable population of g_+ users ($x = 1$) could become unstable if the frequencies of sentence changed so that a became less than b while before it was the other way around. Under this condition, we see that the introduction of even the slightest variation in the population would cause the language of the community to move to one of g_- users, i.e., large scale language change as a result of a bifurcation. It is also interesting to note that syntactic diglossia is permitted within the grammatical and acquisition framework, it is usually eliminated over time unless a is exactly equal to b in such models.

Looking more closely at the grammatical theories and the data, we find that *if there was no pro-drop*, a +V2 grammar tends to be quite stable in conmparison to a -V2 grammar if this is the only parametric difference between the two grammars. Following the analysis in Roberts (1993) and Yang (2000), we may take the two grammars to be:

1. g_+: the +V2 grammar has expressions like **SVO** (subject-verb-object; with verb typically in C and subject in spec-C) and **VS** patterns like **XVSO**, **OVS** and so on.

2. g_-: the -V2 grammar (like Modern French) has expressions like (a) **SVO** (subject-verb-object; with subject in spec-IP) (b) **XSVO** (in general, V> 2 patterns).

Following our analysis above, we see that **SVO** patterns do not count as triggers. The proportion of **XSVO** (trigger for -V2) and **XVSO** (trigger for +V2) patterns in the speech of g_- and g_+ users respectively will determine the evolution of the population. Preliminary statistics (following Yang, 2000) based on the speech of modern $-$V2 (like English and French) and +V2 (like German and Dutch) languages suggest that $a = 0.3$ while $b = 0.2$. Consequently, the +V2 grammar would remain stable.

Let us now consider a +V2 (with pro drop) grammar in competition with a -V2 (with prodrop) grammar. Then we have the following patterns[7]: (Note that +V2 grammars with pro drop will not generate **VO** expressions presumably because subj is in spec-CP and this needs to be filled.)

[7]These patterns are provided for illustrative purposes. In reality, of course, there are many more expressions that are generated by each of the two grammars but these may be deemed irrelevant to the discussion at hand. Consequently, the probabilities provided may be treated as normalized after discarding these irrelevant distributions. More precisely, we are assuming the following. L_+ and L_- each contain a potentially infinite number of expressions. We restrict our discussion to a set $A \subset \Sigma^*$ of expressions where $A = \{\mathbf{SVO}, \mathbf{XVSO}, \mathbf{XVO}, \mathbf{VO}, \mathbf{XSVO}\}$. Then for any element $a \in A$, when we put in values for $P_+(a)$ in our calculations, we actually use the value $\frac{P_+(a)}{P_+(A \cap L_+)}$ and similarly for P_-. There are two other potentially important considerations that we have eliminated from our current discussion. First, we are restricting ourselves to verb-medial grammars. It has been proposed that +V2 systems tend to be more stable in verb-final grammatical systems than in verb-medial ones. We do not explore this issue in part because of our supposition that French was verb-medial throughout. Second, there was a point when subject pronouns started behaving as clitics and it is quite possible that this behavior affected the interpretation of surface expression during the language learning phase and altered the relevant primary linguistic data to weaken V2. We do not consider this issue in this paper.

1. g_+: **SVO; XVSO; XVO**
2. g_-: **SVO; VO; XVO; XSVO**

Let us do a simple calculation. Assume that with probability p the subject is a pronoun rather than a full NP. Further, with probability d, the pronoun is dropped in a pro-drop (null subject) language. If $d = 1$ then the prodrop is obligatory. If $d = 0$ then the language does not allow prodrop. Then we see that

$$\begin{aligned} P_+(SVO) &= 0.7 \\ P_+(XVSO) &= 0.3((1-p) + p(1-d)) \\ P_+(XVO) &= 0.3pd \end{aligned}$$

To clarify the logic of the calculations, let us consider the probability with which **XVSO** would be produced by a speaker of g_+. With probability 0.3 a baseform of **XVSO** would be produced. Now we need to calculate the probability with which this is overtly expressed, i.e., the subject is not dropped. There are two cases: (i) the subject position is filled by a full NP (with probability $1-p$) in which case it cannot be dropped (ii) the subject position is filled with a pronoun (probability p) but this pronoun is not dropped (probability $1-d$). Multiplying this out and adding the two cases, we obtain $P_+(XVSO) = 0.3((1-p) + p(1-d))$. Probability values for the other expressions are obtained similarly.

Now consider probabilities with which g_- speakers produce their expressions. It is simply seen that

$$\begin{aligned} P_-(SVO) &= 0.8(1-pd) \\ P_-(VO) &= 0.8pd \\ P_-(XSVO) &= 0.2(1-pd) \\ P_-(XVO) &= 0.2pd \end{aligned}$$

Given P_- and P_+, we can calculate a and b to be

$$a = 0.3(1-pd)$$

and

$$b = 0.8pd + 0.2(1-pd)$$

From this we see that for $a > b$ we need

$$0.3(1 - pd) > 0.8pd + 0.2(1 - pd)$$

or

$$pd < \frac{1}{9}$$

Thus we see that if $d = 0$, i.e., the language has no prodrop then the dynamics is in the regime $a > b$ and correspondingly, the +V2 grammar is stable as our first analysis showed already. On the other hand, if $d > \frac{1}{9p}$ then a bifurcation occurs and the +V2 grammar becomes unstable. One might then ask, how come a +V2 and pro drop grammar (as old French putatively was) remained stable in the first place? According to this analysis it must be because p was small (so that the product $pd < \frac{1}{9}$). Now notice that if this were the state of affairs, then the only way in which change would come about is if p increased to cross the threshold. While this is happening it is crucial that the null subject is not being lost, i.e., d is not decreasing. By this analysis +V2 is lost before the null subject is lost. If the null subject were lost first, then $d = 0$ and the dynamics would always be in the regime $pd < \frac{1}{9}$ and the +V2 parameter would always remain stable. On this account, +V2 is lost before the null subject was lost. Further, +V2 was lost because of the increase in p, i.e., the use of pronominal subjects in the speech of the times.

The above analysis is empirically anecdotal as we have plugged in plausible numbers for the probabilities. The point of the exercise was to show again how a bifurcation may be at the root of language change and how the conditions for change depend in a subtle way on the frequencies with which expressions are produced. In this case, the product pd is seen to determine the stability of the language. Further, we obtain a linguistic prediction, that +V2 must have been lost before the null subject parameter was lost. The loss of +V2 must have been triggered by the increase in the use of pronominal subjects.

5 Outlook

Variation, Heredity, and Fitness are the fundamental principles of biological evolution by natural selection. We have tried to clarify the principles of language evolution over historical time scales. While there are conceptual similarities, there are also important differences. Crucially, offspring *inherit* their genetic composition from their parents in biological evolution. On the other hand, they *learn* their language from the parental generation at large. Thus language acquisition is the driving force that shapes the evolution of linguistic diversity just as heredity and differential fitness are the forces in the biological case.

To illustrate the conceptual framework, we have examined the situation where two linguistic "types" are in competition with each other. Models of this sort have the same status in language evolution that analogous models of one gene-two alleles have in population genetics. Considerable insight and progress may be made from this point of view by restricting the discussion of language change to a few linguistic parameters at a time.

It needs to be properly understood that a number of simplifying assumptions have been made in the construction of the basic models presented in this paper. By dropping these assumptions a number of variations of this basic model may be obtained. One can then study in a systematic way the effect of multiple (more than 2) grammatical systems in competition, the effect of finite population sizes, of social stratification and neighborhood effects, of complicated generational structure, and so on. Rather than proposing a single unique model for language evolution, what we are suggesting is a general framework within which different explanatory models may be constructed and different issues examined.

The most important take home message of this short paper is the central role of bifurcations in the analysis of language change. Attempts to formalize the process of language change lead to dynamical systems that are typically nonlinear. The state space of such a system is determined by linguistic theory and the update rule is governed by considerations from learning. The parameters of such a dynamical system depend upon the frequencies with which different expressions are used by speakers of the language. Thus we have a and b in the symmetric model developed for syntactic change in French while we have p in the asymmetric models introduced earlier. In each case, we see that a change in these parameters may cause a bifurcation where the dynamics qualitatively changes from one regime to another. Consequently, the stable modes of the linguistic population may be altered and language change may be interpreted in this context. We have investigated several models of language change over the years. Again and again, bifurcations are seen in such models leading one to believe that such bifurcations are real, pervasive, and provide the natural theoretical explanatory construct for the striking patterns of language change that we see in real life.

References

[1] S. Bertolo. *Language Acquisition and Learnability.* Cambridge University Press, 2001.

[2] J. Bresnan. *Lexical-Functional Syntax.* Oxford: Blackwell, Oxford, 2001.

[3] N. Chomsky. *Lectures in Government and Binding.* MIT Press, 1981.

[4] R. Clark and I. Roberts. A computational model of language learnability and language change. *Linguistic Inquiry*, 24:299–345, 1993.

[5] J. D. Fodor. Unambiguous triggers. *Linguistic Inquiry*, 29.1:1–36, 1998.

[6] E. Gibson and K. Wexler. Triggers. *Linguistic Inquiry*, 25(3):407–454, 1994.

[7] A. Kroch. *Syntactic Change (Handbook of Syntax: Baltin and Collins (ed)).* Blackwell, 2001.

[8] D. Lightfoot. *The Development of Language: Acquisition, Change, and Evolution.* Blackwell/Maryland, 1999.

[9] J. Ohala M. Plauche, C. Delogu. Asymmetries in consonant confusion. In *Proceedings of Eurospeech*, Rhodes, Greece, 1997.

[10] S. Mufwene. *The Ecology of Language Evolution.* Cambridge University Press, 2002.

[11] P. Niyogi. *The Computational Nature of Language Learning and Evolution.* http://www.cs.uchicago.edu/ niyogi/, 2002.

[12] C. J. Pollard and I. A. Sag. *Head-Driven Phrase Structure Grammar.* University of Chicago Press, Chicago, 1994.

[13] A. Prince and P. Smolensky. *Optimality Theory: Constraint Interaction in Generative Grammar.* Manuscript, 1993.

[14] I. Roberts. *Verbs and Diachronic Syntax: A Comparative History of English and French.* Kluwer Academic Publishers:Dordrecht, 1993.

[15] K. Stevens and S. Blumstein. Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 76.2:1358–1368, 1978.

[16] C.D. Yang. Internal and external forces in language change. *Language Variation and Change*, 12:231–250, 2000.