Part V

Conclusions

Chapter 14

Conclusions

Parts II, III, and IV of this book have explored the themes of language, learning, and evolution in some detail. In this concluding chapter, we take stock of the situation, assess the progress made so far, and outline important directions for future understanding.

Recall the essential logic of the argument presented here. There are three main observations.

1. Language is composed of units such as distinctive features, phonemes, syllables, morphemes, phrases, and so on. In any particular linguistic system, these have systematic relationships with each other. These relationships may be given a formal characterization (algebraic or statistical). Thus linguistic knowledge and behavior is usefully captured by an underlying computational system (grammar).

2. At any time t there may be variation among the linguistic systems of different mature speakers in the population. This is the synchronic variation (across space) at time t.

3. Learning is the process by which language is transmitted to new speakers — children and others who enter the population by birth or recent migration. This gives rise to evolutionary dynamics at the population level. Thus, we relate learning at the individual level to change and evolution in the linguistic characteristics of the population over generational time.

It is difficult to reason about the relationship between (1),(2), and (3) by verbal arguments alone. For this reason, a number of computational and mathematical models were considered. We do not posit any single model of language evolution. Rather, our goal here has been to lay the foundations for a framework within which a number of different models may be constructed to understand different aspects of the phenomena at hand.

14.1 A Summary of the Major Insights

Let us now review our major results to highlight some of the important insights we have obtained. It is worthwhile to emphasize that many of these insights are only obtained by the construction of a mathematical model that brings questions and issues into sharper focus than before.

14.1.1 Learning and Evolution

Learning at the individual level and evolution at the population level are related. Furthermore, we see that different learning algorithms may 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 have explored many different learning algorithms and worked out their evolutionary consequences. In Chapter 5, we developed a basic understanding of two different linguistic systems in competition with each other. Three learning algorithms were studied. A memoryless learner (the TLA), a batch learner, and an asymmetric cue based learner. These three had different evolutionary dynamics. Not only were the precise equations for the dynamical system different, their qualitative behaviors were also different. The memoryless learner gave rise to a single stable attractor to which the population converged from all initial conditions. The batch learner gave rise to a bistable situation with *two* stable attractors. The cue based learner had two different regimes — one with a single stable attractor and one with two stable attractors. Each of these three learning algorithms satisfies the learnability criterion. In a homogeneous setting with a single target grammar, the learning algorithm identifies the target in the limit as the data goes to infinity. Yet, they have different evolutionary properties.

This theme repeated itself across many chapters. In Chapter 7 on Portuguese, three different algorithms were studied. These were the Galves Batch Algorithm, the Batch Subset Algorithm, and the Online Learning Algorithm (TLA). Different evolutionary dynamics were obtained. In Chapter 8 on Chinese, four different cases were examined. We see that there is a difference depending upon whether learners make a categorical decision or a blending decision in phonetic/phonological acquisition.

We have also seen 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. These polynomial maps were encountered in a number of different settings in Chapters 5, 6, 7, 8, and 13. 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. In particular, we see that there is a qualitative difference in evolutionary behavior between *small* n and *large* n settings. In Chapter 12, we uncovered a similar relationship between evolvability and learning fidelity. Learning fidelity is a measure of how well the learner learns a potential target language and this is naturally related to the amount of linguistic data it receives over the learning period. Thus, a change in the developmental lifecycle of proto-humans (so that a longer "gestation" period was available for language learning) could result in a qualitatively different evolutionary pattern for the emergence of communal languages.

Finally, we 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 theme was explored in Chapter 8 (Chinese), Chapter 9 (models of cultural evolution) and most interestingly in Chapters 12 and 13. We comment on the findings of Chapters 12 and 13 shortly.

Throughout, we have considered examples of language change to ground our abstract discussion in linguistic reality and to provide some sense to the reader as to how linguistic data may be engaged using the approaches described here. To this effect, we discussed phonological changes in Chinese and Portuguese and syntactic change in French and English. An assortment of other changes are scattered across the book to motivate the discussion.

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 has dropped the homogeneous assumption and concentrated on analyzing the implications of learning theory in a heterogeneous population with linguistic variation. Learning theory has not been systematically developed in such a context before.

14.1.2 Bifurcations in the History of Language

A major insight that emerges from the analytic treatment pursued over the preceding chapters has to do with the existence of bifurcations (phase transitions) in the dynamics of language evolution. These bifurcations may be viewed as appropriate explanatory constructs to account for major transitions in language. Throughout this book, we have derived the dynamics of linguistic populations under a variety of assumptions. Again and again, we notice that (a) the dynamics is typically non-linear (b) there are bifurcations and these may be interpretable in linguistic terms as the change of language from one seemingly stable mode to another. We have encountered numerous such examples of bifurcations in this book.

In Chapter 5, we considered models with two languages in competition. For a TLA based learner for learning in the P&P model, we noted that the equilibrium state depended 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 saw that it was 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 saw that there was 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) varied as a function of each other. For models inspired by change in European Portuguese or those inspired by phonological change in Chinese, similar bifurcations arose. In Chapters 12 and 13 we studied the emergence of grammatical coherence. We saw that there was a bifurcation point below which the only stable solution was the uniform solution where all languages are equally represented in the population. Above this bifurcation point, *one-language* solutions were seen to emerge. These correspond to coherent states where the entire population has converged to the same language. This bifurcation point was related to the learning fidelity of individual learners.

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.

14.1.3 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 $> \gamma$). In the light of the discussion on 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.

14.2 Future Directions

It is clear that a number of simplifying idealizations needed to be made to obtain the results and insights discussed above. There are a number of issues that are still poorly understood and a number of directions to take the models towards greater realism. We discuss these below.

1. In the models we have considered there have usually been two different settings. Either children learn from equal exposure to all members in the population (most of the models in this book) or children learn from individual members alone (parents or randomly selected teachers as in Chapter 12). In many cases of interest, there is a more complicated social structure that is best expressed as a network of influences on the learning child. In that case, the population ought to be modeled as a graph where each node of the graph denotes a location or individual and the edges denote the influences that individuals (locations) have on each other. In this book, we have essentially considered complete graphs and degenerate graphs where there are no edges at all. Dynamics when the graph is locally connected were briefly explored in Chapter 10 and then again in Chapter 13, but much more remains to be done. A better understanding is ultimately required of the effect that network topology has on the dynamics of language evolution. This direction will tie together the recent interest in social networks (see, for example, Strogatz, 2003 or Newman et al, 2003) with our interest in language evolution.

2. Two other aspects of the population structure were briefly examined in Chapter 10. One is related to the fact that population sizes are finite. This results in a stochastic process rather than a dynamical system and the relationship between the stationary distribution of this process and the attractors of the dynamical system need to be better understood. In particular, how do the bifurcations that occur in the dynamical systems manifest themselves in the corresponding processes? A second is related to the fact that most of the models have assumed a generational structure that is blocked into discrete generations with dependence only between successive generations. It is of interest to know the effect of more complicated generation structures on learning and evolution.

3. In much of our analysis, we have assumed that the grammatical system

of a human may be characterized by a single computational system (in the sense of Church-Turing). This corresponds to a speaker-hearer being monolingual in the traditional sense of generative linguistics. In the presence of variation, children might become effectively bilingual (multilingual). The effect of this was briefly studied in a few chapters. For example, in Chapter 8 on phonological change in Chinese, we considered both bilingual and monolingual models of learning and observed that they could have different evolutionary consequences. Again, in Chapter 10, some preliminary models of bilingual learning were discussed.

4. Closely related to the previous point is the relationship between first and second language learning and their relative effects on language change. In the models of this book, we have never made a distinction between these two possibly different modes of learning, their interrelationships and their evolutionary consequences. One can consider learning models with two learning phases corresponding to the acquisition of a native language and its characteristics followed by a later period of second language acquisition. Such models would have particular relevance to the analysis of the linguistic adaptation of immigrants in a new society.

5. The general framework presented here makes no commitment to details of linguistic theory or learning theory. Every choice of $\mathcal{H}$ — the class of human grammatical systems and $\mathcal{A}$ — the learning algorithm used by children would give rise to a potentially different population dynamics. Many of the examples were worked out for grammars in the Principles and Parameters tradition of grammatical theory and appropriate learning algorithms in that context. It is possible to provide an analogous development where the entire analysis is carried out within an alternative framework such as Optimality Theory. We leave such alternative analyses to future work.

6. The role of fitness and natural selection is poorly understood at present. We presented an analysis in Chapters 11 and 12 but many problems remain. First off, what is the sense in which knowing a language confers fitness upon the individual that knows it? The particular characterization of communicative efficiency as a proxy for fitness is motivated heavily by an information-theoretic view of language that views the primary purpose of language to be communication. But it may well

be that communicative ability is an accidental side-effect of linguistic competence. It may be that the true function of language is that it confers the ability for thought and reasoning. In that case, communicative success might have little to do with language and the evolutionary pressures would be very different from those developed here. An empirical study of the structure of lexical items already suggests that in the narrow sense of communicative efficiency developed in this book, languages do not seem to be optimized for communication. However, much more remains to be done here.

7. In this book, we have failed to provide any genuine insight into the evolution of *novel* structures. Much of our analysis has focused on the competition between existing structures. We have always assumed a preexisting class $\mathcal{H}$ of possible communication systems and this determines the state space of the dynamical system. The sense in which novelty arises in our framework is the following. One may consider initial conditions of the population that are in some region of the state space where there are many unattested communication systems. Now the dynamical process might unfold in such a way that previously unattested systems might come into existence. We have seen how bifurcations may provide a mechanism for this. However, these unattested systems were possible by construction of the evolutionary process itself. In contrast, one might ask — how does one formulate an evolutionary process where the state space itself is modified by reconstruction so that genuinely new structures are seen to evolve. For example, an important question from a linguistic point of view may be: how did recursion evolve? A high level answer that may be given at this point would claim that recursion evolves in the following stages. In stage one, there are systems with non-recursive rules defined over primitive objects. Primitive communication systems that map *symbol* to *object* like certain kinds of alarm call systems in animal communication possibly belong to this type. In the second stage, one finds the evolution of categories (groups of primitive objects). In the final stage, there are rules defined over categories. Once one evolves rules over categories, recursion would follow almost immediately. A cogent formulation of this general evolutionary sequence is an important problem for future work.

8. Part IV of this book deals with the conditions necessary for the emer-

gence of coherence. We see that for linguistic coherence to emerge, learning fidelity must be high. In general, for learning fidelity to be high, $\mathcal{H}$ needs to be a highly restricted family (in our example model, the cardinality n must be low). So there is clearly some evolutionary pressure for $\mathcal{H}$ to be a small set. Why, then, does not the set $\mathcal{H}$ become smaller and smaller until it becomes a singleton set, consisting only of one language? In that case, $\mathcal{H}$ would eventually consist of only one language eliminating the need for any learning or the possibility of any linguistic variation in the population. We have identified pressures that force $\mathcal{H}$ to be small — both from learning-theoretic and evolutionary considerations. What are the pressures that force $\mathcal{H}$ to be large? A plausible argument is that if $\mathcal{H}$ consists of only one language, then the details of this language must be encoded in the genome sequence and would take up a lot of space. Therefore a large $\mathcal{H}$ corresponds to smaller requirements on the genetic code and greater flexibility in some sense. Clarifying these issues in some quantitative way presents another important direction for future work.

9. We have discussed human language and its origins in some detail over the course of this book but the framework is applicable in general to the evolution of communication systems that are acquired by learning. In the context of the evolutionary origin of human language, relevant data seems to be difficult to come by. It may therefore be more profitable to explore some of the natural evolutionary questions in the context of animal communication where data may be more easily collected. One observes specialized communication systems in a large number of animal species. For example, there is the familiar bee dance which bees use to communicate a source of food to other bees in their hive. The dance follows a figure of eight and has three parameters: (i) the principal axis points towards the direction of food, (ii) the time taken to traverse the figure represents the distance to the source, (iii) the vigorousness of the dance represents the amount of food present. Can one provide a coherent account of how such a specialized communication system arose? Similarly, can one provide accounts of the evolution of songs in birds, or communicating signals in bats? These present significant challenges for the future.

14.2.1 Empirical Validation

While model construction allows us to get a sharper understanding of the questions and issues in language evolution, ultimately these models remain speculative unless they are used to engage empirical facts in a productive way. The empirical validation of these models remains an important ongoing direction of future work

It needs to be properly understood that in a historical discipline like evolution, controlled experiments are difficult to design and so the kind of success one wishes for is not the sort that one sees in certain areas of physics or engineering. Rather, one hopes to get the sort of understanding that one obtains from mathematical models in evolutionary biology. One hopes that one will be able to separate plausible from implausible theories, sort out inconsistencies in reasoning, and generally obtain a deeper qualitative understanding of the phenomena. At the same time, it is worth noting that even in many areas of physics, numerical match to the data is difficult. Nevertheless quantitative models are developed and these play an important role in qualitative understanding. Various aspects of fluid mechanics and turbulence are a good example. It is impossible to predict the motion of a leaf on a stormy day — yet dynamical systems models of the general properties of fluid flow have been developed. Similarly, areas of cosmology, pattern formation (sand piles), soft condensed matter physics and so on present further examples.

As examples of concrete empirical phenomena that have been developed within the mathematical framework of this book, let us recall the case of French. We saw that there was a syntactic change leading to the loss of V2 and loss of pro-drop in the grammar of French. One may make models of this change under different assumptions: (a) speakers are bilingual and there is grammatical variation within each speaker (b) speakers are monolingual and the locus of variation is at the level of the population (c) there is linkage among the grammatical parameters so that they may not be treated as independent, i.e., the nature of the parameterization matters, and so on. Depending on the precise assumptions several different kinds of evolutionary models were explored in Chapters 6 and 10 in some detail. Most interestingly, we found that not all models were consistent with the historically observed trends. From this kind of analysis we were led to conclude that (i) V2 must have been lost before pro-drop was lost (ii) the loss of V2 was triggered by the increase in the frequency with which pronominal subjects were used. Both (i) and (ii) therefore count as predictions that are independently verifiable by further data collection. Similar examples of syntactic and morphological change were provided for English (Chapter 9) and Portuguese (Chapter 7) respectively. We worked through different kinds of models for each of those cases. These models allowed us to check our intuitions about the forces that led to the change and to reason about the interplay between learning and change more generally.

Another example of an empirically motivated study was provided by the example of phonological change in the Wu dialect considered in Chapter 8. We saw that a phonological merger between a diphthong and a monophthong was attested in the Wu dialect of Wenzhou province. As a result, several pairs of rhyming words became homophonous. A number of questions could now be asked. What initiated the change? Why did the change go to completion? Under what conditions do we expect competing forms to coexist in the population? What is the possible effect on population dynamics of noise in the perception and production of speech?

In order to reason about these questions, we considered several different models for the acquisition of phonological form. We saw that if individuals treated phonological forms in a categorical way, then equilibrium states of the population corresponded to pure states where all members used the same form. In contrast, if individuals used both forms then equilibrium states were not pure and there was no tendency for language to change or for the change to go to completion. Thus different hypotheses about individual behavior were separated and these predictions about individual behavior may be empirically falsified. The bifurcations that arise in such models were seen as suitable constructs to account for the actuation problem.

In Chapter 11, we discussed a formulation of communicative efficiency that was based on the notion of information transfer from speaker to hearer. Using a variant of such a formulation, we saw that the lexicon of English was not optimally adapted to the perceptual limitations of humans. In particular, phonetic distinctions that are hard to make seem to bear great functional load. This provided an immediate empirical challenge to the notion that language evolves in the direction of greater communicative efficiency.

We hope that these examples provide a sense to the reader of how one may engage empirical facts. But this has been a largely theoretical book and much more remains to be done. Historical linguistics is filled with examples of language change and competing explanations that account for them. We provided some illustrative examples in Chapter 1 and empirically grounded studies need to be performed in large measure before the approach embodied in this book realizes its full potential.

The empirical validation of models for the origin of language presents an even greater challenge. At the present moment in time, there is very little data that points directly to the different stages in the origin of human language from prelinguistic versions of it. However, there is much more data on other kinds of natural communication systems found in the animal world (see Hauser, 1997 for an account). Perhaps the most empirically fruitful direction of future work would be to sharpen the questions and models of evolutionary thinking by considering case studies of the origin of such communication systems. This presents an important direction of future work.

14.2.2 Connections to Other Disciplines

A central theme of this book is the relationship between local interactions at the individual level and global behavior at the population level. This theme arises in a number of different disciplines and obvious connections emerge.

We have noted at many points the analogies between evolution in linguistic populations and evolution in biological populations. Grammars, like genes, are formal and discrete. Unlike genetic evolution, however, grammars are transmitted from one generation to the next via learning. The evolutionary dynamics that result from this different transmission law has been the central concern of this book. Nevertheless, there are strong synergies between the methods and concerns of evolutionary biology and those of evolutionary linguistics. At some point in the future when the biological basis of grammars are understood, perhaps the two disciplines will be unified more deeply and it will make sense to study gene-grammar co-evolution in human populations.

Statistical physics has long concerned itself with deriving the characteristics of the ensemble from the statistics of individual particles. There is particular interest in the existence and nature of phase transitions and these phase transitions play an important role in describing the behavior of materials at different temperatures. We considered the analogies between language evolution systems and the Ising systems of statistical physics in Chapter 13. In the Ising model, each particle has a spin that takes one of two values. The particles are identified with the vertices of a graph and the connectivity of the graph determines how the particles interact. The temperature T controls the strength of the interaction and one finds a phase transition in the equilibrium spin distribution as the temperature changes smoothly. Above a critical temperature T_c, in the high thermal noise regime, the only equilibrium distribution is one in which as many spins are up as down on average.

Below the critical temperature spontaneous spin alignment may occur.

This phase transition was analogous to the bifurcations we observed in the models of language evolution in Chapter 13. Linguistic agents were like particles, languages were like spins, and learning fidelity (quantified by the parameters a and k) was like temperature. A bifurcation led to the transition between incoherent regimes in which the only stable mode was the uniform mode to coherent regimes in which "alignment of languages" occurred so that a majority of the population spoke a shared language. In this book, we have analyzed the interplay between individual behavior and the ensemble by taking the continuum limit of individual agents, deriving non-linear dynamical systems and invoking the theory of bifurcations in non-linear dynamics. It is our strong intuition, that it is possible to pursue an alternative development using the techniques of percolation theory and statistical mechanics leading to the characterization of language change in terms of phase transitions. We leave this as a possibility for the future.

Linguistic behavior is grounded in our biology yet is closely linked to our cultural identity. Consequently, language evolution may also be viewed as a particular instantiation of cultural evolution. Indeed, many theories of cultural evolution (notably those of Cavalli-Sforza and Feldman) have taken this viewpoint. We considered the relationship between the Cavalli-Sforza theory and ours in Chapter 9. It is possible that some of the tools and techniques described here could also serve theories of cultural evolution in other domains. The evolution of music, writing, or games are particularly good examples since learning plays an important role in all three. However, the form that learning theory must take to best describe the transmission process in each of these domains is probably closer to teaching than to the kind of unconscious learning that takes place in language. More fundamental work will be needed to build on these connections profitably. The game-theoretic analysis of evolutionary processes implicit in Boyd and Richerson may also be used to model the evolution of language. It would be interesting to see the development of theories of language evolution in those terms.

Other areas of the social sciences with which synergies exist are economics and psychology. In economics, there is a long tradition of understanding the macroscopic behavior of economic systems from the behavior of individual economic agents. Examples such as social choice theory, welfare economics, and evolutionary economics spring to mind. A recent and unusual exploration of the interface between economics and linguistics may be found in Rubinstein (2000). Evolutionary thinking is beginning to play a role in psychology with the development of evolutionary psychology where one studies the evolution of cognitive traits. Since language surely counts as an important cognitive trait, the study of language evolution embodied in this book may also be viewed as a contribution towards an understanding of the principles of evolutionary psychology in a concrete and empirically verifiable context.

There are many close ties to computer science and artificial intelligence. A grammar is a computational system and learning is an algorithmic process. The heart of our approach is therefore grounded in the techniques of computer science. This monograph may be viewed as a contribution to mathematical and computational linguistics, a legitimate sub-discipline of artificial intelligence. These connections are obvious and there is no need to elaborate on them any further. It is worthwhile, however, to discuss some other areas of artificial intelligence that have points of contact with the ideas in this book. The areas of artificial life and multi-agent systems deal with the behavior of groups of intelligent agents with a particular focus on emergent phenomena in those communities. This varies from market based models of computational economies (see Boutilier et al, 1997), herding and flocking behavior in robot communities (Mataric, 1998) , and coordination in multi-agent systems (Shoham and Tennenholtz, 1997; Tohme and Sandholm, 1999). Part IV of this book focuses on emergent phenomena in a population of linguistic agents and the results there may also be viewed as a theoretical contribution towards artificial life. The artificial life approach to language evolution is most strongly represented in the work of Luc Steels.

14.3 A Concluding Thought

We have attempted here to provide an evolutionary perspective on language. How and why languages change, why they take the form they do, and how the structure, the acquisition, and the evolution of language are all interdependent.

In the understanding of social and natural phenomena, it is often the case that a historical perspective provides one with a deeper appreciation of why things are the way they are. In this context, it is worth recalling Dobzhansky's famous remark:

> Nothing in biology makes sense except in the light of evolution.
>
> (T. Dobzhansky)

In the natural world, language and communication are grounded in the biology of living organisms. We hope that such an evolutionary perspective will provide a richer understanding of the fundamental nature of human language, and more generally of communication in human, animal, and machine.

Bibliography

[1] N. Abe and M. Warmuth. On the computational complexity of approximating distributions by probabilistic automata. In *Proc. of Computational Learning Theory*, 1992.

[2] D. M. Abrams and S. H. Strogatz. Modelling the dynamics of language death. *Nature*, 424(900), 2003.

[3] J. Aitchison. *The Seeds of Speech: Language Origin and Evolution.* Cambridge University Press, 2000.

[4] G. Altmann, H. van Buttlar, W. Rott, and U. Strau. A law of change in language. In B. Brainerd, editor, *Historical linguistics*, pages 104–115. Bochum:Studienverlag Dr. N. Brockmeyer, 1983.

[5] D. Angluin and M. Kharitonov. When won't membership queries help. *Journal of Computer and System Sciences*, 50:336–355, 1995.

[6] Dana Angluin. Finding patterns common to a set of strings. *Journal of Computer and System Sciences*, 21:46–62, 1980.

[7] Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45:117–135, 1980.

[8] Dana Angluin. Learning regular sets from queries and counterexamples. *Information and Computation*, 75:87–106, 1987.

[9] Dana Angluin. Identifying languages from stochastic examples. Technical Report YALEU/DCS/RR-614, Yale University, March 1988.

[10] A. Ariew and R. Lewontin. The confusions of fitness. *British Journal for the Philosophy of Science*, 55:347–363, 2004.

[11] Martin Atkinson. *Children's Syntax: Introduction to Principles and Parameters Theory.* Blackwell Publishers, Oxford, 1992.

[12] R. Axelrod. *The Evolution of Cooperation.* Basic Books: New York, 1984.

[13] Charles-James Bailey. *Variation and Linguistic Theory.* Center for Applied Linguistics, Washington D.C., 1973.

[14] J. Batali. Computational simulations of the emergence of grammar. In C. Knight J. R. Hurford, M. Studdert-Kennedy, editor, *Approaches to the Evolution of Language: Social and Cognitive Bases*, pages 405–426. Cambridge University Press, 1998.

[15] L. Bauer. *Watching English Change.* Longman, 1994.

[16] Stefan (ed.) Bertolo. *Language Acquisition and Learnability.* Cambridge University Press, 2001.

[17] R. C. Berwick. *The Acquisition of Syntactic Knowledge.* The MIT Press, Cambridge, MA, 1985.

[18] M. Blum and L. Blum. Towards a mathematical theory of inductive inference. *Information and Control*, 28:125–155, 1975.

[19] C. Boutilier, Y. Shoham, and M. Wellman. Economic principles of multi-agent systems. *Artificial Intelligence Journal*, 94(1-2):1–6, 1997.

[20] R. Boyd and P. Richerson. *Culture and the Evolutionary Process.* The University of Chicago Press, 1985.

[21] L. E. Breivik and E. H. Jahr. *Language Change: Contributions to the Study of Its Causes.* Mouton de Gruyter, 1989.

[22] M. Brent. Surface cues and robust inference as a basis for the early acquisition of subcategorization frames. In L. Gleitman and B. Landau, editors, *The Acquisition of the Lexicon*, pages 433–470. MIT Press, 1994.

[23] M. Brent. An efficient, probabilitistically sound algorithm for segmentation and word discovery. *Machine Learning*, 45:71–105, 1999.

[24] M. Brent and T. Cartwright. Distributional regularity and phonotactics are useful for segmentation. *Cognition*, 61:93–125, 1996.

[25] Joan Bresnan. *Lexical-Functional Syntax.* Blackwell Publishers, Oxford, 2001.

[26] E. J. Briscoe. Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2):245–296, 2000.

[27] Helena Britto, Charlotte Galves, Ilza Ribeiro, Marina Augusto, and Ana Paula Scher. Morphological annotation system for automatic tagging of eletronic textual corpora: From english to romance languages. In Oriente, editor, *Anais do VI Simposio Internacional de Comunicacion Social*, pages 582–589. Santiago de Cuba, 1999.

[28] R. Brown and C. Hanlon. Derivational complexity and order of acquisition in child speech. In J. R. Hayes, editor, *Cognition and the Development of Language.* Wiley, NY, 1970.

[29] R. Bush and F. Mosteller. *Stochastic Models for Learning.* John Wiley, 1955.

[30] A. Cangelosi and D. Parisi. *Simulating the Evolution of Language.* Springer-Verlag, 2002.

[31] J. Case and C. Smith. Comparison of identification criteria for machine inductive inference. *Theoretical Computer Science*, 25:193–220, 1983.

[32] L. Cavalli-Sforza. *Genes, Peoples, and Languages.* University of California Press, 2001.

[33] L. L. Cavalli-Sforza and M. W. Feldman. *Cultural transmission and evolution: a quantitative approach.* Princeton University Press, 1981.

[34] Eugene Charniak. *Statistical Language Learning.* The MIT Press, Cambridge MA, 1993.

[35] D. Cheney and R. M. Seyfarth. *How Monkeys See the World: Inside the Mind of Another Species.* University of Chicago Press, Chicago, IL, 1990.

[36] N. Chomsky and M. Schutzenberger. The algebraic theory of context-free languages. In P. Braffort and D. Hirschberg, editors, *Computer Programming and Formal Languages*, pages 118–161. North Holland, 1963.

[37] Noam Chomsky. *Aspects of the Theory of Syntax.* The MIT Press, Cambridge MA, 1965.

[38] Noam Chomsky. *Lectures on Government and Binding.* Dordrecht : Foris Publications, 1981.

[39] Noam Chomsky. *Knowledge of Language: Its Nature, Origin and Use.* Praeger, NY, 1986.

[40] Noam Chomsky. *The Minimalist Program.* MIT Press, Cambridge, MA, 1995.

[41] D. Christian, W. Wolfram, and N. Bube. *Variation and change in geographically isolated communities: Appalachian English and Ozark English.* University of Alabama Press, 1988.

[42] M. Christiansen and S. Kirby. *Language Evolution.* Oxford University Press, 2003.

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

[44] Stephen Crain. Language acquisition in the absence of experience. *Brain and Behavioral Sciences*, pages 597–612, 1991.

[45] Stephen Crain and Rosalind Thornton. *Investigations in Universal Grammar: A Guide to Experiments in the Acquisition of Syntax and Semantics.* The MIT Press, Cambridge MA, 1998.

[46] W. Croft. *Explaining Language Change: An Evolutionary Approach.* Harlow, Essex: Longman, 2000.

[47] J.F. Crow and M. Kimura. *An Introduction to Population Genetics Theory.* Burgess Publishing Company, Minneapolis, MN, 1970.

[48] F. Cucker, S. Smale, and D-X. Zhou. Modeling language evolution. *Foundations of Computational Mathematics*, 4(3), 2004.

[49] Walter Daelemans. Abstraction considered harmful: Lazy learning of language processing. In *Proceedings of the 6th Belgian-Dutch Conference on Machine Learning*, pages 3–12, 1996.

[50] C. de Marcken. *Unsupervised Language Acquisition.* PhD thesis, M.I.T., Cambridge, MA, 1996.

[51] Carl de Marcken. Parsing the LOB corpus. In *Meeting of the Association for Computational Linguistics*, pages 243–251, 1990.

[52] F. de Saussure. Course in general linguistics. In C. Bally, A. Sechehaye, and R. Harris, editors, *Course in General Linguistics: English translation.* Duckworth, London, 1983.

[53] Ferdinand de Saussure. *Course in General Linguistics.* 1916.

[54] M. DeGraff. *Language Creation and Language Change.* MIT Press, 2001.

[55] M. Demetras, K. Post, and C. Snow. Feedback to first language learners: The role of repetitions and clarification questions. *Journal of Child Language*, 13:275–292, 1986.

[56] B. E. Dresher and J. D. Kaye. A computational learning model for metrical phonology. *Cognition*, 34:137–195, 1990.

[57] William S-Y. Wang (ed.). *The lexicon in phonological change.* The Hague: Moulton, 1977.

[58] M. Eigen and P. Schuster. *The Hypercycle: A Principle of Natural Self Organization.* Springer, Berlin, 1994.

[59] J. Elman, E. Bates, M. Johnson, A. Karmiloff-Smith, Domenico Parisi, and Kim Plunkett. *Rethinking Innateness.* MIT Press, 1996.

[60] J. Feldman. Some decidability results on grammatical inference and complexity. *Information and Control*, 20(3):244–262, 1972.

[61] Jerome A. Feldman, George Lakoff, David Bailey, Srini Narayanan, Terry Regier, and Andreas Stolcke. L0 - the first five years of an automated language acquisition project. *Artificial Intelligence Review*, 10:103–129, 1996.

[62] R. Fernandez and A. Galves. Identifying features in the presence of competing evidence: The case of first language acquisition. In *Dynamical Systems: From crystal to chaos*, pages 52–62. World Scientific, 2000.

[63] Roberto Fernandez and Antonio Galves. Identifying features in the presence of competing evidence: The case of first-language acquisition. 1999.

[64] Ronald Fisher. *The Genetical Theory of Natural Selection.* Oxford University Press, 1930.

[65] J. D. Fodor. How to obey the subset principle: Binding and locality. In G. Hermon B. Lust and J. Kornfilt, editors, *Syntactic Theory and First Language Acquisition: Cross-linguistic Perspectives. Vol. 2 Binding, Dependencies and Learnability.* Erlbaum, Hillsdale, NJ, 1994.

[66] J. D. Fodor. Unambiguous triggers. *Linguistic Inquiry*, 29:1–36, 1998.

[67] R. Frank and S. Kapur. On the use of triggers in parameter setting. Technical Report IRCS-92-52, University of Pennsylvania Institute for Research in Cognitive Science, February 1993.

[68] A. Galves and C. Galves. A case study of prosody driven language change. Technical report, UNICAMP-University of Sao Paulo, Brazil, 1995.

[69] W. I. Gasarch and C. H. Smith. Learning via queries. *Journal of the ACM*, 39:649–674, 1992.

[70] E. Gibson and K. Wexler. Triggers. *Linguistic Inquiry*, 25:407–454, 1994.

[71] L. R. Gleitman and B. Landau. *Acquisition of the Lexicon.* MIT Press, Cambridge, MA, 1994.

[72] E. M. Gold. Complexity of automaton identification from given data. *Information and Control*, 37:302–320, 1978.

[73] E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

[74] J. Goldsmith. *Autosegmental Phonology.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1976.

[75] J. Goldsmith. *Autosegmental and Metrical Phonology.* Oxford: Basil Blackwell, 1990.

[76] John Goldsmith. Unsupervised learning of the morphology of a natural language. *Computational Linguistics*, 27(2):153–198, 2001.

[77] A. Gorin, S. Levinson, and A. Gertner. Adaptive acquisition of language. In *Proc. of International Conference on Acoustics, Speech, and Signal Processing*, pages 805–808, 1991.

[78] J. Greenberg. *Language Universals with special reference to feature hierarchies.* Hague:Mouton, 1966.

[79] J. Greenberg. *Language Typology: A historical and analytical overview.* Hague:Mouton, 1974.

[80] J. Greenberg, C. Ferguson, and E. Moravcsik. *Universals of Language.* Stanford University Press, 1978.

[81] P. Gupta and D. Touretzky. Connectionist models and linguistic theory: Investigations of stress systems in language. *Cognitive Science*, 18:1–50, 1994.

[82] Liliane Haegeman. *Introduction to Government and Binding Theory.* Blackwell Publishers, 1991.

[83] M. Halle and W. J. Idsardi. General properties of metrical structure. In J. Goldsmith, editor, *The Handbook of Phonological Theory.* Blackwell, 1995.

[84] M. Halle and J-R. Vergnaud. *An Essay on Stress.* MIT Press, Cambridge, MA., 1987.

[85] Morris Halle. Phonology in generative grammar. *Word*, 18:54–72, 1962.

[86] Morris Halle and William Idsardi. General properties of stress and metrical structure. In Eric Ristad, editor, *DIMACS Workshop on Human Language March 20-22, 1992 (Dimacs Series in Discrete Mathematics and Theoretical Computer Science).* American Mathematical Society, 1992.

[87] M. Halliday. *An Introduction to Functional Grammar.* Edward Arnold, London, 1994.

[88] H. Hamburger and K. Wexler. A mathematical theory of learning transformational grammar. *Journal of Mathematical Psychology*, 12:137–177, 1975.

[89] Michael A. Harrison. *Introduction to Formal Language Theory.* Addison Wesley Publishing Company, Reading, UK, 1978.

[90] M. Hauser. *The Evolution of Communication.* The MIT Press, Cambridge, MA, 1997.

[91] J. A. Hawkins and M. Gell-Mann (eds.). *The evolution of Human Languages (Volume XI of the Santa Fe Institute Studies in the Sciences of Complexity).* Addison-Wesley, 1992.

[92] Bruce Hayes. *Metrical Stress Theory: Principles and Case Studies.* University of Chicago Press, Chicago, 1995.

[93] K. Hirsh-Pasek, R. Treiman, and M. Schneiderman. Brown and hanlon revisited: Mothers' sensitivity to ungrammatical forms. *Journal of Child Language*, 11:81–88, 1984.

[94] J. Hofbauer and K. Sigmund. *The Theory of Evolution and Dynamical Systems.* Cambridge University Press, 1988.

[95] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation.* Addison-Wesley Publishing Company, 1979.

[96] James J. Horning. *A study of grammatical inference.* PhD thesis, Stanford University, Stanford, CA, 1969.

[97] J. Hurford. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, 77(2):187–222, 1989.

[98] J. Hurford and S. Kirby. The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi, editors, *Simulating the Evolution of Language.* Springer-Verlag: Berlin, 2001.

[99] James Hurford and Simon Kirby. Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands and I. Harvey, editors, *Fourth European conference on artificial life.* MIT Press, 1997.

[100] D. L. Isaacson and R. W. Madsen. *Markov Chains: Theory and Applications.* John Wiley and Sons, NY, 1976.

[101] C. Isbell, C. Shelton, M. Kearns, S. Singh, and P. Stone. A social reinforcement learning agent. In *Proc. of Agents: Montreal, Canada*, pages 377–384, 2001.

[102] L. Jenkins. *Biolinguistics: Exploring the Biology of Language.* Cambridge University Press, 2001.

[103] William Jones. *The Collected Works of William Jones in 6 Volumes.* London: Robinson and Evans, 1799.

[104] A. K. Joshi, L. S. Levy, and M. Takahashi. Tree adjunct grammars. *Journal of Computer and System Sciences*, 10(1), 1975.

[105] L. P. Kadanoff. *Statistical Physics: Statics, Dynamics, and Renormalization.* World Scientific, 2000.

[106] M. Kanazawa. *Learnable Classes of Categorial Grammars.* CSLI Publications (Stanford), 1998.

[107] R. Kaplan and M. Kay. Regular models of phonological rule systems. *Computational Linguistics*, 20(3):331–379, 1994.

[108] Ronald M. Kaplan and Joan Bresnan. Lexical-functional grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, MIT Press Series on Cognitive Theory and Mental Representation, chapter 4, pages 173–281. MIT Press, Cambridge, MA, 1982.

[109] M. Kearns and L. Valiant. Cryptographic limitations on learning boolean formulae and finite automata. In *Proc. of 21st Annual ACM STOC.*, 1989.

[110] Michael Kenstowicz. *Phonology in Generative Grammar.* Blackwell Publishers, 1994.

[111] M. Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, 1983.

[112] P. Kiparsky. *Phonological Change.* PhD thesis, MIT, Cambridge, MA, 1965.

[113] Paul Kiparsky. *Explanation in Phonology.* Dordrecht: Foris, 1982.

[114] S. Kirby. *Function Selection and Innateness: The Emergence of Language Universals.* Oxford University Press, 1999.

[115] N. Komarova and P. Niyogi. Optimizing the mutual intelligibility of mutual agents in a shared world. *Artificial Intelligence Journal*, 154:1–42, 2004.

[116] N. Komarova, P. Niyogi, and M. Nowak. The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, 209(1):43–59, 2001.

[117] N. Komarova and M. Nowak. Evolutionary dynamics of the lexical matrix. *Bulletin of Mathematical Biology*, 63(3):451–485, 2001.

[118] N. Komarova and M. Nowak. Language dynamics in finite populations. *Journal of Theoretical Biology*, 221:445–457, 2003.

[119] N. L. Komarova and I. Rivin. Harmonic means, random polynomials, and stochastic matrices. *Advances in Applied Mathematics*, 31(2):501–526, 2003.

[120] A. Kroch and A. Taylor. Verb movement in old and middle english: dialect variation and language contact. In A. van Kemenade and N. Vincent, editors, *Parameters of Morphosyntactic Change.* Cambridge University Press, 1997.

[121] A. Kroch, A. Taylor, and D. Ringe. The middle english verb second constraint: A case study in language contact and language change. In S. Herring, S. Schoesler, and P. van Reenen, editors, *Textual Parameters in Older Language, Current Issues in Linguistic Theory.* John Benjamins: Philadelphia, 2000.

[122] Anthony Kroch. Reflexes of grammar in patterns of language change. *Journal of Language Variation and Change*, 1:199–244, 1989.

[123] Anthony Kroch. Syntactic change. In Mark Baltin and Chris Collins, editors, *Handbook of Syntax.* Blackwell Publishers, 1999.

[124] William Labov. *The Social Stratification of English in New York City.* Washington, D.C.: Center for Applied Linguistics, 1966.

[125] William Labov. *Principles of Linguistic Change, Volume 1: Internal factors.* Blackwell Publishers, Oxford, UK, 1994.

[126] D. Lightfoot. *How to Set Parameters: Arguments from Language Change.* MIT Press/Bradford Books, 1991.

[127] D. Lightfoot. Shifting triggers and diachronic reanalyses. In A. van Kemenade and N. Vincent, editors, *Parameters of Morphosyntactic Change.* Cambridge University Press, 1997.

[128] David Lightfoot. *Principles of Diachronic Syntax.* Cambridge University Press, 1979.

[129] David Lightfoot. *The Development of Language: Acquisition, Change and Evolution.* Blackwell Publishers, 1998.

[130] A. Galves M. Cassandro, P. Collet and C. Galves. A statistical physics approach to language acquisiton and language change. *Physica A*, 263:427–437, 1999.

[131] J. M. Macedonia and C. S. Evans. Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology*, 93:177–197, 1993.

[132] B. MacWhinney. The competition model. In B. MacWhinney, editor, *Mechanisms of Language Acquisition.* Lawrence Erlbaum: Hillsdale, NJ, 1987.

[133] B. MacWhinney. A unified model of language acquisition. In J. Kroll and A. DeGroot, editors, *Handbook of bilingualism:Psycholinguistic Approaches.* Oxford University Press, 2004.

[134] Brian MacWhinney. The childes system. *American Journal of Speech Language Pathology*, 5:5–14, 1996.

[135] A. M. Madeira. On clitic placement in european portuguese. In H. van de Koot, editor, *UCL Working Papers in Linguistics 4, University College London*, 1992.

[136] Christopher D. Manning and Hinrich Schutze. *Foundations of Statistical Natural Language Processing.* The MIT Press, Cambridge MA, 1999.

[137] M. R. Manzini. Second position dependencies (talk given at the 8th workshop on germanic syntax, troms). 1992.

[138] Maja J. Mataric. Coordination and learning in multi-robot systems. *IEEE Intelligent Systems*, pages 6–8, March-April 1998.

[139] R. M. May. *Stability and Complexity in Model Ecosystems.* Princeton University Press, 1973.

[140] A. McMahon. *Change, Chance, and Optimality.* Oxford University Press, 2000.

[141] G. Miller. *The Science of Words.* New York: Scientific American Library, 1996.

[142] J. Milroy. *Linguistic variation and change: On the historical sociolinguistics of English.* Oxford: Blackwell, 1992.

[143] J. Milroy and L. Milroy. Linguistic change, social network and speaker innovation. *Journal of Linguistics*, 21:339–384, 1985.

[144] J. Milroy and L. Milroy. *Real English: The Grammar of English Dialects in the British Isles.* London: Longman, 1993.

[145] W.G. Mitchener. Bifurcation analysis of the fully symmetric language dynamical equation. *Journal of Mathematical Biology*, 43(3):265–285, 2003.

[146] W.G. Mitchener. *A Mathematical Model of Human Languages: The Interaction of Game Dynamics and Learning Processes.* PhD thesis, Princeton University, Princeton, NJ, 2003.

[147] Salikoko Mufwene. *The ecology of language evolution.* Cambridge University Press, 2001.

[148] T. Nagylaki. *Introduction to Theoretical Population Genetics.* Springer-Verlag: Berlin, 1992.

[149] G. Neumann and D. Flickinger. Learning stochastic lexicalized tree grammars from hpsg. Technical report, DFKI, Saarbrucken, 1999.

[150] M. Newman, A. Barabasi, and D. Watts. *The Structure and Dynamics of Complex Networks.* Princeton University Press, 2003.

[151] F. Newmeyer. *Language Form and Language Function.* MIT Press: Bradford Books, 1998.

[152] E. Newport, H. Gleitman, and L. R. Gleitman. Mother, i'd rather do it myself: some effects and non-effects of maternal speech style. In C. E. Snow and C. A. Ferguson, editors, *Talking to Children: Language Input and Acquisition.* Cambridge University Press, 1977.

[153] E. L. Newport and R. N. Aslin. Innately constrained learning: Blending old and new approaches to language acquisition. In S.C. Howell, S.A. Fish, and T. Keith-Lucas, editors, *Proceedings of the 24th Annual Boston University Conference on Language Development.* Cascadilla Press, MA, 2000.

[154] J. Nichols. *Linguistic diversity in space and time.* University of Chicago Press, Chicago IL, 1992.

[155] P. Niyogi. *The Informational Complexity of Learning: Perspectives on Neural Networks and Generative Grammar.* Kluwer Academic Publishers, 1998.

[156] P. Niyogi. The computational study of diachronic linguistics. In D. Lightfoot, editor, *Syntactic Effects of Morphological Change.* Cambridge University Press, 2002.

[157] P. Niyogi. Models of cultural evolution and their application to language change. In Ted Briscoe, editor, *Language Evolution through Language Acquisition.* Cambridge University Press, 2002.

[158] P. Niyogi. Phase transitions in language evolution. In L. Jenkins, editor, *Variation and Universals in Biolinguistics.* Elsevier Press, 2004.

[159] P. Niyogi and R. C. Berwick. Formalizing triggers: A learning model for finite spaces. Technical Report AI Memo-1443, MIT, November 1993.

[160] P. Niyogi and R. C. Berwick. The logical problem of language change. Technical Report AI Memo: 1516, MIT AI Laboratory, 1995.

[161] P. Niyogi and R. C. Berwick. A dynamical systems model of language change. *Complex Systems*, 11:161–204, 1997.

[162] P. Niyogi and R. C. Berwick. Evolutionary consequences of language learning. *Linguistics and Philosophy*, 20:697–719, 1997.

[163] M. Nowak. The evolutionary biology of language. *Philosophical Transactions of the Royal Society of London*, 355:1615–1622, 2000.

[164] M. Nowak, N. Komarova, and P. Niyogi. Evolution of universal grammar. *Science*, 291:114–118, 2001.

[165] M. Nowak, N. Komarova, and P. Niyogi. Computational and evolutionary aspects of language. *Nature*, 417:611–617, 2002.

[166] M. Nowak and K. Sigmund. Chaos and the evolution of cooperation. *Proceedings of the National Academy of Sciences*, 90:5091–5094, 1993.

[167] M. A. Nowak and D. Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14):8028–8033, 1999.

[168] John Ohala. Sound change is drawn from a pool of synchronic variation. In L. E. Breivik and E. H. Jahr, editors, *Language change: Contributions to the study of its causes.* NY: Mouton de Gruyter, 1989.

[169] John Ohala. The phonetics of sound change. In Jones, editor, *Historical Linguistics: Problems and Perspectives*, pages 237–278. Longman, 1993.

[170] M. Oliphant. The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior*, 7:371–384, 1999.

[171] M. Oliphant and J. Batali. Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, 11(1), 1997.

[172] Paulus Orosius. *The Old English Orosius: Janet Bately: editor.* Oxford University Press, 1980.

[173] C. Osgood and T. Sebeok. Psycholinguistics: A survey of theory and research problems. *Journal of Abnormal and Social Psychology*, 49(4):1–203, 1954.

[174] Daniel N. Osherson, M. Stob, and S. Weinstein. *Systems that Learn.* MIT Press, 1986.

[175] Hermann Paul. *Principles of the History of Language (translated from the 1880 German edition by H. A. Strong).* College Park: McGroth Publishing Company, 1970.

[176] M. Piatelli-Palmarini. Evolution, selection, and cognition: From learning to parameter-setting in biology and in the study of language. *Cognition*, 31(1), 1989.

[177] Steven Pinker. *Language learnability and language development.* Harvard University Press, 1984.

[178] Steven Pinker. *The Language Instinct: the new science of language and mind.* William Morrow and Co, NY, 1994.

[179] L. Pitt. Probabilistic inductive inference. *Journal of the ACM*, 36(2):383–433, 1989.

[180] Carl J. Pollard and Ivan A. Sag. *Head-Driven Phrase Structure Grammar.* University of Chicago Press, Chicago, 1994.

[181] Alan Prince and Paul Smolensky. Optimality theory: Constraint interaction in generative grammar. Technical Report TR-2, Rutgers University Cognitive Science Center, 1993.

[182] G. K. Pullum and B. Scholz. Empirical assessment of poverty of stimulus arguments. *Linguistic Review*, 19:9–50, 2002.

[183] T. Regier. Emergent constraints on word-learning: A computational review. *Trends in Cognitive Sciences*, 7:263–268, 2003.

[184] T. Regier, B. Corrigan, R. Cabasan, A. Woodward, M. Gasser, and L. Smith. The emergence of words. In *Proc. of the Cognitive Science Society*, 2001.

[185] Terry Regier. *The Human Semantic Potential: Spatial Language and Constrained Connectionism.* The MIT Press, Cambridge, MA, 1996.

[186] Sidney Resnick. *Adventures in Stochastic Processes.* Birkhauser:Boston, 1992.

[187] J. Rickford. *Dimensions of a Creole Continuum.* Stanford University Press, 1987.

[188] J. Rissanen and E. S. Ristad. Language acquisition in the mdl framework. In E. S. Ristad, editor, *Language Computations.* American Mathematical Society (DIMACS series), 1992.

[189] I. Roberts. *Verbs and Diachronic Syntax: A Comparative History of English and French.* Kluwer Academic Publishers, 1992.

[190] H. Rogers. Godel numberings of partial recursive functions. *Journal of Symbolic Logic*, 23:331–341, 1958.

[191] A. Rubinstein. *Economics and Language.* Cambridge University Press, 2000.

[192] M. Ruhlen. *The Origin of Language: Tracing the evolution of the mother tongue.* John Wiley, 1994.

[193] M. Ruhlen. *On the Origin of Language.* Stanford University Press, 1997.

[194] Jerrold Sadock. *Autolexical Syntax: A Theory of Parallel Grammatical Representations.* University of Chicago Press, 1991.

[195] Y. Sakakibara. Learning context-free grammars from structural data in polynomial time. *Theoretical Computer Science*, 76:223–242, 1990.

[196] W. Sakas. *Ambiguity and the Computational Feasibility of Syntax Acquisition.* PhD thesis, City University of New York, New York, N.Y., 2000.

[197] G. Salvi. La sopravvivenza della legge di wackernagel nei dialettiocci-dentali della peninsola iberica. *Medioevo Romanzo*, 15:177–210, 1990.

[198] Beatrice Santorini. The rate of phrase structure change in the history of yiddish. *Language Variation and Change*, 5:257–283, 1993.

[199] B. Schieffelin and A. Eisenberg. Cultural variation in children's conversations. In R. Schiefelbusch and D. Bricker, editors, *Early language: Acquisition and intervention.* University Park Press, 1981.

[200] Carson T. Schutze. *The empirical base of linguistics: Grammaticality judgements and linguistic methodology.* The University of Chicago Press, 1996.

[201] B. Schwartz and S. Vikner. All verb second clauses are cps. *Working Papers in Scandinavian Syntax*, 38:1–48, 1990.

[202] A. Senghas and M. Coppola. Children creating language: How nicaragual sign language acquired a spatial grammar. *Psychological Science*, 12, 4:323–328, 2001.

[203] Zhongwei Shen. *The dynamic process of sound change.* PhD thesis, University of California at Berkeley, Berkeley, CA, 1993.

[204] Zhongwei Shen. Exploring the dynamic aspect of sound change. *Chinese Journal of Linguistics (monograph)*, 11, 1997.

[205] Y. Shoham and M. Tennenholtz. On the emergence of social conventions: modeling, analysis, and simulations. *Artificial Intelligence Journal*, 94:139–166, 1997.

[206] J. M. Siskind. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61:39–91, 1996.

[207] J.M. Siskind. *Naive Physics, Event Perception, Lexical Semantics, and Language Acquisition.* PhD thesis, M.I.T., Cambridge, MA, 1992.

[208] S.Jain, D. Osherson, J. Royer, and A. Sharma. *Systems that Learn: An Introduction to Learning Theory, 2nd ed.* MIT Press, 1998.

[209] Dan Slobin. *The crosslinguistic study of language acquisition: Vol. I-IV.* Lawrence Erlbaum Associates, NJ, 1985-97.

[210] J.Maynard Smith. *Evolution and the Theory of Games.* Cambridge University Press, 1982.

[211] J.Maynard Smith and E. Szathmary. *The Major Transitions in Evolution.* Oxford University Press, 1995.

[212] W. J. Smith. *The Behavior of Communicating.* Harvard University Press, Cambridge, MA, 1977.

[213] Ray J. Solomonoff. A formal theory of inductive inference. *Information and Control*, 7:1–22, 224–254, March,June 1964.

[214] E. Stabler. Acquiring languages with movement. *Syntax*, 1:72–97, 1998.

[215] L. Steels. Self-organizing vocabularies. In C. Langton, editor, *Proc. of ALIFE V*, 1996.

[216] L. Steels. Social and cultural learning in the evolution of human communication. In K. Plunkett K. Oller, D. Griebel, editor, *Evolution of Communication Systems: A Comparative Approach.* The MIT Press, Cambridge MA, 2001.

[217] L. Steels. Evolving grounded communication for robots. *Trends in Cognitive Science*, 7(7):308–312, 2003.

[218] L. Steels and F. Kaplan. Spontaneous lexicon change. In *Proc. of COLING-ACL: Montreal*, 1998.

[219] L. Steels and P. Vogt. Grounding adaptive language games in robotic agents. In P. Husbands and I. Harvey, editors, *Proc. of the Fourth European Conference on Artificial Life*, 1997.

[220] Kenneth Stevens. *Acoustic Phonetics.* The MIT Press, Cambridge MA, 1998.

[221] S. Strogatz. Exploring complex networks. *Nature*, 410:268–276, 2001.

[222] D. Surendran. The functional load of phonological contrasts. Master's thesis, The University of Chicago, Chicago, IL, 2003.

[223] Henry Sweet. *The Practical Study of Languages.* 1899.

[224] Bruce Tesar and Paul Smolensky. *Learnability in Optimality Theory.* MIT Press, 2000.

[225] F. Tohme and T. Sandholm. Coalition formation processes with belief revision among bounded rational self-interested agents. *Journal of Logic and Computation*, 9(6):793–815, 1999.

[226] R. L. Trask. *Historical Linguistics.* Edward Arnold, 1996.

[227] L. Travis. *Parameters and effects of word order variation.* PhD thesis, MIT, Cambridge, MA, 1984.

[228] William Labov Uriel Weinreich and Marvin Herzog. Empirical foundations for a theory of language change. In W. Lehman and Y. Malkiel, editors, *Directions for historical linguistics*, pages 97–195. University of Texas Press, Austin TX, 1968.

[229] Leslie Valiant. A theory of the learnable. *ACM*, 27(11):1134–1142, 1984.

[230] A. van der Mude and A. Walker. On the inference of stochastic regular grammars. *Information and Control*, 38:310–329, 1978.

[231] Ans van Kemenade and Nigel Vincent. *Parameters of Morphosyntactic Change.* Cambridge University Press, 1997.

[232] V. Vapnik. *Estimation of Dependences Based on Empirical Data.* Springer-Verlag, NY, 1982.

[233] V. Vapnik. *Statistical Learning Theory.* Wiley, NY, 1998.

[234] W. S.-Y. Wang, J.Y. Ke, and J.W. Minett. Computational studies of language evolution. In C. R. Huang and W. Lenders, editors, *Computational Linguistics and Beyond*, pages 65–106. Academia Sinica: Institute of Linguistics, 2004.

[235] William S-Y. Wang. Competing changes as a cause of residue. *Language*, 45:9–25, 1969.

[236] A. Warner. The structure of parametric change and v-movement in the history of english. In A. van Kemenade and N. Vincent, editors, *Parameters of Morphosyntactic Change.* Cambridge University Press, 1997.

[237] D. Watts. *Six degrees: the science of a connected age.* Norton, New York, 2003.

[238] K. Wexler and P. Culicover. *Formal Principles of Language Acquisition.* MIT Press, 1980.

[239] R. M. Wharton. Approximate language identification. *Information and Control*, 26:236–255, 1974.

[240] Patricia M. Wolfe. *Linguistic Change and the Great Vowel Shift in English.* University of California Press, 1972.

[241] Sewall Wright. *Evolution and the Genetics of Populations, Vols 1-4.* University of Chicago Press, 1968-78.

[242] Charles Yang. *Knowledge and Learning in Natural Language.* Oxford University Press, 2002.

[243] Charles D. Yang. *Knowledge and Learning in Natural Language.* PhD thesis, MIT, Cambridge, MA, 2000.

[244] C. Jan-Wouter Zwart. Clitics in dutch: Evidence for the position of infl. *Groninger Arbeiten zur germanistischen Linguistik*, 33:71–92, 1991.