LINGUISTIC EVOLUTION THROUGH LANGUAGE ACQUISITION

EDITED BY

TED BRISCOE

University of Cambridge

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbounre, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Baskerville 11/12.5 pt *System* LaTeX 2ε [TB]

A catalogue record for this book is available from the British Library

ISBN 0 521 66299 0 hardback

Contents

1

Introduction

Ted Briscoe
Natural Language and Information
Processing Group, Computer Laboratory,
University of Cambridge

1.1 Linguistic theory and evolutionary theory

Taking an evolutionary perspective on the origins and development of human language, and on linguistic variation and change, is becoming more and more common, as the papers in Hurford *et al.*(1998) attest. The term 'evolution' now crops up regularly in work emerging from the broadly generative tradition in linguistic theory (e.g. Jackendoff, 1997; Steedman, 2000). The latter development is probably a more or less direct consequence of several influential attempts to reconcile Chomskyan nativism with evolutionary theory, primarily in terms of a gradualist and adaptionist account of the origins and development of the language faculty (e.g. Hurford, 1989; Newmeyer, 1991; Pinker and Bloom, 1990). However, most of the contributions to this book owe more to the complementary but very different insight (e.g. Hurford, 1987, 1999) that not only the language faculty *per se*, but also the origins and subsequent development of *languages themselves* can be fruitfully addressed within the framework of evolutionary theory. Under this view, languages are evolving, not metaphorically but literally, via cultural rather than biological transmission on a historical rather than genetic timescale. This represents a very distinct and quite narrow theme within the broader program of integrating linguistic theory and evolutionary theory, and it is this theme which is primarily addressed by the contributors to this volume.

Evolutionary ideas have had a rather checkered history within linguistic theory despite their close mutual influence in the nineteenth century. McMahon (1994:ch12) provides a brief account of this history and also discusses linguistic work influenced by evolutionary theory during the fifties and sixties. However, the insight that languages *per se* can be studied as (culturally) evolving systems, post the modern synthesis in biology and post mathematical and computational work on dynamical systems, does not seem to have reemerged until the eighties when Lindblom (1986) in phonology, Keller (1984, (1994)) in historical linguistics, and Hurford (1987) in syntactic theory independently articulated this view (using somewhat different terminology). The idea is an instance of the 'universal Darwinist' claim (Dawkins, 1983; Dennett, 1995:343f) that the methodology of evolutionary theory is applicable whenever any dynamical system exhibits (random) variation, selection amongst variants, and thus differential inheritance. In the nineties, this perspective on languages has been espoused enthusiastically and persuasively by non-linguists (e.g. Cziko, 1995; Deacon, 1997). However, it has not had significant impact in mainstream linguistic theory as yet, perhaps partly because work has only recently begun to address questions seen as central to (generative) linguistic theory.

The contributions to this volume are less concerned with questions of linguistic origins or the development of a broad evolutionary account of human language, than with why and how specific syntactic universals evolved (Kirby, Batali, Briscoe), why homonymy and synonymy are present and maintained in vocabulary systems (Steels and Kaplan), the nature of (E-) language syntactic change (Niyogi, Briscoe), the kind of language learning mechanism required to not only acquire an existing linguistic system accurately but also impose further structure on an emerging system (Oliphant, Kirby, Worden), and the (co)evolution of language(s) and this learning mechanism (Turkel, Briscoe). A second and equally important way in which the contributions here represent a tightly circumscribed theme within evolutionary linguistic work is that all utilize a methodology of computational implementation and simulation of (more or less explicit) formal models. For this reason too, there is a close connection with formal generative linguistic theory. Mathematical modeling and/or computational simulation help ensure that theories constructed are complete and precise, and also help with their evaluation by making the assumptions on which they rest fully explicit. This is particularly critical in the context of largely speculative accounts of the prehistoric development of human languages, as without such a methodology there is little to constrain such speculation.

The rest of this introduction describes the key ideas and techniques which underlie the contributions to this book, and, more broadly, the evolutionary approach to linguistic variation, change and development, relating them to current linguistic theory and discussing critical methodological issues. The contribution by Hurford contains a thorough and insightful analysis and comparison of five different computational models of linguistic evolution, two of which are described here (Batali, Kirby), as well as developing a more general framework for such comparisons that could, in principle, be applied to all the work presented here. Therefore, I limit myself here to additional, I hope complementary, remarks and refer the reader to Hurford's contribution for a much more detailed exposition of the general structure of many of the models.

1.2 The formal framework

1.2.1 Generative linguistics

Chomsky (1965) defined grammatical competence in terms of the language of (i.e. stringset generated by) an ideal speaker-hearer at a single instant in time, abstracting away from working memory limitations, errors of performance, and so forth. The generative research program has been very successful, but, one legacy of the idealization to a single speaker at a single instant has been the relative sidelining of language variation, change and development. More recently, Chomsky (1986) has argued that generative linguistics can offer a precise characterization of I-language, the internalized language or grammar of an individual speaker, but has little to say about E-language, 'external' language, which is an epiphenomenon of the I-languages of the individual speakers who comprise a speech community.

Consequently, the study of language change within the generative tradition has largely focused on 'I-language change'; that is, the differences between I-languages or their corresponding grammars internalized by child language learners across generations. And within I-language change on the (parametric) properties of internalized grammars (e.g. Lightfoot, 1979, 1999). The generative approach to language change treats (major) grammatical change as a consequence of children acquiring different grammars from those predominant amongst the adults in the population, perhaps as a consequence of variation in the internalized grammars of these adults. However, theories of language variation, change and development will (minimally) require an account of how the E-language(s) of an adult population can be defined in terms of the aggregate output of these (changing) individuals.

1.2.2 Language agents

A language agent is a idealized model of just what is essential to understanding an individual's linguistic behavior. I use the term 'agent', in common with several contributors to this volume and with (one) current usage in computer science and artificial intelligence, to emphasize that agents are artificial, autonomous, rational and volitional, and that agents are embedded in a decentralized, distributed system, i.e. a speech community.

A language agent must minimally be able to learn, produce and interpret a language, usually defined as a well-formed set of strings with an associated representation of meaning, by acquiring and using linguistic knowledge according to precisely specified procedures. Beyond this, the models of language agents deployed by the contributors differ substantially, depending on theoretical orientation and the precise questions being addressed. Oliphant, and Steels and Kaplan define linguistic knowledge entirely in terms of word–meaning associations in a lexicon, reflecting their focus on the acquisition of vocabulary. Niyogi, Turkel and Briscoe focus on the acquisition of parametrically-defined generative grammars and thus define linguistic knowledge primarily in terms of (sets of) parameter settings. Batali, Kirby and Worden all develop broadly lexicalist models of linguistic knowledge, in which the acquisition of lexical and grammatical knowledge is closely integrated.

All the models provide some account of the acquisition, comprehension and production of (I-) language. Again the details vary considerably depending on the theoretical orientation and questions being addressed. For example Niyogi and Turkel largely assume very idealized, simple accounts of parameter setting in order to focus on the dynamics of E-language change and the genetic assimilation of grammatical information, respectively. The other contributors concentrate on specifying acquisition procedures in some detail, since properties of the acquisition procedure are at the heart of linguistic inheritance and selection. As acquisition is closely bound up with comprehension, most of these contributors also develop detailed accounts of aspects of the comprehension, or at least parsing, of linguistic input. However, none really provide a detailed account of language production, beyond the minimal assumption that linguistic utterances are generated randomly from a usually uniform distribution over the strings licensed by an agent's grammar and/or lexicon.

Additionally, language agents can have further properties, such as the ability to invent elements of language, the ability to reproduce further language agents, an age determining the learning period and/or their 'death', and so forth. For example, the contributors on the development of language or emergence of new traits, often endow their language agents with the ability to 'invent' language in the form of new utterance–meaning pairs, where the utterance can either be essentially an unanalysed atom ('word') or a string with grammatical structure ('sentence'). Invention is again modeled very minimally as a (rare) random process within a predefined space of possibilities, and is one method of providing the variation essential to an evolutionary model of linguistic change and/or development.

1.2.3 Languages as dynamical systems

E-languages are the aggregate output of a population of language users. Such a population constitutes a speech community if the internalized grammars of the users are 'close' enough to support mutual comprehension most of the time. Membership of the population/speech community changes over time as people are born, die or migrate.

Perhaps the simplest model which approximates this scenario is one in which the population initially consists of a fixed number of 'adult' language agents with predefined internalized grammars, and their output constitutes the data from which the next generation of 'child' language learning agents acquires new internalized grammars. Once the learning agents have acquired grammars, this new generation replaces the previous one and becomes the adult generation defining the input for the next generation of learners, and so on. We can define a dynamical model of this form quite straightforwardly. A dynamical system is just a system which changes over time. We represent it by a sequence of states where each state encodes the system properties at each time step and an update rule defines how state s^{t+1} can be derived from state s^t:

$$s^{t+1} = \mathit{Update}(s^t)$$

Time steps in this model correspond to successive non-overlapping generations in the population. Minimally, states must represent the E-language(s) of the current generation of language agents, defining the input for the next generation of learners. The *Update* rule must specify how the internalized grammars of the learners are derived from the E-language input.

Niyogi and Berwick (1997) develop a deterministic version of this model in which each state is defined by a probability distribution over triggers, a finite subset of unembedded sentences from each language defined by each internalized grammar present in the population. The deterministic update rule defines a new probability distribution on triggers by calculating the proportions of the population which will acquire the internalized grammars exemplified in the input data. In this volume, Niyogi describes this model in detail and develops it by exploring the predictions of deterministic update rules which assume that different learners will receive different input depending on their parents or on their geographical location. Niyogi shows how this model makes predictions about the direction and timecourse of E-language change dependent on the learning algorithm and the precise form of the update rule. Throughout, E-language change is modeled as a consequence of a number of 'instantaneous' I-language changes across generations, in common with standard generative assumptions about major grammatical change. However, the population-level modeling demonstrates that the consequent predictions about the trajectory and direction of change are often surprising, very varied, and always sufficiently complex that mathematical modeling and/or computational simulation are essential tools in deriving them.

Niyogi's use of deterministic update rules assumes that random individual differences in the learners' input are an insignificant factor in language change. In his model, learners are exposed to a finite number of triggers randomly drawn according to a probability distribution defined by the current adult population. Sampling variation may well mean that learners will or will not see triggers exemplifying particular internalized grammars present in the adult population. If the number of triggers sampled and/or the size of the population is large, then this variation is likely to be insignificant in defining the overall trajectory and timecourse of E-language change. Therefore, Niyogi models the behavior of an *average* learner in the population. In the limit, the behavior of the overall model will be identical to one in which the behavior of individuals is modeled directly but the population is infinite. The great advantage of this approach is that it is possible to analytically derive fixed points of the resulting dynamical models, and thus prove that certain qualitative results are guaranteed given the model assumptions.

The models utilized by the other contributors are all stochastic in the sense that they model the behavior of individual agents directly and deploy stochastic or random agent interactions. Therefore, there may be sampling variation in learner input. Time steps of the resulting dynamical models are defined in a more fine-grained way in terms of individual agent interactions or sets of such interactions. For example, Batali, Kirby, Oliphant, and Steels and Kaplan all take individual linguistic interactions as the basic time step, so the update rule in their simulations is defined (implicitly) in terms of the effect on E-language of any change in the linguistic knowledge of two interacting agents. In these and most of the other models, language acquisition is no longer viewed as an 'instantaneous' event. Rather agents interact according to their (partial) knowledge of the E-language(s) exemplified in the environment and continue to update this knowledge for some subset of the total interactions allotted to them. Turkel uses a standard (stochastic) genetic algorithm architecture with fitness-based generational replacement of agents so that time steps in his system correspond to non-overlapping generations. However, the fitness of each agent is computed individually based on 10 learning trials between it and another randomly chosen agent in the current population. Briscoe defines time steps in terms of interaction cycles consisting of a set number of interactions proportional to the current population size. Agents interact randomly and a proportion of interactions will involve learners. Once a stochastic model of this type is adopted it is also easy to introduce overlapping generations in which learners as well as adults may contribute utterances to E-language. The stochastic approach provides greater flexibility and potential realism but relies even more heavily on computational simulation, as analytic mathematical techniques are only easily applicable to the simplest such systems. For this reason, it is important that the results of simulation runs are shown to be statistically reliable and that the stochastic factors in the simulation are not dominating its behavior.

Interestingly, though Kirby derives his results via a stochastic simulation of a single speaker providing finite input to a single learner, the critical time steps of his model are generation changes, in which the learner becomes the new adult speaker, and a new learner is introduced. Therefore, it would appear that the analytic model developed by Niyogi and Berwick could, in principle, be applied to Kirby's simulation. The effect of such an application would be to factor out sampling variation in learner input. It should then be possible to prove that the qualitative results observed are guaranteed in any run of such a simulation. Indeed, what we might expect is that, over the predefined meaning space, a single *optimal* grammar, relative to the subsumption based grammar compression algorithm employed, is the sole fixed point of the dynamical system.

1.2.4 Languages as adaptive systems

Niyogi and Berwick (1997) argue that their model of E-language does not need or utilize a notion of linguistic selection between linguistic variants. However, the specific learning algorithm they utilize is selective, in the sense that it is parametric. They examine, in detail, the predictions made by the Trigger Learning Algorithm (TLA, Gibson and Wexler, 1994) embedded in their dynamical model. The TLA is a parameter setting algorithm based on the principles and parameters account of grammatical acquisition (Chomsky, 1981). The TLA selects one grammar from the finite space of possible grammars defined by the settings of a finite number of finite valued parameters. Thus, when faced with variation exemplifying conflicting parameter settings in the input, the TLA selects between the variants by assigning all parameters a unique value. So, selection between variants is a direct consequence of the learning procedure.

It is possible to imagine a learning procedure which when faced with variation simply incorporated all variants into the grammatical system acquired. Briscoe (2000a) describes one such algorithm in some detail. In order to claim that no selection between linguistic variants is happening in dynamical models of the type introduced in the previous section, we would need to demonstrate that the specific learning procedure being deployed by agents in the system was not itself selective in this sense. However, such a learning procedure seems implausible as a model of human language learning because it predicts that the dynamic of language change would always involve integration of variation and construction of larger and larger ‘covering’ grammars of learner input. Loss of constructions, competition between variants, and the very existence of different grammatical systems would all be problematic under such an account.

Once we adopt an account of language learning which is at least partially selective, then it is more accurate to characterize linguistic dynamical systems as *adaptive* systems; that is, as dynamical systems which have evolved in response to environmental pressure. In this case, to be learnable with respect to the learning algorithm deployed by child language learners (whatever this is). The nature of the pressure depends on properties of the learning procedure and need not be 'functional' in the conventional linguistic sense. For example, the TLA selects between variants by either selecting the parameter setting dictated by the last unambiguous trigger (with respect to the relevant parameter) in the input before the end of the learning period or by making an unbiased random guess. Therefore, the relative frequency with which variants are exemplified in learner input is the main determinant of which variants are culturally transmitted through successive generations of language learning agents. However, most of the learning procedures developed by other contributors exhibit various kinds of inductive bias which interact with the relative frequency of variant input to create additional pressures on learnability.

It is striking that with the exception of Turkel's quite idealized account of learning (which is not intended as a serious model of parameter setting), the other contributors all develop learning algorithms which, unlike the TLA, incorporate Ockham's Razor in some form; that is, a broad preference for the *smallest* grammar and/or lexicon ('compatible' with the input). In addition, most of the models remain selective, in the sense defined above with respect to the TLA, in that they bias learning towards acquisition of *unambiguous* word-meaning associations and/or syntactic means of realizing non-atomic meaning representations. Indeed the latter bias is a direct consequence of the former, as alternative encodings of the mapping from meaning to form result in larger descriptions. All the models impose hard constraints in the form of representational assumptions about the kind of grammars and/or lexicons which can be acquired; that is, assumptions about the form of universal grammar. It is in terms of such representational assumptions which incorporate hard inviolable constraints on what can be learnt that the soft, violable constraints or inductive bias in favour of small unambiguous mappings can be stated. As these representational assumptions vary a good deal between the contributions, the precise effect of the bias will also vary. Nevertheless, very broadly, Ockham's Razor creates an additional selection pressure for regularity in linguistic systems, over and above the requirement for frequent enough exemplification in learner input.

One might argue that the incorporation of such inductive biases into these models is no more than a method of ensuring that the simulations deliver the desired results. However, Ockham's Razor has been a central tenet of learning theory for centuries, and in the theory of informational complexity has been formally proved to provide a universally accurate prior or inductive bias over a universal representation language (Rissanen, 1989). In the framework of Bayesian learning, the minimum description length principle, over a given representation language or class of grammars/models, provides a concrete, practical instantiation of Ockham's Razor, which has been used to develop learnability proofs for non-finite classes of grammar (e.g. Muggleton, 1996) and to develop theoretical and computational models of lexical and grammatical acquisition (e.g. Brent and Cartwright, 1996; de Marcken, 1996; Rissanen and Ristad, 1994; Osborne and Briscoe, 1997). Therefore, the learning procedures developed here, which incorporate this principle in some form, are not in any way unusual, controversial or surprising. Indeed, inductive bias has been argued to be essential to successful learning (Mitchell, 1990, 1997), this insight is central to the Bayesian framework, and within the space of possible inductive biases, Ockham's Razor remains the single most powerful and general principle, which under the idealized conditions of a universal representation language has been shown to subsume all other forms of bias (e.g. Rissanen, 1989).

Kirby (this volume, 1998, 2000) extends this insight in several ways arguing that the bias for smaller grammars is tantamount to the assumption that learners generalize from data and will, therefore, be a component of any language learning procedure. He argues that the syntactic systems which emerge in his simulations would emerge given many other possible learning procedures. Oliphant, in the context of word learning, similarly argues that the only kind of learning procedure which will *impose* order on random, inconsistent vocabulary systems is one which prefers unambiguous word-meaning mappings. However, as we have seen above, at root this follows from Ockham's Razor, since this is equivalent to saying that a learner prefers to retain the smallest number of word–meaning associations.

The picture which emerges then, is that languages have adapted to the human language learning procedure, in the sense that this procedure incorporates inductive bias – itself virtually definitional of the concept of learning. Inductive bias creates linguistic selection for more learnable linguistic variants relative to this bias and thus as languages are culturally transmitted from generation to generation via successive child language learners, linguistic systems will evolve that fit, or are adapted to, these biases. However, this picture cannot be the whole truth, for if it were we would predict that all languages should eventually converge to a single optimal system, that change should always be unidirectional, and that variation should decrease and eventually disappear, at least with respect to these biases. However, this is not a realistic picture, variation is maintained and increases in some social contexts (e.g. Nettle, 1999), and unidirectional change in the form of 'grammaticalization' is at best a tendency (e.g. Newmeyer, 1998).

1.2.5 Languages as complex adaptive systems

Evolution is *not* a process of steady improvement along a single trajectory leading to a single optimal solution. Sewall Wright (1931) introduced into evolutionary theory the idea of adaptive or fitness landscapes with multiple local optima or peaks, and this idea has been considerably refined since (e.g. Kauffman, 1993:33f). The modern picture of (co)evolution is of a process of local search or hill climbing towards a local optimum or peak in a fitness landscape which itself inevitably changes. Conflicting selection pressures will cause the fitness landscape to contain many locally optimal solutions, and thus the evolutionary pathways will be more complex and the space of near optimal solutions more varied (Kauffman, 1993:44f). A simple and well-attested example of conflicting selection pressures from biology is the case of 'runaway' sexual selection for a non-functional marker such as the peacock's tail, counterbalanced by natural selection for efficient movement (e.g. Dawkins, 1989:158f). Adaptive systems which change on the basis of interactions between conflicting selection pressures in unpredictable ways, involving positive or negative feedback, with no centralized control are increasingly termed *complex* adaptive systems (e.g Kauffman, 1993).

The idea that there are competing motivations or conflicting pressures deriving from the exigencies of production, comprehension and acquisition has been developed by linguists working from many different perspectives (e.g. Langacker, 1977; Fodor, 1981; Croft, 1990:192f). However, in linguistics little progress has been made in quantifying these pressures or exploring their interaction (Newmeyer, 1998). Computational simulation and mathematical analysis of E-languages, modeled as dynamical systems adapting to such conflicting pressures, provides a powerful new methodology for deriving precise, quantitative and qualitative predictions from the interaction of such conflicting pressures. For example, one perhaps better understood pressure on the evolution of grammatical systems derives from parsability (e.g. Gibson, 1998; Hawkins, 1994; Miller and Chomsky, 1963; Rambow and Joshi, 1994).

A number of metrics of the relative parsability of different constructions have been proposed, both as accounts of the relative psychological complexity of sentence processing and of the relative prevalence of different construction types in attested languages. A metric of this type can be incorporated into an evolutionary linguistic model in a number of ways. Kirby (1999) argues, for example, that parsability equates to learnability, as input must be parsed before it can be used by a learner to acquire a grammar. By contrast, Hawkins (1994:83f) argues that parsability may influence production so that more parsable variants will be used more frequently than less parsable ones (within the space of possibilities defined by a given grammar), and presents evidence concerning the relative frequency of constructions from several languages in support of this position. This would entail that less parsable constructions would be less frequent in learner input, in any case. Briscoe (2000b) reports experiments, using the same simulation model described in this volume, which show that either approach alone or in tandem can, in principle, account for adaptation towards more parsable typological variants.

It is also likely that production pressures, for example for economy of expression, also play a significant role. In general, these have not been quantified to the same extent, at least in work on syntax. However, there are already some interesting computational models. Kirby (2000), for example, extends the simulation and model described in this volume to include a speaker bias towards minimal encoding of meaning representations. Once this is done the grammars in the simulations no longer evolve so inexorably towards optimally regular encoding of the meaning–form mapping, but unstable irregular and less compositional, but nevertheless short mappings repeatedly emerge. If the further assumption is made, that meanings are expressed according to a highly-skewed 'Zipfian' distribution, then irregular, minimal encodings of very frequent meanings emerge and persist stably across generations.

Once we recognise that there are conflicting selection pressures, it is easier to see why language change does not move inexorably (and unidirectionally) towards a unique global optimum. No such optimum may exist, and in any case, change will alway be relative to and local with respect to the current 'position' in the current adaptive landscape. For instance, a canonical SOV grammar might evolve increasingly frequent extraposition because SOV clauses with long or 'heavy' object phrases are relatively unparsable (e.g. Hawkins, 1994:196f). However, SVO grammars will be less likely to do so since long object phrases will mostly occur postverbally anyway and will not create analogous parsing problems. Once such a change has spread, it may in turn create further parsability (or expressiveness or learnability) issues, altering the adaptive landscape; for example, by creating greater structural ambiguity, resulting perhaps in evolution of obligatory extraposition. (It is this locality or blindness in the search for good solutions that makes the evolutionary process more like tinkering than engineering.) In the framework advocated here, we can recognize that such historical pathways can be stereotypical responses to similar pressures arising in unrelated languages, in much the same way that eyes and wings have evolved independently in different lineages many times, without the need to posit a substantive theory of such changes or to see them as deterministic.

1.2.6 Genetic assimilation

So far we have implicitly assumed that the learning procedure and wider language faculty is universal and invariant across the human species. Most of the contributors to this volume focus exclusively on the effects of a universally shared and preadapted (language) learning procedure on the evolution of language itself. Nevertheless, without the assumption of a shared and effective learning procedure across all agents in the population, it would not be possible to demonstrate the emergence and development of consistent and coherent communication systems. For example, Sharpe (1997) demonstrates that vocabulary systems of the type investigated by Oliphant and Steels and Kaplan only emerge under the assumption that all the agents are deploying the same learning algorithm incorporating the same or very similar inductive biases.

The evolution by natural selection of the human (language) learning procedure, and of other elements of the language faculty such as the human parsing and generation mechanisms, has been addressed in a number of recent papers (Pinker and Bloom, 1990; Newmeyer, 1991), and genetic assimilation (e.g. Waddington, 1942), or the so-called Baldwin Effect (Baldwin, 1896), in which changes in a species' behavior (the advent of language) create new selection pressures (the need to learn language efficiently) has been proposed as a plausible evolutionary mechanism through which a language faculty could have gradually evolved. However, this view is certainly controversial; others have proposed saltationist or single step scenarios (e.g. Bickerton, 1998) or argued that preadapted general-purpose learning mechanisms suffice to account for language emergence and subsequent acquisition (e.g. Steels, 1998; Deacon, 1997; Worden, this volume).

The evolutionary perspective on language development and change described above, and the commitment to develop an evolutionarily plausible account of the emergence and subsequent evolution of any putative language faculty, certainly provide new ways of addressing this central issue in Chomskyan linguistic theory. Firstly, under either a gradualist or saltationist account, the presence of (proto)language(s) in the environment is an essential assumption to provide the necessary selection pressure to ensure that a newly emerged faculty persists; if the ability to learn language reliably does not enhance fitness then there would be no selection pressure to maintain such a faculty, and fitness can only be enhanced by it if there is an existing communicative system (e.g. Kirby, 1998). Secondly, if (proto)language precedes the language faculty, then (proto)language must be learnable via general-purpose learning mechanisms. Thirdly, as the historical evolution of languages will be orders of magnitude faster than the genetic evolution of such a faculty, it is quite plausible that languages simply evolved to fit these general-purpose learning mechanisms before these mechanisms themselves had time to adapt to language. As Deacon (1997:109) memorably puts it: "Languages have had to adapt to children's spontaneous assumptions about communication, learning, social interaction, and even symbolic reference, because children are the only game in town ... languages need children more than children need languages." On the other hand, if the language faculty has evolved significantly subsequent to its emergence, then it is of little consequence whether it emerged gradually or by saltation. As Ridley (1990) points out, evolutionary theory tells us more about the maintenance and refinement of traits than their emergence, and the selection pressures subsequent to emergence would be the same given either a saltationist or gradualist account. Fourthly, Pinker and Bloom (1990) and others assume that linguistic universals provide evidence for a language faculty, but if languages evolve to adapt to the inductive bias in the human learning procedure, then linguistic universals need not be genetically-encoded constraints, but instead may just be a consequence of convergent evolution towards more learnable grammatical systems. Again to quote Deacon (1997:116) "universal[s]... emerged spontaneously and independently in each evolving language, in response to universal biases in the selection processes affecting language transmission. They are *convergent* features of language evolution in the same way that the dorsal fins of sharks, ichthyosaurs, and dolphins are independent convergent adaptations of aquatic species."

Worden develops this evolutionary argument against the language faculty, describing a unification-based model of language processing and acquisition and suggesting that a general Bayesian learning algorithm can be used to learn lexical entries in such a model. But the degree to which this is an argument against the existence of or need for a language faculty depends on exactly how domain-independent the unification-based representation language in which linguistic knowledge is couched. Though the representation language is partly encoding conceptual information it is also encoding facts about morphosyntactic realization of meaning (i.e. grammar). Within the context of this more detailed model, Worden is able to make the argument about differential selection pressures on languages and the language faculty and the relative speed of evolution more precise, and tentatively concludes that there would be little pressure for natural as opposed to linguistic selection in line with Deacon's (1997) position.

Turkel, by contrast, simulates evolution of a principles and parameters model of the language faculty and argues that the emergence of a community of speakers endowed with such a faculty, without invoking genetic assimilation, is implausible. The probability of compatible language faculties emerging *de nihilo* in two or more individuals via natural selection is astronomically low. Yet for such a trait to be maintained it must be shared by members of a speech community in order to confer any benefit in fitness. Genetic assimilation provides a mechanism by which a (proto)language using population can gradually converge on a shared language faculty, because individuals able to learn the existing (proto)language slightly more effectively will be selected for over successive generations. Briscoe takes a similar position, presenting a model which integrates Bayesian learning with a principles and parameters account of language acquisition, and arguing that this faculty would be refined by genetic assimilation even in the face of very rapid language change (or 'coevolution').

1.3 Methodological issues

The use of computational simulation and/or mathematical modeling to derive predictions from dynamical models is a vital tool for the exploration of evolutionary accounts of language variation, change and development, and of the development of the language faculty. The behavior of even simple dynamical systems is notoriously complex and often unintuitive, therefore models or theories based entirely on verbal reasoning run a serious danger of being incomplete or not making the predictions assumed. Simulation and modeling force the theorist to be precise enough to specify a complete model and to look at its actual rather than assumed predictions. This places powerful constraints on the development of evolutionary models since it often becomes clear in the process of creating them that some of the assumptions required to make the model 'work' are unrealistic, or that apparently realistic assumptions simply do not yield plausible predictions. However, the mere existence of a 'successful' simulation or mathematical model does not guarantee either the correctness of the assumptions leading to its predictions or of the evolutionary pathway to these predictions.

Evolution is an irreducibly historical process which can be, and often is, affected by accidents, such as population extinctions or bottlenecks, which are beyond the purvue of any rational reconstruction (i.e. model) of an evolutionary process. Since the precise prehistoric pathways that were followed during the emergence and initial development of human language are unknowable, this places a fundamental limit on what we can learn from simulations which (exclusively) address such questions. At the very best such argumentation is irreducibly probabilistic. On the other hand, work in the same framework which addresses historically attested language changes and associated demographic upheavals, such as those occurring during language genesis, is less susceptible to this problem.

A simulation or mathematical model rests on a set of hopefully explicit assumptions just as an argument rests on premises. Often it is not possible to reliably assess the truth of these assumptions or their causal relevance in a prehistoric setting. Thus the predictions made by the model are only as strong as the assumptions behind it. The advantage of models is that *all and only* the critical assumptions required to derive a specific conclusion should be manifest if a good methodology is adopted. The use of computational simulation greatly facilitates the testing of many parameterized variants of a model to explore exactly what is critical. However, it is also important that the initial model adopted abstracts away from as many contingent specific details as possible in order to achieve greatest generality and to derive results from the weakest set of assumptions possible. Ultimately, this is a matter of judgement and experience in the development of such models – there is no 'logic of discovery' – but without such abstraction even computational simulation and exploration of such models will become rapidly intractable.

One example of both the benefits and limits of the methodology is provided by the issue of genetic assimilation of linguistic constraints into the language faculty (discussed in the previous section). Deacon (1997:322f) argues quite persuasively that language change would have been too rapid to create the constant selection pressure on successive generations of language users required for genetic assimilation. However, one simulation in which both language change and genetic assimilation are modeled demonstrates that genetic assimilation still occurs even when language changes are happening as fast as is compatible with maintenance of a speech community (Briscoe, 2000b). The key implicit assumption in Deacon's argument is that the hypothesis space of possible grammars defining the environment for adaptation is sufficiently small that most grammars will be sampled in the time required for genetic assimilation to go to fixation in a population of language users. The model makes clear that if this hypothesis space is large enough then significant portions of it are unlikely to be sampled during this time, so there is constant pressure to assimilate constraints that rule out or disprefer the unsampled grammars. On the other hand, this demonstration, though it undermines Deacon's specific argument, does not guarantee that genetic assimilation of such constraints into the language faculty did, in fact, occur. The model, in conjunction with related work on genetic assimilation (Mayley, 1996) also makes it clear that one critical assumption is that there is correlation between the neural mechanisms underlying language learning and the genetic specification of these mechanisms which will enable the 'transfer' of such constraints to the genetic level. We simply do not know, given our current understanding of both the genetic code and relevant neural mechanisms, whether or to what degree this is the case.

In addition to these general points, there are more specific methodological issues contingent on the type of model adopted. Deterministic models based on analytic techniques, such as that of Niyogi, are methodologically stronger in the sense that predictions derived from them are guaranteed to hold of any specific experimental realization of such models. However, analytic techniques are hard to apply to all but the simplest models. Computational simulation – that is, the running of specific experiments with a model – can be used as an alternative to mathematical analysis. However, the behavior of such simulation runs needs to be considered carefully before conclusions are drawn. If the simulation is stochastic in any way, as most of those presented in this volume are, then we need to be sure that predictions are reliable in the sense that they represent high probability or typical results of such simulations. One basic technique for achieving this is to examine the results from multiple identically-parameterized runs. However, if the qualitative behavior of the model over multiple runs is not absolutely clearcut, then statistical analysis of results may also be required. The advantage of computational simulation is that more realistic models can be explored rapidly in which, for example, there are no fixed points or deterministic attractors in the underlying dynamical system (i.e. no endpoint to evolution). Nevertheless, as this work grows in sophistication, careful statistical analysis of the underlying models will become increasingly important, as is the norm, for example, in population genetics (e.g. Maynard Smith, 1998).

It is sometimes suggested that simulations are too dangerous: "I have resisted the temptation to utilize computer simulations, mostly for reasons of clarity (in my own head – and perhaps also the reader's). Simulations, if they are to be more than mere animations of an idea, have hard-to-appreciate critical assumptions." (Calvin, 1996:8). Behind such sentiments lurks the feeling that simulations are 'doomed to succeed' because it is always possible to build one in which the desired result is achieved. I hope this introduction and the contributions to this volume will convince the reader that, though simulation without methodological discipline is a dangerous tool, methodologically rigorous simulation is a critical and indispensable one in the development of evolutionary dynamical models of language.

1.4 What next?

Though the contributors to this book approach the question of the role of language acquisition in linguistic evolution from a wide variety of theoretical perspectives and develop superficially very different models, there is a deep underlying unity to them all in the realization of the centrality of acquisition to insightful accounts of language emergence, development, variation and change. I hope the reader will recognize this unity and agree with me that this work makes a powerful case for the evolutionary perspective on language. Nevertheless, it should also be clear that much work remains to be done. Methodologically, we have a long way to go in assimilating and evaluating techniques from fields such as population genetics, in which a powerful set of mathematical techniques for studying dynamical systems has been developed. Substantively, we have only begun to scratch the surface of critical issues, such as that of conflicting selection pressures or competing motivations in linguistic evolution, which will take us well beyond the realm of simple models/simulations with fixed points to ones with very complex and dynamic adaptive landscapes. Despite this, I hope the reader will also agree with me that the study of E-languages as complex adaptive systems is a potentially very productive research programme which can be tackled in a methodologically sound way.

References

Baldwin, J. M. (1896). 'A new factor in evolution', *American Naturalist, vol.30,* 441–451.

Bickerton, D. (1998). 'Catastrophic evolution: the case for a single step from protolanguage to full human language' in Hurford, J., Studdert-Kennedy, M. and Knight, C. (eds.), *Approaches to the Evolution of Language,* Cambridge University Press, Cambridge, pp. 341–358.

Brent, M. and Cartwright, T. (1996). 'Distributional regularity and phonotactic constraints are useful for segmentation', *Cognition, vol.61,* 93–125.

Briscoe, E. J. (2000a). 'Evolutionary perspectives on diachronic syntax' in Pintzuk, Susan, Tsoulas, George, and Warner, Anthony (eds.), *Diachronic Syntax: Models and Mechanisms,* Oxford: Oxford University Press, pp. 75–108.

Briscoe, E. J. (2000b). 'Grammatical acquisition: inductive bias and coevolution of language and the language acquisition device', *Language, vol.76.2,* 245–296.

Calvin, W. (1996). *The Cerebral Code,* MIT Press, Cambridge, MA.

Chomsky, N. (1965). *Aspects of the Theory of Syntax,* MIT Press, Cambridge, MA.

Chomsky, N. (1981). *Government and Binding,* Foris, Dordrecht.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use,* Praeger, New York.

Croft, W. (1990). *Typology and Universals,* Cambridge University Press, Cambridge.

Cziko, G. (1995). *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution,* MIT Press, Cambridge, MA.

Dawkins, R. (1983). 'Universal Darwinism' in Bendall, D. S. (ed.), *Evolution: From Molecules to Men,* Cambridge University Press, Cambridge, pp. 403-425.

Dawkins, Richard (1989). *The Selfish Gene,* Oxford University Press, Oxford, 2nd edition.

de Marcken, C. (1996). 'Unsupervised language acquisition', Doctoral dissertation, MIT, EECS.

Deacon, T. (1997). *The Symbolic Species: Coevolution of Language and Brain,* MIT Press, Cambridge, MA.
Dennett, D. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon and Schuster, New York.
Fodor, J. D. (1981). 'Does performance shape competence?', *Phil. Trans. Royal Society of London, vol.B295,* 285–295.
Gibson, E. (1998). 'Linguistic complexity: locality of syntactic dependencies', *Cognition, vol.68,* 1–76.
Gibson, E. and Wexler, K. (1994). 'Triggers', *Linguistic Inquiry, vol.25.3,* 407–454.
Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency,* Cambridge University Press, Cambridge.
Hurford, Jim (1987). *Language and Number,* Blackwell, Oxford.
Hurford, J. (1989). 'Biological evolution of the Saussurean sign as a component of the language acquisition device', *Lingua, vol.77,* 187–222.
Hurford, J. (1999). 'The evolution of language and languages' in Dunbar, R., Knight, C., and Power, C. (eds.), *The Evolution of Culture,* Edinburgh University Press, Edinburgh, pp. 173–193.
Hurford, J., Studdert-Kennedy, M., and Knight, C. (1998). *Approaches to the Evolution of Language,* Cambridge University Press, Cambridge.
Kauffman, S. (1993). *The Origins of Order: Self-Organization and Selection in Evolution,* Oxford University Press, New York.
Keller, R. (1984 (1994)). *On Language Change: The Invisible Hand in Language* (English translation), Routledge, London.
Kirby, S. (1998). 'Fitness and the selective adaptation of language' in Hurford, J., Studdert-Kennedy, M., and Knight, C. (eds.), *Approaches to the Evolution of Language,* Cambridge University Press, Cambridge, pp. 359–383.
Kirby, S. (1999). *Function, Selection and Innateness: The Emergence of Language Universals,* Oxford University Press, Oxford.
Kirby, S. (2000, in press). 'Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity', *IEEE Trans. on Evolutionary Computation, vol.8*
Jackendoff, R. (1997). *The Architecture of the Language Faculty,* MIT Press, Cambridge, MA.
Langacker, R. (1977). 'Syntactic reanalysis' in Li, C. (ed.), *Mechanisms of Syntactic Change,* University of Texas Press, Austin, pp. 57–139.
Lightfoot, D. (1979). *Principles of Diachronic Syntax,* Cambridge University Press, Cambridge.
Lightfoot, D. (1999). *The Development of Language: Acquisition, Change, and Evolution,* Blackwell, Oxford.
Lindblom, B. (1986). 'Phonetic universals in vowel systems' in Ohala, J. and Jaeger, J. J. (eds.), *Experimental Phonology,* Academic Press, Orlando, FL, pp. 13–44.
Mayley, G. (1996). 'Landscapes, learning costs and genetic assimilation' in Turney, Peter, Whitley, D., and Anderson, R. (eds.), *Evolution, Learning and Instinct: 100 Years of the Baldwin Effect,* MIT Press, Cambridge MA.
Maynard-Smith, J. (1998). *Evolutionary Genetics,* Oxford University Press, Oxford, 2nd edition.
McMahon, A. (1994). *Understanding Language Change,* Cambridge University Press, Cambridge.

Miller, G. A. and Chomsky, N. (1963). 'Finitary models of language users' in Luce, R. D., Bush, R. R. and Galanter, E. (eds.), *Handbook of Mathematical Psychology,* vol II, Wiley, New York, pp. 419–491.

Mitchell, Tom (1990). 'The need for biases in learning generalizations' in Shavlik, Jana and Dietterich, Tom (eds.), *Readings in Machine Learning,* Morgan Kaufmann, San Mateo, CA.

Mitchell, Tom (1997). *Machine Learning,* McGraw Hill, New York.

Muggleton, S. (1996). 'Learning from positive data', *Proceedings of the 6th Inductive Logic Programming Workshop,* Stockholm.

Nettle, Daniel (1999). *Linguistic Diversity,* Oxford University Press, Oxford.

Newmeyer, F. (1991). 'Functional explanation in linguistics and the origins of language', *Language and Communication, vol.11,* 3–28.

Newmeyer, F. (1998). *Language Form and Language Function,* MIT Press, Cambridge, MA.

Niyogi, P. and Berwick, R. (1997). 'A dynamical systems model of language change', *Linguistics and Philosophy, vol.17,*

Osborne, M. and Briscoe, E. J. (1997). 'Learning stochastic categorial grammars', *Proceedings of the ACL Comp. Nat. Lg. Learning (CoNLL97) Workshop,* Madrid, pp. 80–87.

Pinker, S. and Bloom, P. (1990). 'Natural language and natural selection', *Behavioral and Brain Sciences, vol.13,* 707–784.

Rambow, O. and Joshi, A. (1994). 'A processing model of free word order languages' in Clifton, C., Frazier, L., and Rayner, K. (eds.), *Perspectives on Sentence Processing,* Lawrence Erlbaum, Hillsdale, NJ., pp. 267–301.

Ridley, M. (1990). 'Reply to Pinker and Bloom', *Behavioral and Brain Sciences, vol.13,* 756.

Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry,* World Scientific, Singapore.

Rissanen, J. and Ristad, E. (1994). 'Language acquisition in the MDL framework' in Ristad, E. (ed.), *Language Computations,* DIMACS Workshop, American Mathematics Association.

Sharpe, T. (1997). *Acquiring a shared vocabulary,* MPhil Dissertation, University of Cambridge.

Steedman, M. (2000). *The Syntactic Process,* MIT Press, Cambridge, MA.

Steels, L. (1998). 'Synthesizing the origins of language and meaning using coevolution, self-organization and level formation' in Hurford, J., Studdert-Kennedy, M., and Knight, C. (eds.), *Approaches to the Evolution of Language,* Cambridge University Press, Cambridge, pp. 384–404.

Waddington, C. (1942). 'Canalization of development and the inheritance of acquired characters', *Nature, vol.150,* 563–565.

Worden, R. P. (1998). 'The evolution of language from social intelligence' in Hurford, J., Studdert-Kennedy, M., and Knight, C. (eds.), *Approaches to the Evolution of Language,* Cambridge University Press, Cambridge, pp. 148–168.