MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL AND COMPUTATIONAL LEARNING
DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES

A.I. Memo No. 1516 July, 1995
C.B.C.L. Paper No. 115

The Logical Problem of Language Change

Partha Niyogi and Robert C. Berwick

Abstract

This paper considers the problem of language change. Linguists must explain not only how languages are learned but also how and why they have evolved along certain trajectories and not others. While the language learning problem has focused on the behavior of individuals and how they acquire a particular grammar from a class of grammars $\mathcal{G}$, here we consider a *population* of such learners and investigate the emergent, global population characteristics of linguistic communities over several generations. We argue that language change follows logically from specific assumptions about grammatical theories and learning paradigms. In particular, we are able to transform parameterized theories and memoryless acquisition algorithms into grammatical dynamical systems, whose evolution depicts a population's evolving linguistic composition. We investigate the linguistic and computational consequences of this model, showing that the formalization allows one to ask questions about diachronic that one otherwise could not ask, such as the effect of varying initial conditions on the resulting diachronic trajectories. From a more programmatic perspective, we give an example of how the dynamical system model for language change can serve as a way to distinguish among alternative grammatical theories, introducing a formal diachronic adequacy criterion for linguistic theories.

This report describes research done at the Center for Biological and Computational Learning and the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Center is provided in part by a grant from the National Science Foundation under contract ASC–9217041. Robert C. Berwick was supported by a Vinton-Hayes Fellowship.

1 Introduction

As is well known, languages change over time. Language scientists have long been occupied with describing phonological, syntactic, and semantic change, often appealing to the analogy between language change and evolution. Some even suggest that language itself is a complex adaptive system (see Hawkins and Gell-Mann, 1989). For example, Lightfoot (1991, chapter 7, pp. 163–65ff.) talks about language change in this way: "Some general properties of language change are shared by other dynamic systems in the natural world...In population biology and linguistic change there is constant flux..... If one views a language as a totality, as historians often do, one sees a *dynamic system.*" Indeed, entire books have been devoted to the description of language change using the terminology of population biology: genetic drift, clines, and so forth[1] However, these analogies have rarely been pursued beyond casual and descriptive accounts.[2] In this paper we formalize these intuitions, to the best of our knowledge for first time, as a concrete, computational, dynamical systems model, and investigating the consequences of this formalization.

In particular, we show that a model of language change emerges as a logical consequence of language acquisition, a point made by Lightfoot (1991). We shall see that Lightfoot's intuition that languages could behave just as though they were dynamical systems is essentially correct, as is his proposal for turning language acquisition models into language change models. We can provide concrete examples of both "gradual" and "sudden" syntactic changes, occurring over time periods of many generations to just a single generation.[3]

Many interesting points emerge from the formalization, some programmatic:

- *Learnability* is a well-known criterion for the adequacy of grammatical theories. Our model provides an *evolutionary* criterion: By comparing the trajectories of dynamical linguistic systems to historically observed trajectories, one can determine the adequacy of linguistic theories or learning algorithms.
- We derive explicit dynamical systems corresponding to parametrized linguistic theories (e.g., the Head First/Final parameter in head-driven phrase structure grammars or government-binding grammars) and memoryless language learning algorithms (e.g., gradient ascent in parameter space).
- We illustrate the use of dynamical systems as a research tool by considering the loss of Verb Second position in Old French as compared to Modern French. We demonstrate by computer modeling that one grammatical parameterization in the literature does not seem to permit this historical change, while another does. We can more accurately model the time course of language change. In particular, in contrast to Kroch (1989) and others, who mimic population biology models by imposing S-shaped logistic curves on possible language changes by assumption, we derive the time course of language change from more basic assumptions, and show that it need not be S-shaped; rather, an S-shape can emerge from more fundamental properties of the underlying dynamical system.
- We examine by simulation and traditional phase-space plots the form and stability of possible "diachronic envelopes" given varying alternative language distributions, language acquisition algorithms, parameterizations, input noise, and sentence distributions. The results bear on models of language "mixing"; so-called "wave" models for language change; and other proposals in the diachronic literature.
- As topics for future research, the dynamical system model provides a novel possible source for explaining several linguistic changes including: (a) the evolution of modern Greek metrical stress assignment from proto-Indo-European; and (b) Bickerton's (1990) "creole hypothesis," concerning the striking fact that all creoles, irrespective of linguistic origin, have exactly the same grammar. In the latter case, the "universality" of creoles could be due a parameterization corresponding to a common condensation point of a dynamical system, a possibility not considered by Bickerton.

2 An Acquisition-Based Model of Language Change

How does the combination of a grammatical theory and learning algorithm lead to a model of language change? We first note that just as with language acquisition, there is a seeming paradox in language change: it is generally assumed that children acquire their caretaker (target) grammars without error. However, if this were always true, at first glance grammatical changes within a population could seemingly never occur, since generation after generation children would successfully acquire the grammar of their parents.

Of course, Lightfoot and others have pointed out the obvious solution to this paradox: the possibility of slight misconvergence to target grammars could, over several generations, drive language change, much as speciation occurs in the population biology sense:

> As somebody adopts a new parameter setting, say a new verb-object order, the output of that person's grammar often differs from that of other people's. This in turn affects the linguistic environment, which may then be more likely to trigger the new parameter setting in younger people. Thus a chain reaction may be created. (Lightfoot, 1991, p. xxx)

We pursue this point in detail below. Similarly, just

[1] For a recent example, see Nichols (1992), *Linguistic Diversity in Space and Time*.

[2] Some notable exceptions are Kroch (1990) and Clark and Roberts (1993).

[3] Lightfoot 1991 refers to these sudden changes acting over a single generation as "catastrophic" but in fact this term usually has a different sense in the dynamical systems literature.

as in the biological case, some of the most commonly observed changes in languages seem to occur as the result of the effects of surrounding populations, whose features infiltrate the original language.

We begin our treatment by arguing that the problem of language acquisition at the individual level leads logically to the problem of language change at the group or population level. Consider a population speaking a particular language[4]. This is the target language—children are exposed to primary linguistic data (PLD) from this source, typically in the form of sentences uttered by caretakers (adults). The logical problem of language acquisition is how children acquire this target language from their primary linguistic data—to come up with an adequate learning theory. We take a learning theory to be simply a mapping from primary linguistic data to the class of grammars, usually effective, and so an algorithm. For example, in a typical inductive inference model, given a stream of sentences, an acquisition algorithm would simply update its grammatical hypothesis with each new sentence according to some preprogrammed procedure. An important criterion for learnability (Gold, 1967) is to require that the algorithm converge to the target as the data goes to infinity (identification in the limit).

Now suppose that we fix an adequate grammatical theory and an adequate acquisition algorithm. There are then essentially two means by which the linguistic composition of the population could change over time. First, if the primary linguistic data presented to the child is altered (due to any number of causes, perhaps to presence of foreign speakers, contact with another population, disfluencies, and the like), the sentences presented to the learner (child) are no longer consistent with a single target grammar. In the face of this input, the learning algorithm might no longer converge to the target grammar. Indeed, it might converge to some other grammar (g_2); or it might converge to g_2 with some probability, g_3 with some other probability, and so forth. In either case, children attempting to solve the acquisition problem using the same learning algorithm could internalize grammars different from the parental (target) grammar. In this way, in one generation the linguistic composition of the population can change.[5]

Second, even if the PLD comes from a single target grammar, the actual data presented to the learner is truncated, or finite. After a finite sample sequence, children may, with non-zero probability, hypothesize a grammar different from that of their parents. This can again lead to a differing linguistic composition in succeeding generations.

In short, the diachronic model is this: Individual children attempt to attain their caretaker target grammar. After a finite number of examples, some are successful, but others may misconverge. The next generation will therefore no longer be linguistically homogeneous. The third generation of children will hear sentences produced by the second—a different distribution—and they, in turn, will attain a different set of grammars. Over successive generations, the linguistic composition evolves as a dynamical system.

On this view, language change is a logical consequence of specific assumptions about:

1. the *grammar hypothesis space*—a particular parametrization, in a parametric theory;
2. the *language acquistion device*—the learning algorithm the child uses to develop hypotheses on the basis of data;
3. the *primary linguistic data*—the sentences presented to the children of any one generation.

If we specify (1) through (3) for a particular generation, we should, in principle, be able to compute the linguistic composition for the next generation. In this manner, we can compute the evolving linguistic composition of the population from generation to generation; we arrive at a dynamical system. We now proceed to make this calculation precise. We first review a standard language acquisition framework, and then show how to derive a dynamical system from it.

2.1 The Language Acquisition Framework

Let us state our assumptions about grammatical theories, learning algorithms, and sentence distributions.

1. Denote by $\mathcal{G}$, a family of possible (target) grammars. Each grammar $g \in \mathcal{G}$ defines a language $L(g) \subseteq \Sigma^*$ over some alphabet Σ in the usual way.

2. Denote by P a distribution on Σ^* according to which sentences are drawn and presented to the learner. Note that if there is a well defined target, g_t, and only positive examples from this target are presented to the learner, then P will have all its measure on $L(g_t)$, and zero measure on sentences outside Suppose n examples are drawn in this fashion, one can then let $\mathcal{D}_n = (\Sigma^*)^n$ be the set of all n-example data sets the learner might be presented with. Thus, if the adult population is linguistically homogeneous (with grammar g_1) then $P = P_1$. If the adult population speaks 50 percent $L(g_1)$ and 50 percent $L(g_2)$ then $P = \frac{1}{2}P_1 + \frac{1}{2}P_2$.

3. Denote by $\mathcal{A}$ the acquisition algorithm that children use to hypothesize a grammar on the basis of input data. $\mathcal{A}$ can be regarded as a mapping from $\mathcal{D}_n$ to $\mathcal{G}$. Thus, acting upon a particular presentation sequence $d_n \in \mathcal{D}_n$, the learner posits a hypothesis $\mathcal{A}(d_n) = h_n \in \mathcal{G}$. Allowing for the possibility of randomization, the learner could, in general, posit $h_i \in \mathcal{G}$ with probability p_i for such a presentation sequence d_n. The standard (stochastic version) learnability criterion (Gold, 1967) can then be stated as follows:

For every target grammar, $g_t \in \mathcal{G}$, with positive-only examples presented according to P as above, the learner must converge to the target with probability 1, i.e.,

$$Prob[\mathcal{A}(d_n) = g_t] \longrightarrow_{n\to\infty} 1$$

[4]In our analysis this implies that all the adult members of this population have internalized the same grammar (corresponding to the language they speak).

[5]Sociological factors affecting language change, affect language acquisition in exactly the same way, yet are abstracted away from the formalization of the logical problem of language acquisition. In this same sense, we similarly abstract away such causes here.