The Computational Nature of Language Learning and Evolution

Partha Niyogi
The University of Chicago

December 28, 2004

Contents

Preface

This monograph explores the interplay between learning and evolution in the context of linguistic systems. For several decades now, the process of language acquisition has been conceptualized as a procedure that maps linguistic experience onto linguistic knowledge. If linguistic knowledge is characterized in computational terms as a formal grammar and the mapping procedure is algorithmic, this conceptualization admits computational and mathematical modes of inquiry into language learning. Indeed, such a view is implicit in most modern approaches to the subject in linguistics, cognitive science, and artificial intelligence.

Now learning (acquisition) is the mechanism by which language is transmitted from old speakers to new. Therefore, the evolution of language over generational time in linguistic populations will depend upon the learning procedure used by the individuals in it. Yet the interplay between learning by the individual and evolution of the population can be quite subtle. We need tools to reason about the phenomena and elucidate the precise nature of the relationships involved. To this end, this monograph presents a framework in which to conduct such an analysis.

Most people can directly observe the learning of language by children and marvel at the phenomenon of language acquisition. In contrast, however, few people have direct experience of the unfolding history of a language. Picking up an Old English text like the Anglo Saxon Chronicles is not always part of our daily existence. If one does this, however, one will find a language that is incomprehensible to the modern English speaker. This leads one immediately to the following question: if in the 9th century A.D., the people in England spoke a language like that in the Anglo Saxon Chronicles, this is the language their children should have learned — and their children after them, and so on. How then, did it come to be that the process of iterative learning by successive generations led to the evolution of English so far from its origins. What does it imply for how English might be a thousand years

from now?

Of course, the problem is not limited to English alone. Language change and evolution is ubiquitous. It happens in most languages, in its syntax, its phonology, and its lexicon. It manifests itself in language birth and death phenomena, in creolization, and in dialect formation. It is happening around us as we speak. More mysteriously, it has happened over evolutionary time scales as the language capacity evolved from prelinguistic versions of it.

There is thus a tension between language learning and language evolution. On the one hand, the learning of language by children is robust and reliable. On the other hand, it cannot be perfect or else languages (barring major migratory effects) would not change. This monograph is an attempt to resolve this tension.

The analytic framework introduced here considers a population of linguistic agents. Linguistic agents are of two types: mature users of a language and learners who acquire a language from the other users. Each learner acquires language based on its own primary linguistic data, i.e., linguistic examples received from other users in the community. By taking taking an ensemble average across learners, one is able to derive the average linguistic composition of the mature speakers of the next generation. Thus the average linguistic composition evolves as a dynamical system. The framework is noteworthy for its shift of emphasis from the individual to the population in the analysis of learning and its evolutionary consequences. Much of language learning theory (often termed learnability theory) focuses on an idealized speaker hearer interaction in a homogeneous linguistic environment. In this tradition, one is concerned with whether the learner will converge to the unique target grammar of the parent as more and more data becomes available. In contrast, we analyze learning algorithms in the case in which the learner is immersed in a heterogeneous linguistic environment. There is no unique target grammar and the learner never converges. Instead, there is a distribution of grammars in the linguistically mature population, and the learner matures after a finite time corresponding to its developmental learning period.

In this setting, our main results may be summarized as follows. First, we elucidate the subtle nature of the relationship between learning and evolution. In particular, we show that different learning algorithms may have different evolutionary consequences. Therefore, one is able to bring to bear both developmental and evolutionary data and arguments to judge the plausibility of various learning algorithms for language acquisition. Second, we find that the dynamics of language evolution are typically non-linear. Fur-

ther, there are often bifurcations that lead to a change in the stability profile of the equilibrium distribution of languages. The parameters associated with such bifurcations are naturally interpretable as the frequency of usage of various linguistic expressions. Thus, much like phase transitions in physics, we argue that the continuous drift of such frequency effects could lead to discontinuous changes in the stability of languages over time. We claim that these bifurcations are the natural explanatory construct for the dramatic patterns of change observed in historical linguistics. Third, we investigate the role of natural selection, communicative efficiency, and learning in the origin and evolution of language. In particular, we investigate the conditions under which shared languages (communicative systems) might emerge. We show that if individuals learn from a single agent in the population, then natural selection is necessary for the emergence of shared languages. On the other hand, if individuals learn from multiple agents in the community (social learning), then shared languages might emerge even in the absence of natural selection.

It is natural to compare linguistic and biological evolution. In biological evolution, one studies how biological (genotypic or phenotypic) diversity evolves under the action of various inheritance mechanisms (sexual and asexual reproduction) and natural selection. In language evolution, one studies how linguistic (syntactic, phonological, etc.) diversity evolves. However, the mechanism of transmission is not inheritance. Rather, it is learning by individual children. Moreover, whereas in biological evolution, one acquires (via inheritance) ones genes from one's parents alone, in linguistic evolution, one might acquire linguistic features from a greater variety of individuals. Further, the sense in which natural selection and fitness may be meaningfully considered in language evolution remains unclear. These similarities and differences have marked the history of both subjects. Since the promulgation of the Indo-European thesis by William Jones, historical linguistics was the pre-occupation of linguists of the nineteenth century. Darwin was clearly influenced by some of these ideas and many times in the *Descent of Man*, he has remarked on these analogies. In the twentieth century, evolutionary ideas were integrated with the genetic and molecular biology revolution. Correspondingly, the traditional questions of nineteenth century linguistics are being reformulated with the insights of modern generative linguistics.

The study of language evolution has a special significance in the scheme of things because it makes it possible for us to transmit information in a non-genetic manner across generations. That is why, as humans, we have such a profound sense of history, culture, and tradition. Learning, rather than

inheritance, is the basis of this transmission of information. It is of interest, therefore, to understand the evolutionary properties of systems where the mechanism of transmission is learning rather than inheritance. This monograph is a contribution to this understanding. More generally, our effort to understand the relationship between the population and the individual is a variation on a theme that cuts across many subjects where one studies the behavior of a complex system of many interaction components. Statistical physics population biology, individual and collective choice in economics, the study of social and cultural norms, are other examples.

This monograph represents a small step towards a larger understanding of the issues in language learning and evolution. This larger understanding will require mathematical models, computer simulations, empirical data analysis, and controlled experiments. The insights will illuminate the nature of communication in humans, animals, and machines. They will have implications for how information is acquired and propagated in linguistics, biology, and computer science.

P.N.

Chicago, Illinois.
December, 2004.

Part I

The Problem

Chapter 1

Introduction

Let us begin with a fact. Shown in 1(a) and 1(b) are two sentences constructed with English words. All native speakers of English today recognize that 1(a) is grammatically well formed ("correct" or "natural") while 1(b) is grammatically ill formed. Following standard practice in the linguistics literature, we have indicated the ill formed expression by a $(*)$.

1 (a) He ran from there with his money.

1 (b) He his money with there from ran. $(*)$

We accept this fact because we know English and this knowledge seems to endow us with the ability to recognize grammaticality and thus separate grammatical sentences from ungrammatical ones. Of course, we weren't born knowing this fact. We learned English as children — presumably from exposure to sentences of English from parents and caretakers. As adults with a mature knowledge of English, we are now able to discriminate between well formed expressions and ill formed ones.[1].

[1]There is often disagreement among researchers and lay people alike regarding the hardness and reliability of grammaticality judgements. Some of this disagreement is well founded and calls for a softer and more nuanced interpretation of grammatical rules. However, from time to time, alarmist have suggested that grammaticality is not a useful notion at all and frequently violated in natural language. Part of this feeling may arise from a confusion between competence and performance issues, between prescriptive and descriptive notions of grammar, between idiolectal and communal languages. Even such alarmists must concede, however, that there are certain expressions that are clearly well formed (such as 1(a)) and some that are clearly ill-formed (such as 1(b)) and about these judgments there may be no reasonable disagreement. This is usually a good starting point from which one can invoke various softer notions of grammaticality possibly using

In normal circumstances children acquire the language of their parents. Thus children growing up today in a relatively homogeneous English speaking environment would learn English from their parents and even if they had not encountered sentences 1(a) or 1(b) before, their judgment on these sentences would agree with that of their parents. That is essentially what it means to learn one's native language.[2] Thus language would be successfully transmitted from parent to child. Indeed, if one polled three consecutive generations of English speakers in an English speaking community, one would find general agreement on the grammatical judgments of 1(a) and 1(b).

Let us imagine an English speaking community today. Let V be the vocabulary, i.e., the set of unique words in English. One can then form strings over V (elements of V^*) and 1(a) and 1(b) are two such strings. Each adult individual in such a community has an internal system of rules (knowledge) that allows him/her to decide which elements of V^* are acceptable and which are not. For individual i let $E_i \subset V^*$ be the set of acceptable sentences for that individual. Correspondingly, I_i is the internal system of rules that characterizes the linguistic knowledge and therefore the extensional set E_i of the ith speaker. The sets E_i might differ slightly from individual to individual but they must largely agree for otherwise speakers would not share the same language[3]. Most of these sets E_i would contain 1(a) but not contain 1(b). Let E denote the intersection of the sets E_i, i.e., $E = \cap_i E_i$. We can interpret E to be the set of sentences that would be considered grammatically acceptable by all people in the community of adults. In fact, we would find 1(a) to be an element of E while 1(b) is not.

Children growing up in such a community would hear the ambient sentences in their linguistic environment from their parents, caretakers, and others they come in contact with. On the basis of such exposure, they too

probability theory as a tool. For a discussion on the role of grammaticality judgments in providing empirical support for various linguistic theories, see Schutze (1996).

[2]One might quibble that children disagree some with their parents. While this is arguably true, this disagreement can never be extreme. Such extreme disagreement would lead to breakdown in successful linguistic communication between parent and child. Note that we use the term parent rather loosely to denote parents, caretakers, and others in the immediate vicinity. Note also that that in linguistically heterogeneous communities, the role of parents may be less important than that of other members. These issues will get clearer as we proceed.

[3]A few remarks are worthwhile. E_i is typically an infinite set for which a finite characterization may be provided by I_i. $E' = \cup_i E_i$ corresponds to the set of all sentences "English speakers" in the community produce. The elements of E_i are observable but the object of fundamental significance is I_i. These distinctions are related to those between E-language and I-language that appear in the work of Chomsky.

would "learn English", i.e., they would acquire a system of rules and correspondingly a language. For the ith such child, let us denote the internal system of rules he/she acquires by I_i^c and the corresponding extensional set by E_i^c. The mechanisms of language acquisition guide the learning child towards the language of the ambient community. Therefore, one ought to find that $E^c = \cap_i E_i^c$ mirrors E. In fact, if one looks at the last hundred years in a relatively homogeneous English speaking community, this seems to be roughly true. Indeed, we are easily able to read English texts from a hundred years ago.

In other words, if all children acquired the language of their parents (read parental generation) and if generation after generation children acquired the language of their parents then language would be successfully transmitted from one generation to the next and the linguistic composition of every generation would look exactly like the linguistic composition of the previous one. A thousand years from now, English speaking communities would still judge 1(a) to be grammatical and 1(b) to be not so. Languages would not change with time.

But they do! In fact, historical linguistics is the study of how, why, when, and in what form, languages change with time.

So let us now go back a thousand years. Shown below is an extract from the *Anglo Saxon Chronicles* taken from the writers of English in 848 A.D. (reproduced from Trask, 1996). The original is italicized and a word for word gloss is provided below.

> Her ... Aelfred cyning ... gefeaht wid ealne, here, and hine
> *Here Alfred king fought against whole army and it*
>
> geflymde and him aefter rad od pet geweorc, and paer saet
> *put to flight and it after rode to the fortress and there camped*
>
> XIII niht, and pa sealde se here him gislas and myccle
> *fourteen nights and then gave the army him hostages and great*
>
> adas, pet he of his rice woldon, and him eac geheton
> *oaths that they from his kingdom would [go] and him also promised*
>
> pet heora cyng fulwihte onfon wolde, and hi paet gelaston
> *that their king baptism receive would and they that did*

It is striking how much the language has changed and at so many different levels that it is barely recognizable as English today. Let us ignore for the moment changes in pronunciation and lexical items and focus instead on the

underlying word order and grammaticality. We have underlined some "odd" portions of the passage for this reason.

Clearly, grammaticality judgments in the ninth century were quite different from those today. The set E describing the language of English speakers in 848 A.D. is quite different from what it is today. There are many points of difference but let us examine a certain systematic difference a little more closely. There are some regularities in the underlying system of rules that characterize "well-formed-ness" (grammaticality) and result in the sets E both of English today and of English in the ninth century. For example, English today has **VO** word order, i.e., the *verb* (**V**) in a verb phrase precedes the *object* (**O**). Thus we have phrasal fragments such as

ate [with a spoon]

kicked [the ball]

jumped [over the fence]

and so on. This fact has received treatment in a variety of linguistic formalisms. For example, getting ahead of ourselves for the moment, we can introduce the notion of the *head* of a phrase which for a verb phrase would be the verb, for a prepositional phrase would be the preposition and so on. English today might be deemed head-first. As a result, in combining words into phrases and ultimately sentences, English speakers put the verb before its object, the preposition before its argument and so on. Some other languages (see Bengali later for example) are the other way around and languages tend to be on the whole fairly systematic and internally consistent on this point. Now consider the following phrasal fragments from English of the ninth century.

pa Darius geseah paet he oferwunnen beon wolde

then Darius saw that [he conquered be would]

(Orosius 128.5)

& him aefterfylgende waes

and [him following was]

(Orosius 236.29)

Nu ic wille eac paes maran Alexandres gemunende beon

now I will also [the great Alexander considering be]

(Orosius 110.10)

Clearly, the language of Old English speakers was underlyingly **OV**. So what went on? These were the kinds of sentences that children presumably heard. The primary linguistic data that children received was consistent with an **OV** type grammar and therefore, this is what we would expect the children to have acquired. If, indeed, English was homogeneous in 800 A.D., and children learned the language of their parents, and their children after them, and so on, why did the language change? These are not changes that are easily explained away by sociological considerations of changing political or technological times, innovations, fads and the like. It is not a word here, an idiomatic expression there, a nuance here, or an accent there — it is deep and systematic change in the underlying word order of sentences — changes that would accumulate over recursions in hierarchically structured phrases leading to such dramatic examples as

> ondraedende paet Laecedemonie ofer hie ricsian mehten swa hie aer dydon
>
> *dreading that Laecedemonians over them rule might as they before did*
>
> "dreading that the Laecedemonians might rule over them as they had done in the past"
>
> (Orosius 98.17)

or

> peh ne geortriewe ic na Gode paet he us ne maege gescildan
>
> *although not shall-distrust I never to-God, that he us not can shield*
>
> "although I shall never distrust God so much as to think he cannot shield us"
>
> (Orosius 86.3)

The phenomena are quite striking and the puzzle is quite real. There are two forces that seem to be at odds with each other. On the one hand we have language acquisition — the child learning the language of its parents successfully. If acquisition is robust and reliable, one would think that language (grammars, linguistic knowledge) would be reliably transmitted from one generation to the next. On the other hand we have language change —

the language of a community drifting over generational time, sometimes just a little bit, sometimes drastically, and sometimes not at all.

And there you have the heart of the problem of historical linguistics. Given that children attempt to learn the language of their parents and caretakers, why do languages change with time? Why don't they remain stable over all time? How fast do they change? In which direction do they change? What are the envelopes of possible change? What are the factors that influence change? These are the kinds of questions that historical linguistics wishes to answer — and indeed, historical linguists over the years have postulated several accounts of documented language change in a number of linguistic subdomains from phonetics to syntax and in a number of different languages of the world.

This book creates a mathematical and computational framework within which to embed those accounts. Such a computational treatment of historical linguistics compels us to make arguments about change precise, work out the logical consequences of such arguments — consequences that might not be obvious from a more informal treatment of the subject. The work in this book is therefore presented as a research tool to judge the adequacy of competing accounts of language change — to aid us in our thinking as we reason about the forces behind such change — to prevent us from falling into the usual pitfalls of Kiplingesque just so stories in an area where data is often sparse and speculation often plentiful. More generally, over the course of this book, we will discuss the themes of learning, communication, language, evolution, and their intertwined relationships. Let us elaborate.

1.1 Language Acquisition

The question of how we come to acquire our native language has received a central position in the current conceptualization of linguistic theory. Learning a language is characterized as ultimately developing a system of rules (a grammar) on the basis of linguistic examples encountered during the learning period. The *language learning algorithm*[4] is therefore a map $\mathcal{A} : D \to g$

[4]The terms learning, acquisition, and development carry different connotations and correspondingly different pictures of the same process. This leads to acrimonious debates and it is safest perhaps to use the more neutral term "map" to denote the procedure that takes linguistic experience (data) as input and produces a computational system (grammar) as output. This map is the learning map, acquisition map, or development map depending upon one's point of view. We will generally use the term learning as well as the metaphors and concepts of learning theory to discuss this map and its consequences.

where D denotes data and g the grammar. What is remarkable about this map is that it involves *generalization*. Of all the different grammars that may be compatible with the data, the child develops a particular one — one that goes beyond the data and one that is remarkably similar to that of its parents in normal and homogeneous environments.

The nontrivial task of generalizing to a grammar from finite data leads to the so called "logical problem of language acquisition". This has received considerable computational attention. Beginning with the work of Gold (1967) and Solomonoff (1964), continuing with Feldman (1972), Blum and Blum (1975), Angluin (1980) on to Jain et al (1998) there exists a rich tradition of research in inductive inference and learnability theory that casts the language acquisition problem in a formal setting that consists of the following key components.

1. *Target Grammar*: $g_t \in \mathcal{G}$ is a target grammar drawn from a class of possible target grammars ($\mathcal{G}$). Grammars are representational devices for generating languages. Languages are subsets of Σ^* where Σ is a finite alphabet in the usual way.

2. *Example Sentences*: $s_i \in L_{g_t}$ are example sentences generated by the target grammar and presented to the learner. Here s_i is the ith example sentence in the learner's data set and L_{g_t} is the target language corresponding to the target grammar.

3. *Hypothesis grammars*: $h \in \mathcal{H}$ are hypothesis grammars drawn from a class of possible hypothesis grammars that learners (children) construct on the basis of exposure to example sentences in the environment. These grammars are then used to generate and comprehend novel sentences not encountered in the learning process.

4. *Learning Algorithm*: $\mathcal{A}$ is an effective procedure by which grammars from $\mathcal{H}$ are selected (developed) on the basis of example sentences received by the child.

These components are introduced to meaningfully discuss the problem of generalization in language acquisition. Consider a somewhat idealized parent child interaction over the course of language acquisition. The parent

It is also worth remarking that the grammar the child develops is probably not the result of conscious meditative deliberation as is the case in developing a strategy for chess. Rather, it is like a reflex — an instinctual reaction to its linguistic environment.

has internal knowledge of a particular language (grammar) so that by his/her reckoning, arbitrary strings of words can be assigned grammaticality values. Thus an English speaking parent would know that sentence 1(a) above was grammatical while 1(b) was not. This language (grammar) is taken to be the target[5] language (grammar) that children must acquire and do in normal circumstances.

In a natural language acquisition setting, children are not directly instructed as to the nature of the grammar that generates sentences of the target language. Rather, they are exposed to sentences of the ambient language as a result of spoken interaction with the world. Thus, their linguistic experience consists of example sentences (mostly from the target language) they hear and this constitutes their so called *primary linguistic data*. On exposure to such linguistic examples, language acquisition is the process by which a grammar is learned (developed, acquired, induced/inferred) so that when novel sentences are produced by parents, children will (among other things) be able to correctly judge their grammaticality and in fact will be able to produce ones of their own as well. This leads to successful ongoing communication between parent and child.

Successful generalization to novel sentences is the key aspect of language acquisition. Thus in our idealized parent child interaction one might imagine that neither sentence 1(a) nor 1(b) was encountered by the learning child over the period of learning English. When the learning period is over, the child's judgment of 1(a) and 1(b) would agree with the parents — the child has been able to go beyond the data to successfully generalize to novel sentences. This is what it means for the child to learn the language of its parents.

Now one has conceptualized the learning procedure of the child as constructing grammatical hypotheses about the target grammar after encountering sentences in the primary linguistic data. Let $h_n \in \mathcal{H}$ be the grammatical hypothesis after the nth sentence. Successful generalization requires at the very least that the learner's hypothesis become closer and closer to the target as more and more data become available. In other words, the

[5]Much of learning theory proceeds with the idealized assumption that there is actually a target grammar. This is a necessary position if one wishes to understand the phenomenon of generalization. This is a reasonable position if one considers an idealized parent child interaction or an idealized homogeneous community. In practice, however, there is always linguistic variation and children acquire a linguistic system from the varied input they receive from the community at large. Therefore, there really is no target in the learning process. Rather, the learning algorithm is a map from data to grammatical systems. In this book we try to understand what happens if we iterate this map in situations that correspond to heterogeneous communities.

learner's hypothesis converges to the target (in some sense indicated here by the distance metric $d(h_n, g_t)$) as the data goes to infinity, i.e.,

$$\lim_{n\to\infty} d(h_n, g_t) = 0 \tag{1.1}$$

Language acquisition is after all a particular cognitive instantiation of a generic problem in learning theory and it is no surprise that the framework here is quite general and applicable to a variety of learning problems in linguistic and nonlinguistic domains. For our purposes, it is important to note that though we have begun on a fairly traditional note with grammars and languages as characterizations of syntactic phenomena, the framework is quite general and is not committed to any particular linguistic theory or even linguistic domain. A number of different aspects of this framework need to be emphasized.

First, $\mathcal{G}$ or $\mathcal{H}$ could represent grammars for syntax in more traditional generative linguistics traditions such as Government and Binding theory (GB), Minimalism, Head driven Phrase Structure Grammar (HPSG), Lexical Functional Grammar (LFG), Tree Adjoining Grammar (TAGS) and so on. It might represent syntactic grammars with less traditional notational systems such as those that arise in connectionist traditions or in recent statistical linguistics traditions.

In the areas of phonology, $\mathcal{G}$ (and correspondingly $\mathcal{H}$) might represent grammars for phonology in any tradition, e.g., Optimality Theory, parameterized theories for metrical stress, Finite State Phonology and so on.

As a matter of fact, $\mathcal{G}$ need not even be a class of symbolic grammars. It might be a class of real valued functions characterizing the decision boundary in some acoustic-phonetic-perceptual space between two phonemic classes. Such a decision boundary also needs to be learned by children in order to acquire relevant phonetic distinctions and build up a phonological system.

Second, the example sentences (where s_i denotes the ith example sentence) might be strings of lexical items, annotated lexical strings, parse trees of example sentences, (form, meaning) pairs such as pairings of syntactic structure with semantic representation and so on. In the case of phonology, they may be surface forms, acoustic waveforms, stress patterns and the like.

Third, the learning algorithm will undoubtedly depend upon the representations used for grammars in $\mathcal{H}$ and examples s_i. Learning algorithms vary from parameter setting algorithms in the Principles and Parameters tradition, constraint reranking algorithms in Optimality Theory, parameter estimation methods based on statistical criteria like Expectation Maximiza-

tion (EM), Maximum Entropy and related methods, gradient descent and Backpropagation in neural networks and so on.

Thus, depending upon the domain and the phenomena of interest, an appropriate notational system for grammars and a cognitively plausible learning algorithm is used in formal explorations in the study of language acquisition. We will encounter several such instantiations over the course of the book.

Finally, the question of generalization characterized by the convergence criterion in Eq. 1.1 can be studied under a number of different notions of convergence. The entire framework can be probabilized so that sentences are now drawn according to an underlying probability distribution. One can then study convergence on all data sequences, on almost all data sequences, strong and weak convergence in probability and so on. The norm in which convergence takes place can vary from extensional set differences (the $L_1(\mu)$ norm where μ is a measure on Σ^* and languages are indicator functions on Σ^*) to intensional differences between grammars as defined by the distance between Godel numberings in an enumeration of candidate grammars.

The resulting learning theoretic frameworks vary from the Probably Approximately Correct framework of Valiant (1984) and Vapnik (1982) to the inductive inference framework of Gold (1967). The necessary and sufficient conditions for successive generalization by a learning algorithm has been the topic of intense investigation by the theoretical communities in computer science, mathematics, statistics, and philosophy. They point to the inherent difficulty of inferring an unknown target from finite resources and in all such investigations, one concludes that *tabula rasa* learning is not possible. Thus children do not entertain every possible hypothesis that is consistent with the data they receive but only a limited class of hypotheses. This class of grammatical hypotheses $\mathcal{H}$ is the class of possible grammars children can conceive and therefore constrains the range of possible languages that humans can invent and speak. It is Universal Grammar (UG) in the terminology of generative linguistics.

Thus we see that there is a learnability argument[6] at the heart of the

[6]This is usually articulated as the Argument from Poverty of Stimulus (APS). There are strong and weak positions one can take on this issue and this has been the subject of much debate and controversy. The theoretical implausibility of tabula rasa learning and the empirical evidence relating to child language development suggest that $\mathcal{H}$ is a proper subset of the set of unrestricted rewrite rule systems (equivalent to Turing Machines). What the precise nature of $\mathcal{H}$ is and whether it admits a low dimensional characterization is a matter of reasonable debate. Over the course of this book, we work with certain

modern approach to linguistics. The inherent intractability of learning a language in the absence of any constraints suggests that the only profitable direction is to try and figure out what the appropriate constraints are. Linguistic theory attempts to elucidate the nature of the constraints that underlie $\mathcal{H}$; psychological learning theory concentrates on elucidating plausible learning algorithms $\mathcal{A}$ and together they posit a solution to the problem of language acquisition.

Language acquisition is the launching point for our discussion on language change. If language acquisition is the mode of transmission of language from one generation to the next, what are its long term evolutionary consequences over generational time? How do these relate to the historically observed trajectories of language change and evolution? This is the primary issue that we will attempt to resolve over the course of this book.

1.2 Variation — Synchronic and Diachronic

A ubiquitous fact of human language is the variation that exists among the languages of the world. At the same time, the fact that language is *learnable* suggests that this variation cannot be arbitrary. In fact, theories of Universal Grammar attempt to circumscribe the degree of variation possible in the languages of the world. Since $\mathcal{H}$ characterizes the set of possible grammatical hypotheses humans can entertain, at any point in time or space, each natural language corresponds to a particular grammar g belonging to $\mathcal{H}$.

For example, shown below are two sentences of Bengali (Bangla) with a word for word translation.

2(a) o or paisa niye shekhan theke dourolo.
He his money with there from ran.

2(b) o dourolo theke shekhan niye or paisa. (*)
He ran from there with his money.

Clearly Bengali has a different system of grammaticality rules from English today so that unlike English, 2(b) is deemed ill formed while 2(a) is well formed. Even if one ignores the fact that the two languages use different lexical items, it is easy to recognize that they use different linguistic (syntactic, in this case) form to convey precisely the same meaning.

plausible choices for illustrative purposes.

The variation across languages might occur at several different levels. For a start, they might have different lexical items. Further, the system of rules that determine grammaticality might consist of phonetic, phonological, syntactic, semantic, pragmatic and other considerations. Two languages might have different lexicons but similar syntactic systems as is the case for Hindi and Urdu, two languages spoken in large parts of South Asia. They might also have similar lexicons but different syntactic systems as is often the case for dialects of the same language. Or they might share similar lexical and syntactic properties yet have have very different phonological systems as is the case for the different "accented" Englishes spoken around the world. While the modules governing the different aspects of the grammatical system of a language all need to be specified to define a fullblown grammar in UG, in particular inquiries of linguistic phenomena, one considers $\mathcal{H}$ to cover the variation that is relevant depending upon the domain under consideration.

Our discussion so far has been as if languages have an existence that is independent of the individuals that speak them. Perhaps it is important to clarify our point of view. $\mathcal{H}$ denotes the set of possible linguistic systems that humans may possess. In any community, let the ith individual possess the system $g_i \in \mathcal{H}$. In a homogeneous community most of the g_i's are similar and one might say that these individuals speak a common language and terms like English, Spanish, etc. refer to these communally accepted common languages. In general, though, there is always variation and these variants may be referred to as different idiolects, dialects, or languages based on social and political considerations. This variation refers to the *synchronic* variation across individuals in space at any fixed point in time.

This book concerns itself with variation along a different dimension — the variation in the language of spatially localized communities over generational time. Thus one could study the linguistic behavior of the population of the British isles over generational time and as we remarked in our opening section, this has shown some striking changes over the years. Indeed, historical phenomena and *diachronic* variation are properly the objects of study in historical linguistics and this book presents a computational framework in which to conduct that study. Since we try to understand and reason about the possible behaviors of linguistic systems changing with time, we will be led to a dynamical systems framework and will derive several such dynamical systems over the course of this book.

The starting point for the derivation of such dynamical systems is a class of grammars $\mathcal{H}$ and a learning algorithm $\mathcal{A}$ to learn grammars in this class. To see the interplay between the two in a population setting, imagine for

a moment that there were only two possible languages in the world, i.e., $\mathcal{H} = \{h_1, h_2\}$ defining the two languages $L_{h_1} \subset \Sigma^*$ and $L_{h_2} \subset \Sigma^*$ over a finite alphabet Σ.

Consider now a completely homogeneous linguistic community where all adults speak the language L_{h_1} corresponding to the grammar h_1. A typical child in this community receives example sentences and utilizing a learning procedure $\mathcal{A}$ constructs grammatical hypotheses. Let us denote by h_n the grammatical hypothesis the learning child has after encountering n sentences. Suppose that each child is given an infinite number of sentences to acquire its language so that $\lim_{n\to\infty} d(h_n, h_1) = 0$, i.e., the child converges to the language of the adults. This happens for all children and the next generation would consist of homogeneous speakers of L_{h_1}. There would be no change.

Now consider the possibility that the child is not exposed to an infinite number of sentences but only to a finite number N after which it matures and its language crystallizes. Whatever grammatical hypothesis the child has after N sentences, it retains for the rest of its life. Under such a setting, if N is large enough, it might be the case that most children acquire L_{h_1} but a small proportion ϵ end up acquiring L_{h_2}. In one generation, a completely homogeneous community has lost its pure[7] character — a proportion $1 - \epsilon$ speak the original language while a proportion ϵ speak a different one.

What happens in the third generation? Will the proportion ϵ grow further and eventually take over the population over generational time? Or will it decrease again? Or will it reach a stable ϵ^*? Or will it bounce back and forth in a limit cycle? It will obviously depend upon how similar the two languages L_{h_1} and L_{h_2} are, the size of N, the learning algorithm $\mathcal{A}$, the probability with which sentences are presented to the learner and so on. In order to reason through the possibilities, one will need a precise characterization of the dynamics of linguistic populations under a variety of assumptions. We will consider several variations to this theme over the course of this book.

Even a simplified setting like this is not without significant linguistic applications. In a large majority of interesting cases of language change, two languages or linguistic types come into contact and their interaction can then be tracked over the years through historical texts and other sources.

[7]It is worth noting that one need not necessarily consider starting conditions that are homogeneous. The dynamics will relate the linguistic states of any two successive generations. One may then consider these dynamical systems from any initial condition — those that relate to mixed states corresponding to language contact may be of particular interest.

For example, in the case of English, it is believed that there were two variants of the language — a northern variant and a southern one that differed in word order and grammaticality and contact between these two led to one variety sweeping through the population. We will consider this and several other cases in greater detail over the course of this book.

1.3 More Examples of Change

The case of English syntactic change with which we opened this chapter is only one of a myriad of cases of historical change across linguistic communities of the world for which documented evidence exists in various forms. It is important therefore to recognize that there is a genuine phenomenon at hand here and the pervasiveness of such phenomena is important to emphasize. Let us consider here some more examples drawn from different linguistic subsystems and different geographical regions of the world. They present interesting puzzles to work on.

1.3.1 Phonetic and Phonological Change

The earliest studies of historical change were often in the domain of sound change — phonetic and phonological changes occurring in various languages and the Neogrammarian enterprise of the early twentieth century brought it to the center stage of historical linguistics.

The Great English Vowel Shift

In the Middle English period from the fourteenth to the sixteenth century, the long vowels of English underwent a cyclic shift so that pronunciations of words using these long vowels changed systematically. A simplified version of the cyclic shift of vowels is shown below (for more details, see Wolfe, 1972):

Back Vowels

The back vowels are produced with the tongue body at the back of the vocal cavity resulting in a lowered first formant (Stevens, 1998). We consider in this system the following four vowels: (i) the diphthong $/a^u/$ as in the modern English word "loud" (ii) $/aw/$ as in the modern English word "law" (iii) $/o:/$ as in the word "grow" and (iv) $/u:/$ as in "boot". The pronunciations of the words in the ME (Middle English) phonological system went through the following cyclic shift.

$$/a^u/ \rightarrow /aw/ \rightarrow /o:/ \rightarrow /u:/ \rightarrow /a^u/$$

$/a^u/ \rightarrow /aw/$	$/aw/ \rightarrow /o:/$	$/o:/ \rightarrow /u:/$	$/u:/ \rightarrow /a^u/$
law	grow	boot	loud
saw	mow	moot	proud
bought	hose	loose	house

Table 1.1: A partial glimpse of the vowel shift in Middle English. Words which share the same vowel are shown in each column. Each of these words went through a systematic change in pronunciation indicated by the vowel shift shown in the top row. Thus "grow" (pronounced $/gro:/$ today) was pronounced $/graw/$ before. Words pronounced with an $/o:/$ before are pronounced with an $/u:/$ today (words in the third column).

Thus, the word "law" (pronounced $/law/$ today) was pronounced differently as $/la^u/$ in M.E. Consider Table 1.1.

A similar cyclic shift occurred for front vowels as well. Thus at one point in time (before the fourteenth century) speakers in England pronounced words in a particular way using a vocalic system that was in place at the time. Consider a random child growing up in such an environment. Such a child would have presumably heard "house","mouse","proud","loud" all being pronounced with the vowel $/u:/$. Why would they not learn the same pronunciation?

One might argue that the actual pronunciation of words is sometimes sloppy and therefore listeners might misperceive the pronunciations of the words. However, sloppy pronunciation might have a random distribution around the canonical pronunciation and in that case it is not clear at all that such random mispronunciation effects would have a directional effect and systematicity that would accumulate over generations. Even if a few children misconverged, what is the guarantee that the new pronunciation system will actually spread through the population over generational time?

One might reasonably argue that there was either language variation or language contact resulting in two pronunciation systems existing in the population at some time and competition between these two systems led to the gradual loss of one over time. In the absence of a deeper analysis, this would again be handwaving our way around the problem.

Then there is the matter of the cyclical nature of the change. Such cyclic changes are often referred to as *drag chains* in the historical linguistics literature. Because a particular vowel shifts, i.e., a pronunciation changes, it leaves a gap in the vowel system with an UN utilized vowel. At the same time,

unless other vowels shift too, a number of homophonous pairs will be created leading perhaps to possible confusion. Imagine for a moment that /o : / changed to /u : / in word pronunciations. Therefore "boat" and "boot" would be a homophonous pair (we are considering modern pronunciations here to make the point). In order to eliminate confusion, perhaps, speakers and listeners will feel compelled to shift the pronunciation of "boot". This might now create new confusions ("boot" with "bout", for example) which need to be eliminated leading to further changes and so on in a chain reaction to the first change from /o : / to /u : /. Again, a number of questions arise. Why, for example, don't speakers and listeners simply exchange the vowels /o : / and /u : /. That would fill the gap in the vowel system, eliminate homophony, and present a satisfactory solution.

In order to reason coherently through the various possibilities without resorting to handwaving arguments, one will need to tease apart several notions: (i) individual learning by children (ii) tendencies by speakers, listeners, and learners to avoid gaps and reduce homophonies (iii) the fact that words are used with varying frequencies and some vowel mergers might have greater consequences for communication than others and (iv) the effect of all of the above at a population level leading to systematicities in population behavior. It is almost impossible to examine the interplay between these factors by verbal argument alone. One will be compelled, therefore, to consider computational models in the spirit of those developed over the course of this book.

Phonological Mergers and Splits

Phonological mergers occur when two phonemes that are distinguished by speakers of the language stop being distinguished. This implies that certain acoustic-phonetic differences are no longer given phonological significance by users of the language. The reverse process occurs when a phoneme (typically allophonic variations) splits into two. Some examples of historical change along this dimension are illustrated below.

Sanskrit, Hindi, and Bengali

In Sanskrit, there were (and are) three different unvoiced strident fricatives that vary by place of articulation. These are shown below with the point of constriction of the vocal tract in producing these sounds varying from the front of the cavity to the back from 1 through 3.

1. /s/ alveolar-dental as in *sagar* (sea)
2. /xh/ retroflex as in *purush* (man)
3. /sh/ postalveolar as in *shakti* (energy)

Phonological mergers have occurred in two descendents of Sanskrit — Hindi and Bengali. In Hindi, the retroflexed and postalveolar fricatives have merged into a single postalveolar (palatal) one so that the "sh" in *purush* is pronounced identically to the "sh" in *shakti*. Thus there are only two strident (unvoiced) fricatives in the phonological system of the language. In Bengali, all three have merged into a single palatal fricative so that the fricative in *sagar*, *purush*, and *shakti* are all pronounced identically. The words in question are Sanskrit originals that have been retained in the daughter languages with altered pronunciations.

Interestingly, the orthographic system used in writing preserves the distinction between each of the three different fricatives so a different symbol is used for the fricatives in 1,2,and 3 although they are pronounced in the same way by Bengali speakers. Similarly, Hindi inherited the Devanagari orthographic system of Sanskrit and distinguishes the fricatives in the written form although 2 and 3 have merged.

An example of a similar merger can be considered from Spanish where an ancestral form of the language had both /b/ and /v/ as distinct phonemes. In the modern version of the language, these phonemes are merged. However, the old spelling has been retained so that *boto* (meaning "dull") and *voto* (meaning "vote") are spelled differently yet both are pronounced with a word initial /b/ by modern Spanish speakers.

Wu Dialect in Wenzhou province

Zhongwei Shen (1997) describes two detailed studies of phonological change in the Wu dialects. We consider here as an example the monophthongization of $/o^y/$ resulting in a phonological merger with the rounded front vowel /o/. This sound change is apparently not influenced by contact with Mandarin and is conjectured to be due to phonetic similarities between the two sounds. These two phonological categories were preserved as distinct by many speakers, but over a period of time, the distinction was lost and their merger created many homophonous pairs.

Thus, the word for "cloth" — $/po^y/^{42}$ — now became homophonous with the word for "half" — $/po/^{42}$ and similarly, the word for "road" — $/lo^y/$ became homophonous with the word for "in disorder" — $/lo/^{11}$. A list of 35 words with the diphthong $/o^y/$ is presented in Z. Shen (1993) and some

$/po^y/^{42}$	"cloth"
$/do^y/^{31}$	"graph"
$/mo^y/^{31}$	"to sharpen"
$/to^y/^{42}$	"jealous"
$/so^y/^{42}$	"to tell"

Table 1.2: A subset of the words of Wu dialect that underwent change over the last one hundred years. The vowels were all diphthongs that changed to monophthongs. The numeric superscript denotes the tonal register of the vowel (unchanged).

of these are reproduced in Table 1.2.

The phonetic difference between the two sounds lies in movements of the first and second formants. Both of the sounds in question are long vowels. The monophthong $/o/$ has a first formant at around 600 Hz. and a second formant at 2200 Hz. The diphthong $/o^y/$ has a first formant that starts around 600 Hz and gradually drops down to 350 Hz while the second formant increases slightly above 2200 Hz. The change from the diphthong to monophthong can in principle be gradual with no compelling phonetic reason to make this change abrupt.

Each word participating in the change has two alternative pronunciations in the population. An *original* pronunciation using the diphthong and an *altered* pronunciation using the monophthong. At one point all speakers used the original pronciation. Gradually speakers adopted the other pronunciation and today, everyone uses the monophthongized pronunciation of the word. We consider this example in some detail later in the book (Chapter 8). In particular, we will examine several plausible learning mechanisms and work out their evolutionary consequences for the case when two distinct linguistic forms are present in the population. By doing so, we will arrive at a better understanding of the stable modes of the linguistic population and under what conditions a switch from one stable mode to another might happen.

Other Assorted Changes

A wide variety of phonetic and phonological changes have been studied and discussed in the rich literature on historical linguistics and language change.

Let us briefly consider a few more examples for illustrative purposes.

Yiddish is descended from Middle High German (MHG) which is itself descended from Old High German (OHG). In OHG, words could end in voiced obstruents, e.g., *tag* for "day". A change occured from OHG to MHG so that word final voiced obstruents were devoiced. (See discussion in Trask, 1996). Thus *tag* became *tac*, and *gab* ("he gave") became *gap*, *weg* ("way") became *wec*, *aveg* ("away") became *avec* and so on. This can be expressed as the rule

$$[+\text{obstruent } +\text{voice}] \longrightarrow [+\text{obstruent } -\text{voice}] \mid \_ \#$$

Of course, voiced obstruents that are not in word final position remain voiced. Thus the plural forms of the words are *tage* ("days") and *wege* ("ways"). Modern German retains this rule. In Yiddish, on the other hand, the forms of the same words are *tog* ("day"), *weg* ("way") and so on. At the same time, words without alternations[8] such as *avec* are pronounced with a voiceless stop.

Consider now the sequence of transformations from OHG to MHG to Yiddish. The devoicing rule was *added* from OHG to MHG. Now one could postulate (i) that a new rule was added in Yiddish so that word final unvoiced consonants were voiced (ii) the devoicing rule that was added in MHG was simply lost again. If (i) were true, then it would not explain why *avek* remain unvoiced. Therefore (ii) must be closer to the truth. A plausible explanation is that words where alternations provide clues as to their underlying form (such as *tac-tage*) were reanalysed as voiced. Words without alternations suggesting the possibility of a voiced underlying form were analysed as unvoiced. Since the devoicing rule was lost anyway, the reanalyzed form was not subject to devoicing and this explains the modern form of Yiddish words.

A number of issues now arise. Rules are part of the phonological grammatical system. Why would rules arise and be lost? One explanation might lie in variation existing in the population. Perhaps, some portion of the population had the devoicing rule and some did not. Given conflicting data, it is possible that some children acquired the devoicing rule while others did not. How might learning by children, frequency of usage of different forms, and variation in the population interact to create the circumstances under which

[8]Root words that had inflections where the relevant obstruent was voiced in some cases but not in others are referred to as alternations. Thus *tac-tage* and *wec-wege* are alternations.

a rule might be gained and the circumstances under which a rule might be lost? We need ways for thinking about these issues in order to sharpen our understanding of the factors involved.

Like the Chinese example of Shen described earlier, another example of a linguistic change in progress comes from William Labov's pioneering study of vowel centralization in Martha's vineyard. In the speech of Martha's vineyard, the diphthongs /*ai*/ and /*au*/ as in "light" and "house" are centralized. This is unusual for New England (though quite common in Canada) and Labov studied a large number of subjects of varying ages with respect to the degree of centralization of each of the diphthongs. A measure of centralization called the centralization index was constructed and could be plotted for each subject by age. Strikingly, it was observed that centralization decreased with age with the oldest group having the lowest CI. The youngest group however had a low CI index too suggesting that centralization increased over time and then started decreasing again. This can be related to occupations, social stratification, degree of identification with the island and serves as an example of social forces interacting with linguistic forces that has been studied in a quantitative manner in the sociolinguistic tradition pioneered by Labov (see Labov (1994) for an account.).

1.3.2 Syntactic Change

As we have discussed before, changes in the grammatical properties of linguistic populations occur in many different linguistic domains and here we review some cases of syntactic change.

French

Old French had a number of properties including (i) V2 — the tendency of verbs (finite) to move to second position in matrix clauses (ii) *pro-drop* — the ability to drop the pronominal subject from a sentence without sacrificing the grammaticality of the resulting expression.

Let us examine the case of *pro-drop* for a moment. In some languages of the world, like Modern English for example, it is required that the pronominal subject of a sentence be present in the surface form for the sentence to be deemed grammatical in that language. Thus in the English sentences below, 3(a) is grammatical while 3(b) is not.

3(a) He went to the market.

3(b) Went to the market. (*)

Modern Italian on the other hand, allows one to drop the subject if the putative subject can be unambiguously inferred by pragmatic or other considerations. Thus both 4(a) and 4(b) (meaning "I speak") are grammatical.

4(a) Io parlo.

4(b) Parlo.

Or consider another Italian sentence with pro-drop.

Giacomo ha detto che ha telefonato.

Giacomo has said that (he) has telephoned.

It has been suggested that this aspect of syntactic structure defines a typological distinction between languages of the world with some allowing *pro*-drop and others not.

It turns out that Old French used to allow *pro*-drop while Modern French does not. Consider the following two sentences taken from from the discussion on French change in Clark and Roberts (1993).

Ainsi s'amusaient bien cette nuit.

thus (they) had fun that night.

and

Si firent grant joie la nuit.

thus (they) made great joy the night.

Both these sentences are ungrammatical in modern French. Again we are led to the usual puzzles. At one time, French children would have had enough exposure to the language of their times that they would have learned that *pro*-drop was allowable and acquired the relevant grammatical rule, much as Italian children do today. Why then did they stop acquiring it? May be a few didn't acquire it, the frequency of usage became rare, it triggered the rule in less and less children as time went on, and ultimately it died out. This story needs to be made more precise with data, models, and a deeper understanding of the interaction of learning, grammar, and population dynamics. Clark and Roberts (1993) and Yang (2002) have taken steps in this direction and we will revisit this problem later in the book.

Yiddish

Yiddish is the language of Jews of eastern and central Europe and is descended from medieval German with considerable influence from Hebrew and Slavic languages as well. Like English and French, Yiddish too underwent some remarkable syntactic changes leading to different word order formations in the modern version of the language.

One particular change had to do with the location of the auxiliary verb with respect to the subject and the verb phrase in clauses. Following Chomsky (1986) one might let the auxiliary verb belong to the functional category INFL (that bears inflectional markers) and thus distinguish between the two basic phrase structure alternatives as in 5(a) and 5(b).

5(a) $[\text{Spec } [\text{VP INFL}]]_{\text{IP}}$

5(b) $[\text{Spec } [\text{INFL VP}]]_{\text{IP}}$

The inflectional phrase (IP) describes the whole clause (sentence) with an inflectional head (INFL), a verb phrase argument (VP) for this INFL head and a specifier (Spec). The item in specifier position is deemed the subject of the sentence. In modern English, for example, phrases are almost always of type 5(b). Thus the sentence (6)

(6) $[\text{John } [\text{can } [\text{read the blackboard }]_{\text{VP}}]]_{\text{IP}}$

corresponds to such a type with "John" being in Spec position, "can" being the INFL-head and "read the blackboard" being the verb phrase. If we deem structures like 5(a) to be INFL final and structures like 5(b) to be INFL medial, we find that languages on the whole might be typified according to which of these phrase types is preponderent in the language.[9]

Interestingly, Yiddish changed from a predominantly INFL final language to a categorically INFL medial one over the course of a transition period from 1400 A.D. to about 1850 A.D. Santorini (1993) has a detailed quantitative

[9]It should be mentioned that while this typological distinction is largely accepted by linguists working in the tradition of Chomsky (1981), there is still considerable debate as to how cleanly languages fall into one of these two types. For example, while Travis (1984) argues that INFL precedes VP in German and Zwart (1991) extends the analysis to Dutch, Schwartz and Vikner (1990) provide considerable evidence arguing otherwise. Part of the complication often arises because the surface forms of sentences might reflect movement processes from some other underlying form in often complicated ways. But this is beyond the scope of this book.

Time Period	INFL-medial	INFL-final
1400-1489	0	27
1490-1539	5	37
1540-1589	13	59
1590-1639	5	81
1640-1689	13	33
1690-1739	15	20
1740-1789	1	1
1790-1839	54	3
1840-1950	90	0

Table 1.3: Relative numbers of INFL-medial and INFL-final structures in clauses with simple verbs (at different points in time). Taken from the study of the history of Yiddish in Santorini (1993).

analysis of this phenomenon and shown below are two unambiguously INFL final sentences of early Yiddish (taken from Santorini, 1993). Such sentences would be deemed ungrammatical in the modern categorically INFL medial Yiddish.

ds zi droyf givarnt <u>*vern*</u> (Bovo 39.6, 1507).

that they there-on warned were

ven der vatr nurt doyts leyan <u>*kan*</u> (Anshel 11, ca. 1534).

if the father only German read can

To illustrate this point quantitatively, a corpus analysis of Yiddish documents over the ages yields the following statistics shown in Table 1.3. Clauses with simple verbs are analyzed for INFL-medial and INFL-final distributions of phrase structures.

More statistics is available in Santorini (1993) but this simple case illustrates the clear and unmistakable trend in the distribution of phrase types. It is worthwhile to mention here that while Santorini (1993) expresses the statistics within the notational conventions of Chomsky (1986), almost any reasonable grammatical formalism would capture this variation and change with two different grammatical types or forms in competition with one gradually yielding to the other over generational time. Again one might wonder

about the causes of such a change, the stable grammatical modes of populations, the directionalities involved, and the like. As quantitative measures of the sort described here are made to characterize the historical phenomena at hand, one is led irrevocably towards quantitative and computational models to attain a deeper understanding of the underlying processes involved.

1.4 Perspective and Conceptual Issues

This book is a computational treatise on historical and evolutionary phenomena in human language. At the outset it may not be entirely clear that there are meaningful computational questions and that such a computational treatment is possible, profitable, or necessary in the discourse on historical linguistics. After all, one does not typically study human social and political history with computational tools. On the other hand, evolutionary biology is today a heavily mathematized discipline. In fact the mathematization of evolutionary biology began in the early part of the twentieth century to resolve the apparent conflict between the ideas of Mendel, Darwin and other evolutionary thinkers — conflicts that were difficult to resolve by verbal reasoning alone.

Human language is interesting because it is in part cultural and in part biological. That part which is biological belongs more readily to the natural sciences and is amenable to a treatment by the usual modes of inquiry in the natural sciences. We have tried to illustrate in previous sections some of the examples of language change that belong to this domain and some of the issues that arise in the study of such phenomena for which a computational analysis becomes necessary. The overall rationale behind such an approach and the possibility of a computational treatment rests on three aspects of language that are central to our point of view and worth highlighting separately.

1. Language has form. The linguistic objects of distinctive features, phonemes, syllables, morphemes, words, phrases, and sentences have reasonably concrete representations and display systematic regularities that give language form. This formal aspect separates it from amorphous cultural convention and makes it amenable to study by formal or mathematical means. Indeed, the discipline of formal language theory evolved in part to provide the apparatus to describe this formal structure and associated linguistic phenomena. Interestingly enough, grammars, automata, and languages are central also to investigation

in logic and computer science and many of the ideas we present in this book are possible to articulate only because of this link between computer science and linguistics.

One might quibble about the details of this form, about grammaticality judgements, about competence and performance issues, about functionality issues. One might argue that the true goal of language is communication after all, that the meaning of sentences is paramount and their form not all that sacrosanct. However, one will still have to concede that we don't speak word salad. Of all the different ways of conveying the same meaning, a particular language will choose a limited number of ways to give form to that meaning. Thus English chooses 1(a) while Bengali 2(b) — it could easily have been the other way around and indeed in 800 A.D. it was. When one moves away from syntactic to phonological phenomena the link between form and meaning becomes even more remote and it is in some ways easier to recognize this strict yet arbitrary formal aspect of language in phonological systems.

2. Language is learned. Unlike other modalities like vision or olfaction, where the role of learning is unclear beyond some plasticity in the neural apparatus, language is clearly and indisputably learned. When we are born we don't know language. We are exposed to linguistic data and we learn it. In many ways, it fits quite neatly into the framework of learning from examples and in fact the field of formal inductive inference arose to study the tractability of the problem of language acquisition.

   The ability to learn has been a central topic of investigation in artificial intelligence and a variety of computational tools ranging from abstract theory to computer simulation have been brought to bear in this enterprise.

3. Languages vary. Variation across the languages of the world is a ubiquituous fact of human existence. In many ways it might have been quite convenient if they did not vary at all. If there was one pure, fit, language that was hardwired in our genes and we all grew up speaking the same language. While this is not true, in some ways perhaps it is not far from the truth for while we are not born with the details of a particular language, it is likely that we are born with the class $\mathcal{H}$ that limits possible variations in some sense. This book attempts to create

the computational framework for studying diachronic variation.

Thus the mathematical and computational tools that will be utilized to characterize each of these aspects of language are

1. *Formal Language Theory* and related areas to describe linguistic form and linguistic structures.

2. *Learning Theory* to characterize the problem of language acquisition and learning.

3. *Dynamical Systems* to characterize the diachronic evolution of linguistic populations over time.

In the rest of the book we will see how these different areas of mathematics come together in our computational approach to the problem. As we proceed, we will need to tease apart several issues that need to be kept in mind for a complete treatment of historical phenomena in linguistics. Indeed, historical linguists have considered these at various points in time.

1.4.1 The Role of Learning

Clearly language is acquired by children — most significantly from the input provided by the previous generation of speakers in the community. The idea that language change is contingent on language learning has been a long standing one. As early as the nineteenth century we have the following scholars

> ..the main occasion of sound change consists of the transmission of sounds to new individuals
>
> (Paul, 1891, pp.53-4)

and more strikingly, from the British linguist, Henry Sweet, we have

> ...if languages were learned perfectly by the children of each generation, then language would not change: English children would still speak a language as old atleast as Anglo Saxon and there would be no such languages as French or Italian.
>
> (Sweet, 1899, pg. 75)

More recently, Halle (1962), Kiparsky (1965), Weinreich et al (1968), Wang (1969,1991), Ohala (1993) have invoked it in explicit or implicit ways in the phonological domain. Similarly, in syntax, Lightfoot (1979, 1998),

Roberts (1992), Kroch (1989,1999), Yang (2002) among others have argued this connection strongly. This book contributes to the effort to explore systematically the precise nature of the relationship between language acquisition and language change.

1.4.2 Populations Versus Idiolects

Isolated instances of mislearning or idiosyncratic linguistic behavior is clearly of little consequence unless it spreads through the community over time to result in large scale language change. In any meaningful discourse on language change, one therefore needs to distinguish between the population and the individuals in it. Individual speaker-hearers (language users) might vary from each other at any single point in time and this characterizes the synchronic variation in the population at that point in time. However, one can also discuss average characteristics of the population as a whole and in some sense, when one talks about a language changing with time, one is talking about the average characteristics of the population changing over successive generations. After all, an individual occupies only one generation.

Historical linguistics often confuses this issue. Part of the reason is that our data about language change often comes from individual writers. Strong trends in different individual writers over successive generations are certainly suggestive of larger scale population level effects but don't necessarily imply it. Mufwene (2001), Labov (1994), and Shen (1997) have in various ways emphasized this difference. Shen (1997) provides the source of the Wenzhou data that is discussed in a later chapter. The data arose by explicitly sampling multiple people in the population for each generation. An important goal of this book is to explore the relationship between change at the individual level and change at the population level.

1.4.3 Gradualness Versus Abruptness (or the S-shaped curve)

The rate and time course of language change has been the object of study and speculation by historical linguists for some time. Since most linguistic changes are ultimately categorical ones, the possibility exists for a language to change categorically — and therefore abruptly — from one generation to the next. Empirical studies of the process have always yielded, however, a more graded behavior and much has been made of the so-called S-shaped curve denoting the change in linguistic behavior (average population behavior, typically) over successive generations. Bailey (1973) discusses the

S-curve:

> A given change begins quite gradually; after reaching a certain point (say, twenty percent), it picks up momentum and proceeds at a much faster rate and finally tails off slowly before reaching completion. The result is an S-curve, ... (1973, p. 77)

Similarly, we have Osgood and Seboek (1954) discuss the S-shaped nature of change while introducing the notion of community (population) and the possibility of change being actuated by children (learning):

> The process of change in a community would most probably be represented by an S-curve. The rate of change would probably be slow at first, appearing in the speech of innovators, or more likely young children; become relatively rapid as these young people become the agents of differential reinforcement; and taper off as fewer older and more marginal individuals remain to continue the old forms.

Weinreich, Labov, and Herzog (1968) also discuss the S-shaped curve thus:

> ...the progress of language change through a community follows a lawful course, an S-curve from minority to majority to totality (1968, p. 133).

As we see, for some time now, there has been a discourse on the importance and pervasiveness of the S-shaped change in historical linguistics. Of course, the "knee" of the S could be sharp reflecting a sudden transition from one linguistic usage to another or it could be gradual over many centuries. Lightfoot (1998) argues that many of the changes in English syntax from Old to Middle to Modern English were actually quite categorical and sudden.

Why should the changes be S-shaped? A historicist account would claim this to be one of the "historical laws" that govern language change with time. An alternative position — and one we explore in this book — would consider this to be an epiphenomenon. We attempt to derive the long term evolutionary consequences of short term language learning by children. As a result we develop some understanding of when we expect trajectories to be S-shaped. The collection of quantitative historical (or pseudo-historical) datasets along with an in interest in explaining qualitative S-shaped behavior has prompted researchers in recent times (Kroch, 1989; Shen, 1997) to explore mathematical models of the phenomena. We discuss them at a later point in this book.

1.4.4 Different Time Scales of Evolution

It is worth noting that there are two distinct time scales at which one can study the evolutionary history of linguistic systems. One time scale corresponds to historical linguistics, i.e., the period after modern humans arose and the human language faculty was in place. Much of our discussion so far has been at this time scale. We have seen how the linguistic systems of humans living in different geographical regions have undergone change (evolution) over time.

A second time scale corresponds to the origin and evolution of the human language faculty from pre-linguistic versions of it that may have presumably existed in our pre-human ancestors. In a discussion of the major transitions of evolution, Maynard Smith and Szathmary (1995) consider the evolution of human language to be the last major transition.

These two time scales present interesting similarities and differences. In both, one needs to concern oneself with population dynamics, individual learning, social networks, and linguistic systems. However, it is likely that on historical time scales, natural selection (differential reproductive fitness based on communicative advantage) is less important than it is on evolutionary time scales. Another matter of signficant importance is the range of data available to empirically ground theories and explanations. If one is interested in human language, much more data is available at historical time scales while almost none is available at evolutionary time scales. For studies at evolutionary time scales, therefore, one will probably have to resort to cross-species comparative studies across different animal communication systems(see Hauser, 1997 for this point of view). What can be said about human language as a result of such comparisons remains unclear. In parts II and III of this book, our discussion is mostly about human language and examples from various linguistic systems are provided. The discussion in part IV, however, has considerable relevance to both animal and artificial communication systems and should be read with this thought in mind.

1.4.5 Cautionary Aspects

Long term change of a language over time is complicated by several compounding factors. First, socio political considerations often enter the picture. The undue influence of one person or group of persons on society might result in the propagation of their linguistic preference across the population over time. Prestige, power, influence are difficult to formalize and model precisely

and often best left alone in this regard. We will concentrate in this book on those kinds of phenomena for which we believe a linguistic rather than extra-linguistic (sociological) explanation is possible or likely. Nevertheless, we are acutely aware that explanatory possibilities from socio-political considerations need to be carefully considered at all times for "naturalistic" explanations are often proferred too eagerly when the underlying causes reside elsewhere. As a matter of fact, the interaction of social forces with linguistic considerations is explored fully in the kind of quantitative sociolinguistic work pioneered by Weinreich, Herzog, and Labov (1968) and discussed at length in Labov (1994).

A second complication arises from the nature of the data available and the testability of theories. Because the discipline is inherently a historical one, it is hard to replay the tape of life or conduct experiments of any sort[10]. At the same time, historical records often show clear patterns of regularity — with data so strikingly regular and abundant that the force of the phenomena become compelling. Of course this problem is not peculiar to linguistics and is shared by all scientific disciplines that focus on historical phenomena from evolutionary biology to cosmology. In recent times, the collection of large electronic corpora of linguistic facts and documents (see, for example, the collections of the Linguistic Data Consortium) has also provided fruits for historical linguistics. The Penn Helsinki corpus of Middle English consists of a million parsed sentences from a variety of texts in the Middle English period from which it is possible to collect statistics on the frequency of various kinds of constructions and track their change with time. This is beginning to be repeated for a number of other languages. For example, texts of European Portuguese in the period from 1600 A.D. to 1800 A.D. have been collected and are beginning to be annotated and made available in computer format as part of the Tycho Brahe project (Galves and Galves; see http://www.ime.usp.br/∼tychobrahe). Field studies of languages changing with time in the last 50 years have been conducted in a variety of languages from creoles (see Mufwene, 2001; DeGraff, 2001; Rickford, 1987), sign languages (Senghas and Coppola, 2001), Chinese (Shen, 1997), American English dialects (Labov, 1994; Christian et al, 1988), British English (Milroy and Milroy, 1985, 1993; Bauer, 1994) and so on that provide the empirical base on which historical linguistics and language change is founded. It is not possible to do justice to the variety of such field studies and we will cite and

[10]See Ohala (1993) for an interesting suggestion of laboratory experiments simulating sound change.

deal with only a limited number of case studies over the course of this book.

1.5 Evolution in Linguistics and Biology

Each of the issues discussed above arises albeit in a different form in the domain of evolutionary biology.

Heritability and modes of transmission of genetic information leading to the similarity between children and parents is a crucial feature of biological populations. A variety of such modes of transmission from sexual to asexual reproduction have been considered by evolutionary biologists. The interaction of genetically transmitted information and inputs from the environment contributes to the developmental sequence of a biological organism and its ultimate mature form. In the case of human language, the mode of transmission is learning rather than genetic reproduction. Learning by children results in linguistic similarity between parent and child. However, parents are not the only influence on the child's linguistic development and the general linguistic environment of the growing child plays a role as well. Thus while sexually reproducing biological organisms inherit their genetic makeup from their parents, children acquire their language based on the linguistic composition of the parental generation at large. For this reason, it is likely that language evolution is more like epidemiology or ecology than like Mendelian genetics.

Population thinking pervades all of biology. The individual organism and the population of which the organism is a part are distinguished and the entire field of population biology attempts to work out the population dynamics resulting from individual interactions that may vary from biological reproduction to strategies for survival in predator prey systems. From gradualness to punctuated equilibria, biologists have pondered the various dynamical aspects of changing populations (see Hofbauer and Sigmund, 1988 or Nagylaki, 1992 for mathematical reviews). Because individual learning is the mechanism by which language is transmitted from the speakers of one generation to those of the next, the theory of learning will play a central role in the development of the evolutionary models that we consider in this book. As a result, the precise nature of these models and the mathematics that surrounds their analysis are quite distinct from that encountered in the literature on evolutionary biology.

The importance of the population as an object of study results in an emphasis on observing and characterizing variation and typology within the

population. Evolutionary biologists since the time of Darwin have been interested in the diversity of biological life forms — how it arises, what maintains it, and how it evolves. Correspondingly, linguists have always been interested in linguistic diversity in space and time.

In this context, it is interesting to reflect upon Lewontin's (1978) review of sufficient conditions for evolution by natural selection. These are

1. There is variation in ... behavioral traits among members of a species (the principle of variation);

2. The variation is in part heritable ... in particular, offspring resemble their parents (the principle of heredity);

3. Different variants leave different numbers of offspring either in immediate or in remote generations (the principle of differential fitness)

In the case of language evolution, one is interested in linguistic behavior. As we have noted already, in any population there is variation in linguistic behavior. This variation is in part inheritable though the mechanism of inheritance is based on learning — not just from the parents but from a larger group of people.

The case of differential fitness is a tricky point. There is no obvious sense in which speakers of one linguistic variant are reproductively more successful than speakers of another in recent historical times. It is also not clear that natural languages are getting "fitter" in any sense over historical time though this may be a matter of some debate. Fitness may be viewed, however, as the differential transmission of linguistic variants and the role of communicative efficiency, principles of "least effort" (Zipf, 1948) in production and perception of speech, and the differential ease of learnability of different features of a language will all need to be sorted out.

When one considers evolutionary scenarios in prehistoric times and questions like the origin of language in modern humans or the evolution of different kinds of communication systems in the animal world in general, then it is quite possible that communicative ability plays a role in survival or mate selection and therefore has direct bearing on reproductive success.

1.5.1 Scientific History

A little bit of historical perspective on these parallels and differences between language and biology is helpful. Since the discovery of the relatedness of the members of the Indo European family of languages by William Jones

(see *Collected Works*, Volume III) 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.

> The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously parallel.* But we can trace the formation of many words further back than that of species, for we can perceive how they actually arose from the imitation of various sounds. We find in distinct languages striking homologies due to community of descent, and analogies due to a similar process of formation. The manner in which certain letters or sounds change when others change is very like correlated growth. We have in both cases the re-duplication of parts, the effects of long-continued use, and so forth. The frequent presence of rudiments, both in languages and in species, is still more remarkable.
>
> ... Languages, like organic beings, can be classed in groups under groups; and they can be classed either naturally according to descent, or artificially by other characters.
>
> ... The survival or preservation of certain favoured words in the struggle for existence is natural selection.
>
> (*Descent of Man*, C. Darwin, 1871)

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 and Greek (languages 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.

At various points, since the promulgation of evolutionary theory by Darwin and Mendel, linguists have taken up the analogy between biological evolution and language change. For example, the German scholar August

Schleicher did much work on Indo European linguistic tree reconstruction and was influenced both by Linnaeus' taxonomy and Darwin's ideas. In 1863, he published a manuscript entitled *The Darwinian Theory and the Science of Language.* Similarly, the Norwegian linguist Otto Jesperson was also influenced by the Darwinian approach and advanced the view that there was an evolutionary scale and languages were on the whole getting "fitter" and more efficient with the passage of time. In recent times, with the rise of generative linguistics that has placed language as a distinct cognitive and therefore biological trait, these analogies have become more precise. For example, Lightfoot (1998), Kroch (1999), McMahon (2000), Piatelli-Palmarini (1989),Pinker (1994), Aitchison (2000), Wang (1991), Mufwene (2001), Croft (2000), Jenkins (2001), among others have elaborated on this connection.

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. However, historical and evolutionary points of view are vastly more vigorous in biology today than they are in linguistics. It is time, perhaps to change that.

Clearly, biological systems are extremely complex ones and biological form is arguably even harder to characterize quantitatively than linguistic form. The forces shaping biological evolution are by no means simpler than those shaping linguistic evolution. Yet biologists have recognized the utility of computational and mathematical thinking to reason through the morass of possibilities and seeming tautologies to understand important trends. For more than seventy years now, evolutionary biology has been steadily mathematized with a wide range of models ranging from probability theory to game theory. For a random sampling of this aspect of the field, see Fisher, 1930; Seawall Wright, 1968; Crow and Kimura, 1970; Maynard Smith, 1982; Hofbauer and Sigmund, 1988.

Given this, and given that many aspects of linguistic inquiry were greatly mathematized by the Chomskyan revolution in the 1950s, it is somewhat surprising that the study of language evolution has avoided mathematical

analysis until recently. Over the last decade, however, a significant body of work has begun to emerge on computational approaches to the problem opening up the way to such modes of inquiry into historical linguistics and language evolution. (References are too numerous to mention. A partial list includes Yang, 2002; Niyogi and Berwick, 1997; Steels, 2001; Clark and Roberts, 1993; Freedman and Wang, 1996, Shen, 1997; Briscoe, 2000; Batali, 1998; Kirby, 1999; Hurford, 1989, Hurford and Kirby, 2001; Nowak and Krakauer, 1999; Nowak et al, 2001; Cucker, Smale, and Zhou, 2004; Abrams and Strogatz, 2003; Wang, Ke, and Minett 2004; Cangelosi and Parisi, 2002; Christiansen and Kirby, 2003. Since 1996, an International Conference on Language Evolution has been held every two years. See also the website http://www.isrl.uiuc.edu/amag/langev/ for more references.)

Finally, no account will be complete without noting that there has also been a distinct tradition in the study of cultural evolution that has many potential points of contact with language evolution. Pioneering mathematical accounts of cultural evolution have been proposed by Cavalli-Sforza and Feldman (1981), Boyd and Richerson (1985), and Axelrod (1984). Of these, the first mentioned work has greatest overlap with the approach in this book. Chapter 9 has been devoted to similarities and differences between the two approaches. More recently, at a very different level of analysis, an empirical study of the similarities between genes, people, and languages is reported in Cavalli-Sforza (2001). Ruhlen (1994,1997) in a number of scholarly works following in the tradition of Greenberg (1966,1974,1978) constructs phylogenetic trees using techniques from classical comparative linguistics and influenced by perspectives from biological and cultural evolution.

1.6 Summary of Results

As we have noted above, there are many similarities and differences between evolutionary processes in linguistics and biology. Rather than dwell too much on analogies, we will develop the logic of language evolution on its own terms over the course of this book. Let us reiterate our point of view again.

1. Linguistic behavior and underlying linguistic knowledge may be characterised as a formal system. Let $\mathcal{H}$ represent the range of such formal systems that humans possibly possess.

2. Variation within any population (at time t) may be characterized by a probability distribution P_t on $\mathcal{H}$. For any $h \in \mathcal{H}$, $P_t(h)$ represents the

proportion of individuals using the system h.

3. An individual child born within this population will acquire language based on a learning algorithm $\mathcal{A}$ that maps its primary linguistic data onto linguistic systems (elements of $\mathcal{H}$). The distribution of data that the typical child receives will depend upon the distribution of linguistic types in the previous generation P_t and the mode of interaction between the child and its environment.

(1), (2), and (3) taken together will allow us to deduce a map $P_t \to P_{t+1}$ that characterizes how linguistic variation evolves over time.

The interplay between learning by individuals and change (evolution) in populations is subtle. We do not currently have good intuitions about the precise nature of this relationship and the possible forms it could take. Progress on this question is key to developing good theories of how and why languages change and evolve. It is extremely difficult to make progress by verbal arguments alone and therefore it makes sense to study this question in some formal abstraction. So although we try to engage linguistic facts throughout this book, much of our discussion remains abstract.

Another aspect of the book is its focus on mathematical models where the relationship between various objects may be formally (provably) studied. A complementary approach is to consider the larger class of computational models where one resorts to simulations. Mathematical models with their equations and proofs and computational models with their programs and simulations provide different and important windows of insight into the phenomena at hand. In the first, one constructs idealized and simplified models but one can now reason precisely about the behavior of such models and therefore be very sure of one's conclusions. In the second, one constructs more realistic models but because of the complexity, one will need to resort to heuristic arguments and simulations. In summary, for mathematical models the assumptions are more questionable but the conclusions are more reliable — for computational models, the assumptions are more believable but the conclusions more suspect.

1.6.1 Main Insights

Let us summarize the main insights that emerge from the investigations conducted in this book.

Learning and Evolution:

Learning at the individual level and evolution at the population level are related. Furthermore, we see that different learning algorithms have different evolutionary consequences. Thus every theory of language acquisition also makes predictions about the nature of language change. Such theories may therefore be tested not only against developmental psycholinguistic data but also against historical and evolutionary data.

Over the course of this book, we explore many different learning algorithms and their evolutionary consequences. We see that the evolutionary dynamics can depend in subtle ways on whether learning operates with online memoryless algorithms or global batch algorithms. Similarly, there is a difference between symmetric algorithms like the trigger based algorithms and asymmetric algorithms like the cue based algorithms. While both satisfy learnability criteria, they have different evolutionary profiles. In the context of P&P based algorithms, there is some debate as to whether there are default, *marked* states or not during the acquisition process. Such marked states would give rise to asymmetric learning algorithms and their different evolutionary consequences may then be judged against historical data.

We also see the role of critical age periods (the maturation parameter) in learning and evolution. If the learning stops and the mature language crystallizes after a number n of examples have been received, we see that the evolutionary dynamics are characterized by degree n polynomial maps. Although such high degree polynomial maps may have complicated behavior in general, in the particular case of language, they operate in bounded parameter regimes. Thus, though bifurcations typically arise, chaos typically does not.

We also see the differences between learning algorithms that learn from the input provided by a single individual (parent, teacher, or caretaker) versus algorithms that learn from the input provided by the community at large. This point is elaborated shortly.

We have considered some examples of language change to provide linguistic grounding and to give a sense to the reader of how linguistic data may be engaged using the approaches described here. Particular note should be taken of the treatment of phonological changes in Chinese and Portuguese and syntactic change in French and English. An assortment of other changes are scattered across the book.

Finally it is worth noting that much of learning theory in language acquisition is developed in the context of the classical Chomskyan idealization of an "ideal speaker hearer in a homogeneous linguistic environment". As a result one typically assumes that there is a target grammar which the learner

tries to reach. This book drops the homogeneous assumption and analyzes the implications of learning theory in a heterogeneous population with linguistic variation. Learning theory has not been systematically developed in such a context before.

Bifurcations in the History of Language:

A major insight that emerges from the analytic treatment pursued here is the role of bifurcations[11] in population dynamics as an explanatory construct to account for major transitions in language. When one writes down the dynamics one would expect in linguistic populations under a variety of assumptions, again and again, one notices that (a) the dynamics is typically non-linear (b) there are bifurcations (phase transitions) and these may be interpretable in linguistic terms as the change of language from one seemingly stable mode to another. There are numerous examples of such bifurcations in this book.

In Chapter 5, we introduce models with two languages in competition. For a TLA based learner for learning in the P&P model, we see that the equilibrium state depends upon the relationship of a with b where a and b are the frequencies with which ambiguous forms are generated by speakers of each of the two languages in question. If $a = b$, then no evolutionary change is possible. If $a < b$, then one of the two languages is stable, the other is unstable. For $a > b$ the reverse is true. We thus see that it is possible for a language to go from a stable to an unstable state because of a change in the frequencies with which expressions are produced. In a cue based model of learning, also discussed in that chapter, we see that there is a bifurcation from a regime with two stable equilibria to one with only a single stable equilibrium as k (the number of learning samples) and p the cue frequency vary as a function of each other. For models inspired by change in European Portuguese or those inspired by phonological change in Chinese, similar bifurcations arise. In chapters on the emergence of grammatical coherence, we

[11] Bifurcations may be recognized as phase transitions in a number of dynamical models in physics and biology. The most familiar instances of phase transitions are those that lead to changes in the state of materials, e.g., water turning into ice or an iron bar becoming a permanent magnet (the Ising and Potts models). In both cases, statistical physics models may be constructed and temperature is the parameter that is varied in such models. It is seen that a critical threshold in temperature separates the qualitatively different regimes of behavior of the material. Thus temperature may change continuously across this threshold leading to a discontinuous change in the state of the material. For a more precise discussion of this point, see Chapter 13.

see how there is a bifurcation point below which stable solutions include all languages (polymorphism) and above which stable solutions contain a single shared language in the community at large.

These results provide some understanding of how a major transition in the linguistic behavior of a community may come about as a result of a minor drift in usage frequencies provided those usage frequencies cross a *critical threshold*. The usage frequencies may drift from one generation to the next but the underlying linguistic systems may remain stable. But if these usage frequencies cross a threshold, then rapid change may come about. Thus a novel solution to the actuation problem (the problem of what initiates language change) is posited.

Natural Selection and the Emergence of Language:

In part IV of this book, we shed some light on the complex nature of the relationship between communicative efficiency and fitness, social connectivity, learnability, and the emergence of shared linguistic systems. For example, we study the emergence of grammatical coherence, i.e., a shared language of the community in the absence of any centralized agent that enforces such coherence. Two different models are considered in Chapters 12 and 13. In one, children learn from their parents alone. In the other, they learn from the entire community at large. In both models, it is found that coherence emerges only if the learning fidelity is high, i.e., for every possible target grammar g, the learner will learn it with high confidence (with probability $> \gamma$). After examining conditions for learnability, we see that the complexity of the class of possible grammars $\mathcal{H}$, the size of the learning set n, and the confidence γ are all related. For a fixed n, if γ is to be large, then $\mathcal{H}$ must be small. Thus in addition to the traditional learning theoretic considerations, we see that there may be evolutionary constraints on the complexity of $\mathcal{H}$ — the class of universal grammar. In order to stably maintain a shared language in a community, the class of possible languages must be restricted, something like universal grammar must be true.

A second insight emerges from considering the difference in the two models of Chapters 12 and 13. We see that if one learns from parents alone, then natural selection based on communicative fitness is necessary for the emergence of a shared linguistic system. On the other hand, if one learns from the community at large, then natural selection is not necessary. Now in human societies, the social connectivity pattern ensures that each individual child receives linguistic input from multiple people in the community. In such

societies, it is therefore not necessary to postulate mechanisms of natural selection for the emergence of language. On the other hand, in those kinds of animal societies where learning occurs in the "nesting phase" with input primarily from one teacher, one may need to invoke considerations of natural selection. This is the case for some bird song communities, for example.

1.7 Audience and Connections to Other Fields

This book is a computational treatment of the interplay between learning, language, and evolution. The subject matter and ideas reside at the boundary of several disciplines that form the intended audience for this monograph. Our target audience includes the following:

1. The linguist ought to be interested from a variety of perspectives. Those studying language acquisition can now examine the long term evolutionary consequences at the population level of acquisition at the individual level. Those studying historical linguistics will find here a computational framework to investigate and explain the phenomena of their discipline. Finally, those interested in the origin of language and how evolutionary considerations may possibly shape the structure of language, will find some new tools with which to reason about their theories.

2. The computer scientist in domains like computational linguistics and artificial intelligence are introduced here to a new domain of study with its own phenomena that have received little computational attention in the past. While computational linguistics is a fairly old discipline with its roots in mechanical translation, in recent times the focus has largely been on computational models of learning and parsing. Computational work in historical or evolutionary linguistics has been non-existent in the past though a gradual stream of work beginning in the mid-nineties has started to gain momentum. Many subdisciplines of artificial intelligence might find an interest in this work. Those in the areas of artificial life and artificial societies might be interested in the behavior of societies of linguistic agents and their dynamics. While this monograph concentrates on those cases for which analytic understanding is possible, a larger set of phenomena might be realistically pursued within the framework of agent based simulations. This book does not pursue such an approach.

A grand challenge in artificial intelligence is to understand how humans acquire language and to get a computer to do the same. Work in this tradition has ranged from abstract theories of language acquisition discussed earlier (Wexler and Culicover, 1980; Osherson et al, 1986; Blum and Blum, 1975; Gold, 1967; Feldman, 1972; Angluin, 1988; Sakakibara, 1990) to computational work (Berwick, 1985; Feldman et al, 1996; Regier, 1996; Siskind, 1992). An unusual and previously unexploited window into the phenomenon of language acquisition is provided by the facts of language change.

Variation and variability among speakers is possibly the primary source of difficulty for computers to automatically process natural languages. This arises at all levels from speech recognition to language understanding to language translation. Why does this variation arise? What constrains it? What propagates it? These are important questions to resolve and a better understanding of variability is critical to progress in spoken language systems. Here we take a fundamentally historical view towards linguistic variation — one that has almost never been taken in the synchronic view of variation that is implicit in natural language processing.

3. The mathematician interested in dynamical systems will find a variety of concrete iterated function maps that arise in the study of historical linguistics and some of which are of considerable mathematical difficulty. The study of dynamical systems has been fed by problems in physics, biology, economics but never before from linguistics as far as we know.

4. The evolutionary biologist may be interested in a new domain with much of the same character. We have already touched upon the similarities and differences between evolutionary processes in language and biology. Researchers in animal communication, signalling, and ethology may find synergies between their own perspectives and tools and those developed here.

5. Social scientists like anthropologists interested in culture and its transmission, economists interested in bounded rationality and its evolutionary consequences, and evolutionary psychologists interested in evolutionary perspectives on cognitive behavior will find a parallel here in the behavior of linguistic learners as they learn language and evolve over time.

6. This book discusses the relationship between the macroscopic behavior of a linguistic population and the microscopic behavior of the linguistic agents in this population. As a result, a theme that runs across this book is the analysis of the emergent properties of a complex system of several interacting components. This theme arises in different ways in studies of pattern formation in biology, physics, and various social sciences. Researchers interested in the study of complexity and complex phenomena may find new applications in the analysis of language as a complex adaptive system.

1.7.1 Structure of the Book

The rest of the book is organized as follows. In the next several chapters, we introduce the basic dynamical system model for studying language change. This is developed over two parts. The starting point for our narrative is the problem of language acquisition. In Part II of this book (Chapters 2, 3, and 4) we discuss the philosophical problem of inductive inference that lies at the heart of the language learning problem. We introduce frameworks for the analysis of learning algorithms that will play a useful role in later chapters. In Part III, we begin by deriving the dynamics of linguistic populations (Chapters 5 and 6) that forms the core model for much of the rest of the book. We show how models of language change depend upon the learning algorithm, the class of grammars, and sociolinguistic considerations of frequency of language use.

We continue in Part III by applying our model (Chapters 6, 7, 8, 9, and 10) to various special cases. As a result, variations and extensions of the basic model are fleshed out. Many of these chapters have a running discussion of relevant linguistic phenomena that provide the motivation for this entire exercise.

In Part IV we consider the trickier problem of the origin of linguistic communicative systems. We explore in some detail two themes. First, we consider the matter of communicative efficiency and discuss a probabilistic formulation for two linguistic agents communicating in a shared world. Second, we examine the role of fitness based on communicative efficiency in language evolution. We consider evolutionary models with fitness in Chapter 12 where individuals reproduce at differential rates that are proportional to their communicative success. In Chapter 13, we consider language models without fitness but with social learning where individuals learn from everybody. In both cases, we study when and how linguistic coherence emerges

and relate this coherence threshold to the learning fidelity of the individual learner.

In summary, our goal in this book is to understand the nature of the relationship between individual language learning, grammatical families, and population dynamics. In a nutshell, we wish to understand how the distribution of different grammatical types in a population will evolve under a variety of learning algorithms and modes of interaction. Chapters 2 through 13 are an exploration of this theme in some mathematical detail. Computational and mathematical modeling forces us to be precise in our reasoning as we proceed.

In Part V, we conclude. We take stock of the situation, outline our essential results, and suggest directions for future work.