Natural Language From Artificial Life

Simon Kirby
Language Evolution and
    Computation Research Unit
Department of Theoretical and
    Applied Linguistics
University of Edinburgh
Adam Ferguson Bldg.
40 George Square
Edinburgh EH8 9LL
Scotland
simon@ling.ed.ac.uk
http://www.ling.ed.ac.uk/
    ~simon

Abstract   This article aims to show that linguistics, in particular the study of the lexico-syntactic aspects of language, provides fertile ground for artificial life modeling. A survey of the models that have been developed over the last decade and a half is presented to demonstrate that ALife techniques have a lot to offer an explanatory theory of language. It is argued that this is because much of the structure of language is determined by the interaction of three complex adaptive systems: learning, culture, and biological evolution. Computational simulation, informed by theoretical linguistics, is an appropriate response to the challenge of explaining real linguistic data in terms of the processes that underpin human language.

1   Introduction

What can artificial life offer linguistics, and why is language a particularly appropriate target for ALife methods?

In this article I will review a relatively small subset of the work that has emerged over the last 15 years or so at the intersection of linguistics and artificial life. In particular, this article will deal with the models that have shed light on the origins of syntax in human language. This should not be taken as an exhaustive review—for example, I will not be covering the excellent work that has been undertaken on phonetics and phonology (see, for example, [35]), critical periods for language acquisition [55, 59], or language change (14, 29–31, 45, 62, 81, 111). Nor will the extensive literature on models of animal communication be the main target of this review (although some of the work will be discussed in Sect. 3). Instead, I hope that the selection of studies presented here will provide clear answers to the questions above and offer support for my view that individual-based computational modeling will be the core of any future research framework for an explanatory linguistics.[1]

In the next section, I will review some of the important features of language that make it the most complex natural system that we are aware of. The following sections review some of the models in the literature in a roughly chronological fashion (as it turns out, both in terms of publication date and evolutionary age of the target of explanation). The topics covered are, in order, the biological evolution of innate simple signaling systems; cultural evolution of simple signaling systems that are learned rather than innate; the grounding of signals and the modeling of flexible meaning spaces; the emergence

---

[1] At least one other excellent review of the literature on computational modeling of the evolution of language exists [100]. Another source for a more in-depth look at the field is [22]. The specific goal of this article, however, is to look at ALife research from the point of view of *linguistics*.

of complex structured languages both through "negotiation" and cultural evolution; and models that tackle coevolution of languages and language learning mechanisms. Finally, the last section will give a personal viewpoint on what needs to be done and what breakthroughs we are likely to see in the next decade.

2   The Three Complex Dynamical Systems in Linguistics

What's so special about language? Is it really the most difficult problem in science?

So far in its short history, artificial life has mostly concerned itself with the study of natural phenomena that would be considered to be the remit of biology and ethology. For example, in Langton's edited volume surveying the field in 1995 [71], there are models of molecular evolution [95], of cooperation in ecosystems [74], and of morphogenesis [91]. There have also been attempts to use ALife techniques in social science domains [39] and even to model traffic jams [78]. One rationale for the ALife approach in these domains is that they all involve complex synergistic interactions that, when taken together, lead to emergent behaviors that are difficult to predict. In complex dynamical systems, verbal theorizing often leads to incorrect predictions because our intuitions about the links between local interactions and global behavior are notoriously unreliable. Furthermore, the classical alternative—analytic mathematical modeling—may require the kinds of idealization that will necessitate the removal of the very network of interactions that give rise to the target of explanation.

ALife provides the answer to the methodological problem of linking theory and prediction for complex dynamical systems involving multiple interacting components. Microsimulation models, carefully designed, give the theorist a tool for testing ideas about how natural phenomena such as bird flocks, cell division, and so on, may come about. In this view of the value of artificial life, human language is an ideal topic for exploration using ALife models.

Consider a few of the things we know about human language:

1. It is a nontrivial communication system in that it allows us the potential for producing a signal for a range of meanings with infinite extent. In other words, it is likely that the majority of communicative events that we engage in every day involve the production of signals that the hearer has never encountered before, and yet we have a reasonable expectation of being understood.

2. Language is symbolic. That is, the signal associated with a particular meaning is likely to be arbitrary with respect to the system of convention that set up that association.

3. No one is born with the ability to communicate using language.

4. Almost every member of our species can be a master at this system by around 6 years old without being taught (in the sense of deliberate instruction), and without being given any reliable feedback about their communicative performance [17].

5. Our ability to learn a language at all tails off as we get older. If we have not learned a language by puberty we probably never will [72].

6. Our eventual ability to communicate in the language of the community into which we are born appears to have very little to do with who our biological parents are;

nor does it appear to be predictable from any other measure of individual difference.

7. Nearly all languages appear to be equally functional with respect to communication.

8. There is no unequivocal evidence of any other species in possession of a communication system even close to the complexity of human language [50].

9. The most able nonhuman at human language is a pygmy chimpanzee named Kanzi, who was not taught language directly and was raised in a mixed human–chimpanzee environment with a lot of functional communicative interaction [94].

10. There are around 6,000 different languages on Earth. Many of these show discernible similarities due to a common history. Others show similarities that cannot be explained in this way. There is a great deal of diversity in language types, yet it is clear that the logically possible range of variation is not reflected cross-linguistically [32].

11. Languages change over a relatively short historical time scale. No language is completely static.

12. With oral tradition, written texts, and the enabling of direct instruction and description, language allows us to have the most complex cultural transmission system of the natural world, enabling the development of technologies that dwarf in complexity those of our closest biological species.

13. Despite its complexity, humans can manage fairly well without fully syntactic language. Many people are able to cope with little disadvantage in communities that employ a language they are not proficient in, for example.

14. In some cases, languages can emerge in a very short space of time (e.g., one generation) in communities of people who do not have any shared language (or indeed any languages at all) [7, 93].

15. Many aspects of language appear to be localized in particular areas of the cortex and may be damaged by, for example, stroke.

16. Formally, the natural language syntax appears to be at least context-free, and there are constructions in some languages that seem to require context-sensitive descriptions.

17. Context-free languages are not learnable in the limit in the most general case [41]. However, if learners have more information available to them than simply samples of the set of grammatical strings of the language the constraint is weakened [43].

18. Native speakers are still able to make "grammaticality judgments" about sentences that they will never hear, use, and may find extremely difficult to understand.

19. The principles that underlie *some* of the patterns of grammaticality within and across languages do not have any clear parallel in any nonlinguistic human behavior.

20. Although much of the structure of human language appears to be adapted to the task of communication between humans, some properties of syntax appear to be positively dysfunctional [73].

To come up with a truly *explanatory* linguistics, we would need to (a) understand the origins of this unique behavior, and (b) derive the properties of this behavior from the properties of nonlinguistic systems.[2] It turns out that this is a very challenging task. Part of the reason for this is clear if we look, for example, at what processes give rise to the object we call "English."

Firstly, we must understand that the behaviors that make up most of English are the mass of short-lived utterances that are being exchanged all the time around the globe every day. These utterances consist of a stream of sound waves (which we can analyze as a sequence of phonemes—minimal contrastive units of a language's sound system) paired with a communicative intention. In this view, English is a fundamentally dynamic object. Furthermore, we can view English as a property of the particular wiring and state of the brains of the millions of adult English speakers in the world. These two views of English can be termed the E-language and I-language views, respectively.

Not only is English at any one point in time intrinsically dynamic (being made up as it is of moving air, and firing neurons), the manner in which it persists over time means that it is liable to change in radical ways. The brain-state that is English can only arise through observation by children of the utterances produced by others. Obviously, this process is not spontaneous: Throughout this process children's language changes enormously.

Through this language learning process, E-language is transformed into I-language. When language learners produce utterances, the transformation is reversed and the process of transmission is repeated. As English is being transmitted on a historical time scale, its population is changing: New members are born, old members die, birth rates change, groups migrate, other language users come into contact with English users, and so on.

In addition to language learning on an ontogenetic time scale, and language evolution on a glossogenetic (i.e., historical) time scale, we need to consider the manner in which the brains of language users are shaped by biological evolution on a phylogenetic time scale. We have the brains we do by virtue of a combination of our genetic endowment (which specifies the ways in which our brains will grow in response to the environment) and the environment in which we find ourselves (which interacts with our genetic program to determine the final state of our brains). As the brain was evolving, the genes that controlled its development were under selection pressure. It is possible that this pressure changed as the very behaviors that the brain supported emerged.

To summarize, languages are learned by observing others. The only way language can persist is for it to be learnable; this leads to languages evolving culturally as they are passed from learner to learner. The languages that result from this cultural evolution process result in a change in the selection pressures on the genes that specify the learning mechanisms for language. This means that the properties of the learning mechanism may change, ultimately affecting how languages are learned (Figure 1).

English, therefore, is the result of a staggeringly complex interaction of (at least) three complex adaptive (dynamical) systems. It (and all the other languages of the world) are therefore perfect topics for the next generation of ALife models. In the next

---

2 It should be noted that this is not the only view of what counts as an explanatory linguistics. Chomsky notably treats the goal of coming up with an elegant framework of linguistic descriptions that can be generalized to predict the grammaticality of unseen utterances and that can be employed for the description of any language as an explanatory goal. In [26], Chomsky makes a comparison to the discovery of the periodic table. The periodic table was an elegant description that led to developments in physical theory to account for its existence. In a sense, physics was required to explain the periodic table. Contrast this with the alternative possibility: that the periodic table had been derived from physical principles and then had been compared with chemical data.

In a sense, Chomskyan explanation can be thought of as running in reverse to the way I am expounding. As for the origins of language, Chomsky has suggested that not enough is known to make this question worth pursuing.
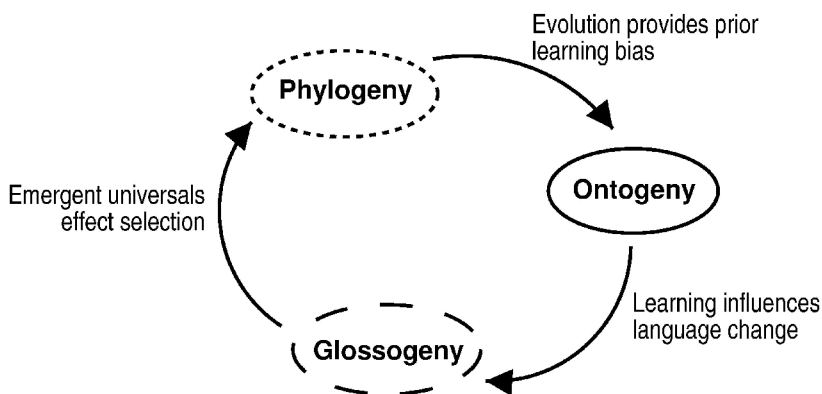
Figure 1. Language is the result of an interaction between three complex adaptive systems that operate on different time scales: the time scale of biological evolution (phylogeny), the time scale of individual learning (ontogeny), and the time scale of language change (glossogeny).

few sections, I will summarize a line of research that leads from simple animal signaling systems to the coevolution of languages and the language acquisition device to show the sort of work that can be done.

3   Innate Signaling Systems

What are the simplest systems of communication, and what does ALife tell us about their evolution?

The vast majority of communication that we are aware of is the result of purely innately coded behaviors, that is, systems that involve no learning whatsoever. Examples of this sort of communication can be found all through the natural world. Systems as diverse as the attraction of insects to particular flowers, bee dances, and vervet monkey alarm calls are communicative systems that are solely the result of evolution by natural selection. Oliphant [85] reviews the work carried out within the framework of artificial life on innate communication and draws out the main themes in this research within his own idealized simulation model. In this section, I will summarize his viewpoint.

Before we look in general at what an artificial life approach to innate communication might look like, we should define what constitutes communication. This is itself an area of some debate. In particular, there is a tension between definitions of communication employed by philosophers concerned primarily with language (e.g., [36]) and ethologists looking at animal communication (e.g., [110]). Oliphant, following the non-computational framework of Krebs and Dawkins [70], suggests that it is useful to look at communication as a simultaneous combination of *exploitation* and *manipulation*:

An act of communication is a causal chain of events, whereby one individual, the sender, exhibits a behavior in response to a particular situation, and a second individual, the receiver, responds to this behavior. Such an interaction is communicative if it involves manipulation on the part of the sender and exploitation on the part of the receiver. ([85, p. 14])

Exploitation in this context refers to cases where the behavior of an individual is determined to some extent by the fact that another individual can be expected to behave in a particular way in a particular situation, and that behavior has been observed. If

I were to avoid entering a colleague's office because I heard her sneezing, this could be considered an exploitative interaction. The sneezing is a behavior I recognize is expected if my colleague has a cold—my decision to use this signal to avoid her is based on this relationship between a sneeze and having a cold.

Manipulation, on the other hand, refers to cases where the behavior of an individual is determined to some extent by the expectation that another individual will respond in a particular way to that behavior. We can think of the English expression "boo!" as a manipulative signal as it is sometimes used to make people jump in fright. In this case, the person shouting "boo!" is manipulating the fact that a sudden noise can trigger a flight response in the hearer.

It should be clear that the cases of communication that we are familiar with all fit within Oliphant's definition. The fact that a flower is a particular color arises out of both manipulation of the insects it is attracting *and* exploitation by the insects of this signal. Similarly, the bee's dance is both a manipulation of the response of its fellow bees (to fly in a particular direction) and an exploitation by the hive of the arriving bee's dance. In both these cases, it is evolution by natural selection that gives rise to this adaptive circle of manipulation and exploitation via signaling. In other words, the two behaviors are the result of genetically hardwired responses to situations.

Several artificial life models have been proposed to explore what is required for such a system to emerge (e.g., [1, 18, 37, 75, 82, 109]). In the work of MacLennan and Burghardt [75], for example, each agent had access to a local environment, which was in a random state, and a global environment that was accessible for each agent to view and also alter. The global environment thus becomes a conduit for communication about each agent's local environment. After a "speaking" agent places a symbol in the global environment, it is given a fitness reward if the next agent to act gains knowledge about the speaker's local environment. In addition, the "hearer" also gains a fitness reward for this successful transfer of information. Ackley and Littman [1] use a similar framework where agents communicate about hidden parts of their local environment, but they break the symmetry of fitness payoffs by not rewarding speakers. They show that communication can also evolve in this setting under certain conditions that we will return to later.

Rather than model specific ecological conditions, Oliphant [85] sets out a generalized approach to innate signaling systems.[3] We assume a set S of low-cost behaviors (signals) that are observable and can be distinguished by others.[4] There is also a set M made up of pairs of environmental states and appropriate responses to those states (meanings). An agent's communication system is described in terms of two probability functions: $s(\mu \in M, \sigma \in S)$, the transmission function, and $r(\sigma \in S, \mu \in M)$, the reception function. $s(\mu, \sigma)$ for a given meaning and signal gives the probability that the agent will produce the signal σ for the meaning μ. Conversely, $r(\sigma, \mu)$ gives the probability that the signal σ will be interpreted as the meaning μ by the agent.

When is a communication system useful? To answer this question, we need a measure of communicative accuracy for a population of agents. Firstly, Oliphant calculates the expected probability that a transmitting agent with transmission function, s, will be understood by a receiving agent with reception function, r. This is the average of the product of signaling and receiving probabilities for each possible signal-meaning pair:

$$ca(s, r) = \frac{1}{|M|} \sum_{\mu \in M} \sum_{\sigma \in S} s(\mu, \sigma)\, r(\sigma, \mu)$$

---

3 Oliphant is not unique in taking this approach. Others [18, 37, 82] also use generalized models of innate signaling to explore the conditions under which innate communication evolves.

4 See, for example, [83] for discussion of what difference it makes if we change assumptions about the cost of signaling.

S and R are defined to be the average probability functions for a population of n agents:

$$S(\mu, \sigma) = \frac{1}{n} \sum_{i=1}^{n} s_i(\mu, \sigma)$$

$$R(\sigma, \mu) = \frac{1}{n} \sum_{i=1}^{n} r_i(\sigma, \mu)$$

So, the average communicative accuracy of a population is simply $ca(S, R)$. The op-
timally communicating population would have $ca(S, R) = 1.0$ since, in this case, for
every meaning and every pair of agents a signal would be sent that would be interpreted
correctly as that meaning.

   To look at the evolution of communication Oliphant encodes agents' transmission
and reception functions as a string of genes. Under the simplification that s and r can
only return 0.0 or 1.0, the genome of an agent is an ordered string of signals that the
agent can produce (one for each meaning), followed by a string of meanings (one for
each signal the agent might perceive). For example, if an agent produced the signal a
for meaning 1 and the signal b for meaning 0, but interpreted both signals as meaning
0, then this agent's genome would be $ba00$. [Incidentally, a population made up of
clones of this agent would have $ca(S, R) = 0.5$.]

   A typical evolutionary simulation using this genetic encoding scheme would involve
a population of agents with random genomes being repeatedly paired off in communi-
cation "games." In each game a random meaning would be chosen, one of the agents
would produce the signal associated with that meaning (according to its genetically en-
coded transmission function), and the other agent would interpret that signal according
to its reception function. Two outcomes are possible: Either the original meaning and
the interpreted meaning are the same or they are different. The results of these games
would be used in some way in the simulation to assign a fitness score to each agent.
This fitness score would then be used to select the genetic make-up of the following
generation.

   The key question is now: Under what specific simulation parameters will $ca(S, R)$
end up equaling 1? There are two key ways in which the general simulation model can
be parameterized:

1. How is fitness calculated in response to success or failure in the communication
   game?

2. How are agents selected to become communicative partners?

   With regard to the former, Noble [83] divides the fitness calculation into two compo-
nents: P_S and P_R. These correspond to the fitness effect on the sender, and the fitness
effect on the receiver. In addition, Noble varies the costs of signaling and responding
to signaling: C_S and C_R. These costs can be varied in a limited way by the agents (e.g.,
an agent may choose to make a "loud" or "quiet" signal, incurring different costs).
Noble's model has many parameters that he varies systematically in a range of evo-
lutionary simulations. Rather than reproduce his results here, I will follow Oliphant's
simpler parameterization of the space of possible ecological scenarios by looking at
two alternatives:

   Mutual benefit. Both speaker and hearer benefit from successful communication.

   Altruistic behavior. Only the hearer benefits from successful communication.

From this perspective, the simulation of Ackley and Littman [1] is different from the other two described earlier in that they modeled an altruistic situation as opposed to a mutualistic one.

Using this type of framework, Oliphant concludes that it is relatively easy for evolution to tune an innate communication system in situations of mutual benefit. However, in settings where speakers do not benefit from communication, the evolutionary simulation is not guaranteed to succeed. These results are confirmed by Noble [83]. To understand when a perfect communicative accuracy score *can* be achieved in these situations, the choice of game partners needs to be examined:

> *Random.* Communication games take place between randomly picked members of the population.

> *Spatial organization.* The population has *spatial structure*, and communicative partners are chosen that are close together.

In the generalized model with random game-partner choice, altruistic communication does not emerge. However, with a spatially organized population, Oliphant shows that an optimal $ca(S, R)$ can be achieved even in cases where the speaker does not benefit (as in the case of [1]). He argues that this is because spatial organization is equivalent to *kin selection*, a mechanism known to promote the evolution of altruistic behavior [44]. This is because when you are more likely to talk to your neighbors, you are more likely to talk to your kin (under the assumption that the children of agents are located close to each other). Although this parallel seems sensible, DiPaolo [38] points out that Oliphant does not actually formally test the equivalence of spatial organization and kin selection. He also shows that, in some circumstances, there are alternative mechanisms that are not related to kin selection by which spatial organization can promote the evolution of optimal communication.

The use of artificial life modeling techniques for studying the evolution of innate communication is well established. Most researchers would probably agree that evolution by natural selection can tune a simple communication system under reasonable ecological assumptions. In many ways it seems that any further models of innate communication would be unnecessary. However, we must be very careful not to let the attractive simplicity of the general model described here blind us to other issues that deserve study. Throughout this section I have used terms like "tune" to describe the work that evolution is doing in this model. We have shown the conditions under which evolution can select optimal mappings between signals and meanings. However, models such as this assume that there already exist agents who are naturally inclined to map signals and meanings at all. To put it another way, these agents have given to them on a plate a dedicated channel through which they send information, and cognitive systems that naturally pair this information with meanings.

Can artificial life techniques help us to understand the origins of the communication channel itself? Recent work by Quinn [92] suggests that they can. In Quinn's model, mobile agents are placed in a featureless environment in pairs. Each agent has a set of rough sensors that help it "see" the other agent if it is close (the agents are actually Khepera robots). To simplify somewhat, the agents evolve according to an evolutionary scheme that rewards the distance that the center point between the two agents moves. What makes this evolutionary task interesting is that the optimal strategy requires coordination between the two agents. Optimally, both agents should move in the same direction to maximize their joint displacement from the starting condition.

Notice that, in this model, there is no dedicated communication channel. All the agents have as potential output is their own movement, and all they have as their

potential input are readings from their proximity sensors. Neither are there any obvious "meanings" in the simulation. However, communication evolves in Quinn's simulation. The typical behavior of two evolved agents is as follows:

1. Each agent rotates anti-clockwise until it faces the other.

2. The first one to face the other moves forward to close the range between them.

3. Once it is in range, it starts moving backward and forward a small amount, while staying close to the agent it has approached.

4. This continues until the other agent also becomes aligned.

5. At this point, the agent that got aligned last starts reversing, and the other follows.

In summary, then, the agents in Quinn's model have an innate signaling system consisting of one signal—the oscillating movement when close to the other agent—which we might gloss in English as "after you."

   This section has looked at the ways in which artificial life can illuminate questions relating to the evolution of simple innate signaling—the sort of communication that is common in the natural world. It is striking, however, that human language is predominantly very *unlike* one of these communication systems (even a cursory glance back at the list of features given in Sect. 2 should convince us of that). The next section turns to one of the most striking differences between human language and most other communication systems: learning.

4   Culturally Evolved Signaling and the Iterated Learning Framework

How can we use artificial life simulations to uncover the requirements for a learned communication system?

At the start of this article, I argued that one of the most interesting things about human language is the fact that it can be understood in terms of an interaction of complex adaptive systems. The work covered in the previous section only makes reference to one of these systems—adaptation through natural selection—since for innate simple communication, this provides the only relevant dynamic. As such, these studies are of only limited interest to those wishing to understand the origins of language.

   Fundamentally, in human languages the mapping between signals and meanings is learned rather than being determined genetically.[5] An important question at this point is whether this actually changes anything—in other words, perhaps we can use the results from studies of innate communication systems and somehow apply them directly to their learned analogue. There are some indications that this could be a sensible strategy. There have been attempts to model language acquisition and language change (over a historical time scale) using genetic algorithms [30, 56]. Biological evolution is simply optimizing the meaning–signal mapping to a communicative function (under the assumption that communication offers a selective advantage), and surely individual learning is performing a functionally identical optimization?

   I wish to argue here that this simplistic approach is unwise, and that, at least for the most relevant types of learning, we cannot carry over what we have learned from looking at innate communication to learned communication in any trivial way. For a start, a learned communication system is subject to two dynamical processes acting

---

5 I am not saying here that *only* human languages involve learned symbolic mappings (although this has been claimed by Oliphant [87]). More study is needed to understand exactly what learning mechanisms are used in other species' communication systems.

together (as pointed out in Sect. 2). Not only does the system develop over the lifetime of the agent in response to experience, but the experience that the agent learns from is itself the output of other agents' learned behavior. If we are to look for equivalences between innate and learned systems, how does the process of biological evolution map onto the twin processes of ontogenetic and glossogenetic (i.e., cultural) evolution?

Another problem concerns *adaptation*. The biological evolution of innate communication is a process of adaptation to a particular function: reproductive fitness (and therefore, indirectly, communicative success). We cannot take it for granted that either learning or cultural evolution are adaptive mechanisms that seek optimal solutions with regard to communication, however intuitively appealing that may appear. It is certainly true that a learning scenario could be envisaged that would appear to do just that. For example, if a communication system was learned by an agent using a form of reinforcement or error-driven learning in which the reinforcement or error signal was the result of an attempt to communicate, then we could argue that the behavior of learning agents might reflect that of evolving agents. Unfortunately, there is no evidence that human language learning proceeds with any reliable feedback on performance, let alone an error signal relating to communicative success [17].

Turning to cultural evolution, there have been many attempts to treat this and biological evolution as specific instantiations of the same mechanism (e.g., [33]) ever since Darwin mentioned the parallels in *The Descent of Man* [34]. Furthermore, the functionalist approach to linguistics directly builds in notions of adaptation to communication into linguistic analysis, and the mechanisms behind this have been related to cultural evolution (see [63] for discussion). A problem with this approach based on the analogy between language change and biological evolution is that it is actually rather weak. In fact, as Kirby and Hurford [68] point out, language change as an evolutionary process violates the central dogma of molecular biology (that there is no reverse translation from soma to germ line). Rather, there is a repeated reverse translation from E-language to I-language via induction in linguistic transmission.

It seems sensible, therefore, to study learned communication using a different methodology than that used for innate communication. This does not mean we have to throw everything away that has been developed previously. For example, we will use the same formalism as discussed in the previous section for simple signaling systems. The transmission and reception functions $s(\mu, \sigma)$ and $r(\sigma, \mu)$ for an individual agent are, however, not directly provided by the genome of that agent. Rather, the probabilities returned by these functions are determined by the experience of that agent and a learning algorithm. This does not deny a role for biological evolution, since the particulars of the learning algorithm must come from somewhere—this interaction between learning and evolution is discussed in Section 7.

The interesting feature of learned communication systems is that the introduction of learning leads naturally to another complex system emerging: culture. As long as there is the potential for a learning agent to be "younger" than a speaking agent (i.e., to have been an active member of the population for a shorter time) then there will be a *historical* process of information transmission through repeated learning and performing of behaviors. Notice that this does presuppose the idea that there will be population dynamics in the model. This does not need to be the case. For example, the work of Batali tends to use static populations of learners [3, 4]. To discriminate between these kinds of models and ones in which there is population replacement, Batali uses the term *social coordination* rather than *cultural transmission/evolution*. Section 6 deals with this model in more detail.

Culture, then, relies on the input to a learning agent being the output of similar learning agents. Although an enormous amount of work has been done looking at modeling learning, rather surprisingly there is very little looking at this type of *iter-*

ated learning (IL). Typically, machine learning looks at cases where the data to be learned from comes from "outside" the system in some sense (e.g., it is given by the experimenter or the environment). This is unfortunate, since iterated learning has some fascinating properties (as we shall see later) and is an excellent way of thinking about human language.

To facilitate analysis of models of cultural transmission, Kirby and Brighton [12, 13, 65, 68] set out a generalized *iterated learning framework*. The framework consists of four components:

1. A meaning space

2. A signal space

3. One or more learning agents

4. One or more adult agents

In this framework, adult agents are given random sets of meanings for which they must produce signals. Learning agents sample the resulting meaning–signal pairs and use this to build a model of the population behavior. At some point (typically once a particular number of meaning–signal pairs has been sampled by a learner), learners will become adults. In this framework, it is common to maintain a fixed number of adults and learners, so the maturing of a learner will entail removal of an adult (modeling death) and the introduction of a new learner (modeling birth). Importantly, in most instances of an IL model, each learner directly after birth will be identical throughout the simulation. In other words, there is no biological variation or heredity in the simulation. Also, typically, the probability of an agent dying is the same for all agents, meaning there is no equivalent of natural selection.

There has been rather less work in the artificial life field on learned simple communication, perhaps because it is not clear where this type of behavior is found in the natural world (since most, if not all, nonhuman signaling systems are innate). However, see [54] and [60] for early examples.

One of the key features that differentiates the various artificial life simulations that broadly fall within the IL framework is the type of learning they use. As I have already argued, we should prefer IL models that do not rely on task feedback or an error signal passing between agents and should use instead some form of *observational learning* (using solely the sample of meaning–form pairs that adults produce to induce a representation of the communication system). Oliphant [86] sets out a taxonomy of observational learning strategies based on how they change the language being transmitted by IL:

Acquisition. A learning strategy is said to be capable of acquisition if it is able to learn the system of an optimally communicating population.

Maintenance. A learning strategy is capable of maintenance if it is able to learn the system of an optimally communicating population even in the presence of noise.

Construction. A learning mechanism is capable of construction if it *fails* to learn the system of a suboptimally communicating population and moreover if a population of constructors will generate an optimal communication system from randomness.

The distinction between acquisition, maintenance, and construction underlines the importance of looking at IL. On first looking at modeling learning organisms, ALife researchers might turn to the machine learning literature to find a model of learning to

use. However, they are likely to choose an algorithm that is capable of *acquisition* but not necessarily maintenance or construction. An agent only capable of acquisition is never going to be a good model of an organism with a learned communication system because the researcher would have to remove all noise from the system and also start the simulation with an optimal communication in place. Of far more interest are those learning strategies that count as constructors. In machine learning terms, these will be those that have a prior bias that results in generalizations that map unique signals to each meaning.

By far the most thorough study of what is needed to be a constructor is one by K. Smith [98]. Smith uses simple associative networks to model an agent's representation of the population's mapping from meanings to signals. These networks act as both reception and transmission function for the agents.[6] In this form of associative network, meanings and signals are given localist representations (i.e., in the meaning and signal layers, only one node is given an activation of 1, and all the others are given an activation of 0). The two layers of the network are fully connected with each other, and the weights are all integers. To retrieve an association, either a signal or meaning node is activated and the activation is multiplied through the weights. A localist representation is calculated by a winner-take-all scheme, and the resulting activation is thresholded to 1.

Learning is implemented by adjusting the weights on the connections between each meaning node and signal node. There are four possible input–output pairs for a given node:

1. Both meaning and signal activated

2. Only meaning activated

3. Only signal activated

4. Neither activated

Smith constrains learning rules for each of these cases to either increase the weight by 1, decrease it by 1, or leave the weight alone. This gives $3^4 = 81$ possible learning rules in total. These can be represented as a four-long vector:

$$\left\langle \left\{ \begin{matrix} 0 \\ 1 \\ -1 \end{matrix} \right\}, \left\{ \begin{matrix} 0 \\ 1 \\ -1 \end{matrix} \right\}, \left\{ \begin{matrix} 0 \\ 1 \\ -1 \end{matrix} \right\}, \left\{ \begin{matrix} 0 \\ 1 \\ -1 \end{matrix} \right\} \right\rangle$$

where the positions in the vector correspond to the input–output pairs given above.

By putting a homogeneous population of agents with a particular learning rule in an IL framework with a random initial language, Smith categorizes all of the possible associative network learning rules in terms of Oliphant's taxonomy. Fifty of the 81 total rules failed to acquire an optimal system at all (i.e., they could be considered nonlearners). The remaining 31 were capable of acquisition; of these 18 could also maintain an optimal system against noise, and 9 of these were constructors. Interestingly, one of the constructors, $\langle 1, -1, -1, 0 \rangle$, can be considered to be a form of Hebbian learning [52]. In a further study, Smith [97] also models a heterogeneous population of the

---

6 Actually, it is common in IL models for reception to play a very small role. This is because the focus of the model is on how the information inherent in the communication system is passed on culturally. The only relevant mechanisms are transmission and learning. Modeling reception is still important, however, because without it the researcher cannot calculate the $ca(S, R)$ score for the population.

various learning rules that is under biological selection pressure. This is an example of a *coevolutionary* model,[7] which are discussed in the context of syntactic systems in Section 7. The main conclusion we can reach from Smith's [97] research is that the interactions between learning, culture, and evolution are nontrivial. In particular, genetic drift plays a major role in ensuring that enough constructors dominate the population at some point to "jump start" a process of cultural convergence, and then biological selection.

Other authors have used different techniques to see which learning algorithms are constructors. For instance, a number of papers follow up Hurford's [54] work in noting that construction is facilitated by learning algorithms that induce $r(\sigma, \mu)$ through observation, and then use some form of inversion technique to infer $s(\mu, \sigma)$. This general strategy has been termed *obverter* by Oliphant and Batali [88]. Obverter can be thought of as a *Bayesian* strategy for communication that is trying to maximize the probability of success at getting a meaning across to a hearer on the assumption that that hearer is (a) like you, and (b) has been exposed to similar data as you. The particular way obverters invert their reception function is by transmitting, for each $\mu \in M$ the $\sigma \in S$ that maximizes $r(\sigma, \mu)$. In other words, obverters produce the signal that, if they hear it, they would most likely understand to be the correct meaning. In this way, obverter agents are using their own reception function to model the reception functions of the rest of the population. Oliphant and Batali [88] prove the optimality of obverter for simple communication systems.

Throughout this section we have been looking at observational learning where agents learn by inducing a model of population behavior from a sample of meaning–signal pairs. Critics of this approach quite rightly point out that the ready availability of signals with meanings neatly attached to them reduces the credibility of any results derived from these models (we will return to this point later in this article). Interestingly, there are alternatives to giving meanings to agents on a plate. The proponents of task-oriented feedback as opposed to pure observational learning (e.g., Steels, Kaplan, McIntyre, and Van Looveren [101, 103]) do not in fact use signal–meaning pairs to train their agents. In a sense, they are moving beyond looking at the evolution of the mapping between signals and meanings into looking at the origins of meanings themselves.[8] It is to these results we now turn.

5   Grounding and the Origin of Meanings

> *What should ALife agents talk about? Can the semantics of a language be learned and evolve as well as its lexicon?*

In a number of publications, Harnad (e.g., [46]) has focused on what he calls the "grounding problem."

> Suppose you had to learn Chinese as a first language and the only source of information you had was a Chinese/Chinese dictionary! This is more like the actual task faced by a purely symbolic model of the mind: How can you ever get off the symbol/symbol merry-go-round? How is symbol meaning to be grounded in something other than just more meaningless symbols? This is the symbol grounding problem. [46]

---

7 In fact, the model can be seen as a combination of models of learning from [98] and cultural transmission from [99].

8 We have already seen an approach that does not assume a predefined set of meanings (or indeed, signals) in the work of Quinn [92]. However, his approach currently only works for evolving innate communication systems.

Any simulation that simply treats meanings as symbols that can be handed over to agents along with their associated signal is not providing any solution to the grounding problem.[9]

In response to this, some researchers stress the importance of placing agents in a rich environment so that their communication is about something. There are essentially two approaches to this (although the boundaries are somewhat blurred) depending on whether the environments are simulated or real. In the latter case, communicating agents are robots (e.g., [9, 104, 108]); in the former they are similar to the software agents we have already looked at, but usually with a number of ways of interacting with their environment [20].

Cangelosi and Parisi [21], in part as a response to Harnad's point about grounding, describe a model in which an innate communication system evolves in a population of neural networks that forage for mushrooms. The connection weights in their networks are set by evolution using a genetic algorithm with a fitness function based on the energy of the agents. An agent's energy depends on the mushrooms that agent eats. Some of the mushrooms are edible and some are poisonous, and the mushrooms are distinguished by the perceptual properties they present to the network, with all edible mushrooms sharing some perceptual similarities and all poisonous ones also being perceived in similar ways to each other.

One part of the output of the network controls the movement of the agents, and after evolution, Cangelosi and Parisi found that the agents were able to recognize the two types of mushroom and avoid eating those that were poisonous. This task, however, does not involve any communication. To set up a situation in which communication was relevant, the authors ran a simulation in which the perceptual properties of the mushroom were not necessarily available to the forager (specifically when the mushroom was far away). However, walking along with the forager is another conspecific (randomly chosen from the population) that can always perceive the mushroom. This "guide" does not control its own movement (its movement output is ignored), but instead, the other output units are given to the forager as input. In other words, the guide produces a signal that the forager has access to.

In this evolutionary scenario, an innate communication system evolves. Whenever an agent perceives a mushroom in the poisonous class it emits a particular signal, and it will emit a different signal whenever it perceives a mushroom in the edible class. Furthermore, whenever an agent receives a signal it will act appropriately.

This simulation is interesting mainly because the meanings in the simulation are arguably grounded in the perceptions of the agents. Certainly, the categories "poisonous" and "edible" are not provided in advance for the agents, but instead emerge because of their ecological relevance.[10] That said, the perceptual system of the agents is fairly rigid, and the meanings that emerge are the only ones that are conceivable within the simple environment. On top of this, the communication system is purely innate (although later work by Cangelosi [19] adds a learning component). Could a communication system emerge without biological adaptation that is grounded in an environment complex enough that many conceivable meanings are possible?

---

9 Not that this is necessarily the wrong thing to do, of course. It is important, in building artificial life models, that we do not try to solve everything at once. A sensible ALife methodology is one in which the computational model is merely an instantiation of a theory about the domain of enquiry (to assist the theoretician in generating predictions to test). If the particular theory being tested (e.g., that learning biases affect emergent languages in iterated learning) does not mention grounding, then the model need not solve the grounding problem.

10 The alert reader may also notice that Cangelosi and Parisi's simulation also demonstrates the emergence of communication in the absence of benefit for the sender, kin-selection, or a spatially organized population. Oliphant [85] notes this interesting feature of the simulation and argues (as Cangelosi and Parisi themselves note) that the sharing of the hidden-unit representational space by both the signal-emitting and movement-generating output nodes favors the emergence of an optimal system in this case.
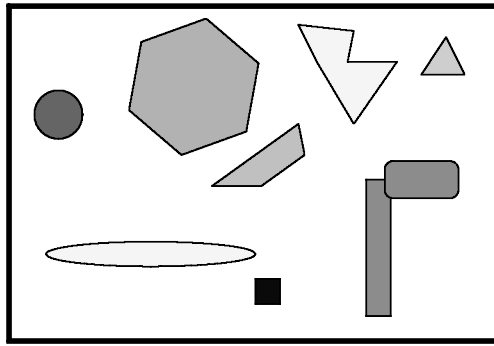
Figure 2. A schematic of a typical scene for the talking heads experiment. Note that the real scene would be subject to environmental conditions such as reflections, variation in light, and so forth.

Luc Steels and his colleagues have attempted to answer this question by using robots to embed communicative agents in the real world. They define various games that can be played between pairs of robots with respect to the world they are able to sense: the discrimination game, the guessing game, the naming game, and so on. For example, in the "talking heads" experiment [101], two robot-controlled cameras try to communicate about a scene in front of them. The scene consists of a set of colored shapes on a white board (see Figure 2). This scene exists in the real world and is thus subject to a lot of variability and noise such as glare from lights, changes in ambient luminosity, and so on.

Each robot camera contains a software agent (which, in fact, only remains resident in that camera for a short time–this means many more agents than cameras can be simulated). The agent has various dimensions (e.g., vertical position, horizontal position, size, color, etc.) along which it may make discriminations among the objects in the scene. It is able to refine internal structures (discrimination trees) in such a way as to represent particular objects in the scene along these dimensions. For example, an agent might think of the square in Figure 2 as the "object on the bottom" or "the small object" or "the square object" or even "the small square object at the bottom." Which way an agent will actually represent the square depends entirely on that agent's life experience, that is, the particular discriminations that it has refined over time.

In a guessing game between two agents, one is picked as the speaker and the other the hearer. After low-level visual segmentation, a *topic* is chosen at random from the shapes on the board (the *context*). The speaker then attempts to find a representation for that object using its discrimination trees. If it is successful, then this representation (a meaning) is looked up in the agent's lexicon of meaning–signal pairs. A relevant signal is produced and passed to the hearer.

Now, the hearer attempts to *point* at the shape that it thinks the speaker is trying to name. Pointing is carried out by sending information about where the camera is looking. The speaker is then able to send a *success* or *failure* signal to the hearer along with (if necessary) a further hint at the correct object by pointing itself. If the game is unsuccessful, both agents make changes to their internal representations to improve the chances of future successful communication. This can involve refining their discrimination trees and/or refining their lexicons.

It may seem to be giving the agents far too much to allow them to tell each other if they have been right or wrong with a guess, and even allow them to point to the correct object when there is failure. However, it is important to realize that by not predefining a meaning space, Steels has increased the complexity of the coordination problem for the agents enormously. This is because it is not only the meaning–signal *mapping* that is evolving in the agents, but also the set of distinctions that the community is making.

As an illustrative example, consider the square in Figure 2 again. It is quite possible for two agents to communicate perfectly about this topic but have quite different lexicons. For example, let's say a speaker had just sent the signal "mulipa." The hearer could pair "mulipa" with a node in a discrimination tree corresponding to *bottom*, but the speaker could actually have a node in a discrimination tree corresponding to *black* that was paired with the same word. The hearer might not even make any distinctions based on color, let alone have the color *black* paired with this word, but despite this, communication is successful. Eventually, however, it is likely to fail (see [96] for further analysis).

In [102] the results of the talking heads experiment are discussed in terms of *semiotic dynamics*. The authors' point is that, by allowing agents to create their own meaning spaces, they have introduced a new dynamic into the cultural evolution of languages that is not present in other simulations. The results of the experiments show that there are rich patterns of variation, both within the population and over time. New meanings can always emerge, and multiple meanings appear to compete with each other. Importantly, this apparently chaotic dynamic does not mean that the agents are not capable of communicating with each other.

At this point, the sceptic might wonder why we care about dynamic meaning spaces at all. What about the point that we should avoid using a secondary error-signal (either from another agent, or from success or failure at a task)? Does the talking heads experiment tell us anything about *real* language?

Firstly, there are good linguistic reasons to prefer the talking heads model of semantics over, say, the semantics in Oliphant's simulations. Bowerman and Choi [11] in a study of the acquisition of spatial vocabulary in English, Dutch, and Korean show that a universal set of atomic semantic concepts is unlikely, but it is exactly this sort of set that many computational models rely on.

Secondly, it may be possible to implement a model with flexible meaning spaces that uses a more realistic approach to learning. A. Smith [96] proposes an alternative approach to constructing mappings between discrimination trees and signals based on obverter. Results at the moment seem to be equivocal, but there are a number of approaches that could make it simpler for agents to acquire communication systems using pure observational learning. Smith suggests that a well-known principle from the literature on child language acquisition—the principle of contrast [76]—could aid in discrimination tree and lexicon construction. Another point that Smith makes is that children do not seem to be free to make *any* meaning distinction they like. It is likely that there are preexisting biases for particular distinctions that may aid the search for a shared language (for a well-known example, we could look at the work on color-term universals by Berlin and Kay [6]).

Although models of language evolution such as the talking heads simulations seem to take us closer to some of the specifically *human* aspects of human language (i.e., flexible semantics), little has so far been said about the most strikingly unique aspect of language, its syntax. The next section reviews some of the recent work in this area and shows how it builds on results from simulations of learned simple communication systems.

6   Syntax from Iterated Learning

How does the uniquely structured mapping between meanings and strings of symbols evolve in humans? And why is it unique?

Much is made of the uniqueness of human language. Typically this uniqueness is related to the syntactic nature of language as a communication system. Exactly how to

Table 1. Compositional English and holistic equivalents.

Compositional	Holistic
Walked	Went
I greet you	Hi
Died	Bought the farm
I thought I saw a pussy cat	*Bark*

characterize the syntax of language is an enormous problem and constitutes a large proportion of the research in theoretical linguistics. Of course, this in turn poses a problem for the ALife researcher proposing an explanation for language origins. Exactly what are we trying to explain? This is addressed in the literature on language origins by setting out a particular simplified characterization of the general properties of human language syntax (see, for example, Pinker & Bloom's [90] list of features of language, or Newmeyer's [80] explanation for polystratality of syntax), or by picking a principle from a particular model of syntax and studying that in more detail (e.g., both [79] and [28] look at the subjacency principle from generative grammar).

In this section, I will take the former approach and attempt to characterize one of the most fundamental properties of human language syntax: compositionality. In linguistic semantics (e.g., [77]), the term is used to refer to the way in which the meaning of an utterance is typically *composed* through some function of the meanings of parts of that utterance and the way in which those parts are put together. Notice that the notion of semantic compositionality allows for recursion in language, since the parts of the utterance themselves may have meaning through a compositional function.

Table 1 shows the difference between compositional English sentences on the left with *holistic* equivalents on the right with similar meanings[11] (the last one is the Vervet alarm call for leopards—see, for example, [24]).

Before the emergence of compositionality can be explored in ALife simulations, agents need to be able to use signals with some kind of internal structure and represent meanings with structured representations. For example, the signals could be strings of symbols, and the meanings could be a set of feature–value pairs. A compositional mapping in this case would be one in which similar strings map onto similar meanings. Conversely, a *holistic* mapping is one in which the similarity structure in one space is not preserved in the other (see Figure 3).

A simple example of this kind of model is given in [68]. In this simulation, signals and meanings are represented as 8-bit vectors. The agents are simple feed-forward networks with 8 input units, 8 hidden units, and 8 output units. The inputs are signals and the outputs meanings. The backpropagation of error training algorithm is used to model learning in these agents.

Because they are strictly feed-forward, there is a problem with this approach. Although the networks after training can map received signals onto meanings, they cannot do the reverse. To solve this problem the obverter strategy discussed earlier in this article can be used. To produce a signal s from a meaning m, a network finds the signal that would maximize the likelihood of correct reception if the agent was talking to

---

11 Of course, in some circumstances, the phrase "bought the farm" is more compositional (i.e., when referring to someone buying a farm). Idioms generally appear to have compositional syntax but in fact have a noncompositional relationship with their meaning. In many ways this could be compared with some complex animal signals such as the song of whales or birds and the long calling of gibbons (assuming these do not have a compositional relationship with their meanings).
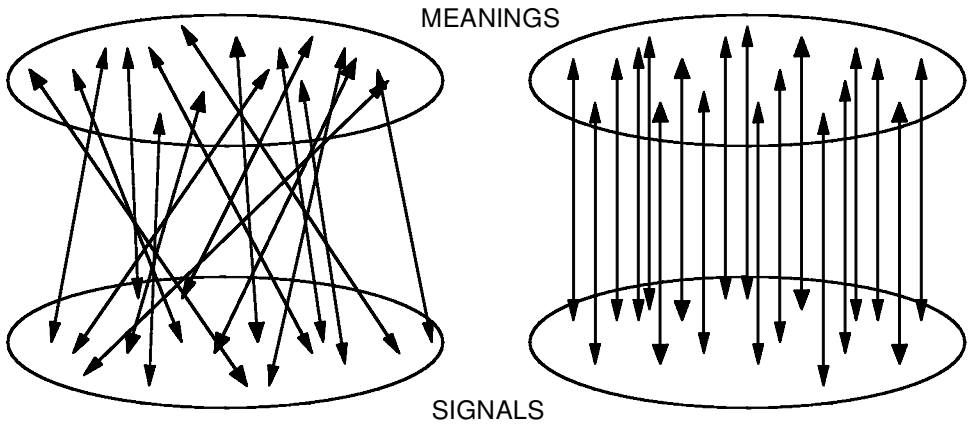
Figure 3. Two types of mapping between a signal space and a meaning space: holistic on the left, and compositional on the right.

another agent identical to itself.

$$s_{\text{desired}} = \underset{s}{\text{argmax}}\, P(s \mid m) \tag{1}$$

$$= \underset{s}{\text{argmax}}\, \frac{P(m \mid s)P(s)}{P(m)} \tag{2}$$

$$= \underset{s}{\text{argmax}}\, P(m \mid s) \tag{3}$$

$$= \underset{s}{\text{argmax}}\, C(m \mid s) \tag{4}$$

where $C(m \mid s)$ is the confidence that the network has in the mapping $s \rightarrow m$.

In other words, find the signal that maximizes the network's confidence in the given meaning. To calculate $C(m \mid s)$, we treat the real-numbered network outputs $o[1, \ldots, 8]$ as a measure of confidence in the the meaning bits $m[1, \ldots, 8]$.

$$C(m[1 \ldots 8] \mid o[1 \ldots 8]) = \prod_{i=1}^{8} C(m[i] \mid o[i]) \tag{5}$$

$$C(m[i] \mid o[i]) = \begin{cases} o[i] & \text{if } m[i] = 1, \\ (1 - o[i]) & \text{if } m[i] = 0. \end{cases} \tag{6}$$

With these agents, an iterated learning simulation can be conducted as follows:

1. An initial population is set up consisting of two randomly initialized networks, a speaker and a hearer.

2. A certain number of random meanings are chosen from the set of binary numbers 00000000 to 11111111, with replacement.

3. The speaker produces signals for each of these meanings by applying the obverter procedure.

4. This set of signal–meaning pairs is used to train the hearer network using the backpropagation of error learning algorithm.[12]

5. The speaker is removed, the hearer is designated a speaker, and a new hearer is added (with randomly initialized weights).

6. The cycle repeats.

With this model (and with many others), the results are determined by the number of meaning–signal pairs the learners are given. I refer to this value as the size of *bottleneck b* through which the language evolving in the simulation must pass from generation to generation. There are essentially three regimes of behavior with this model:

1. With low b, the language is unstable and inexpressive.

2. With high b, the language is fairly stable and expressive, but takes a long time to reach the point of maximum expressivity.

3. With an intermediate value of b, however, a completely stable and maximally expressive language emerges rapidly.

An examination of the languages that arise in each case shows that the medium bottleneck sizes result in compositional languages. A compositional language in this case is one in which each bit of the signal vector determines the state of one bit of the meaning vector (and vice versa).

Why does this result emerge? One way of looking at the dynamics of this system is from the point of view of the language, not the agents. The most stable languages will be those that are learnable. If too few examples are seen by the learner, then all languages will be equally unstable. However, if there are enough examples to learn from, the most stable languages will be those that can be reliably learned from different subsets of the utterances they generate (imagine trying to reconstruct the mappings in Figure 3 from just a few of the lines showing—you would be more successful with the mapping on the right). These languages are going to be the ones that are generalizable (see [57]). For the network, it is trivial to form generalizations that reflect the structure of the signal and meaning bit vectors. As the number of training examples increases, however, the pressure on the language to be generalizable decreases, and the network can tolerate languages that involve a degree of memorization.

This is a simple demonstration of the interesting dynamics that arise from the iterated learning framework when there is structure in the meaning and signal spaces. A criticism of this kind of model is that it lacks generality. It is possible that this compositionality result is due to some element of the simulation rather than to languages adapting to the learning task. However, there are an increasing number of models in the literature that replicate this basic result (see [58] for a deeper review):

Batali, 1998 [3]. This model uses a population of recurrent neural networks. The signals are strings of letters, and the meanings are bit vectors with some internal structure. There is no population turnover in this simulation,[13] but the agents converge on a system of strings that are made up of substrings that reflect the

---

12 The learning algorithm used has a learning rate of 0.1 and no momentum term. Each learner is presented with 100 randomized epochs of the data set.

13 Tonkes [106] classifies this simulation and Batali's other work as a *negotiation model* in contrast with the *iterated learning model*. This difference in terminology reflects the fact that Batali has no generational turnover.

meaning-space structure. The result is remarkably reminiscent of the morphology of natural languages. These results are replicated in [42].

Kirby, 2000 [64]. In a similar fashion, this paper uses strings and structured meanings. However, the languages are learned by a heuristically driven grammar inducer rather than a neural network. Once again, the language evolves such that utterances are made by combining words into sentences. To get this symbolic approach to work, there is the need for some form of random *invention.* This is never needed in the network simulations, because they are always able to produce an output.

Kirby, 2002 [66]. These results are extended by the introduction of hierarchically structured meaning representations (similar to predicate logic). The meaning space is now potentially infinite, so a bottleneck smaller than the whole meaning space is guaranteed. The result is the emergence of recursion in the simulations.

Batali, 2002 [4]. This paper also uses symbolic representations. Like the earlier paper by Batali, there is no population turnover. Learning is implemented using an instance-based learning algorithm that generalizes through analogy from structures the agent has memorized. Some interesting structures emerge that are reminiscent of passives and reflexives in real languages, for example. Another strength of this paper is the way in which the semantics are structured. Each meaning is a "flat" bag of simple predicates. Interestingly, the emergent languages are recursively structured even though the semantics are not.

Teal and Taylor, 1999 [105]. In this paper, there is no meaning space. Agents induce finite-state machines to generate sequences of letters using the *minimum description length* (MDL) approach. The goal of this paper is to show that languages are stable if they are compressible.

Kirby, 2001 [65]. In all the previous papers, every meaning is equally probable. This paper demonstrates that a skewed distribution over meanings produces a different result. If speakers prefer short strings, then languages emerge that are partly compositional and partly holistic. The holistic parts of the language are the highly frequent parts. This result mirrors what is found in natural languages.

Brighton and Kirby, 2001 [13]. This paper employs the MDL approach of Teal and Taylor [105] but with a modified form of finite-state machine that is able to represent meanings. These are used to predict the circumstances under which compositionality will be stable. The result (confirmed in the mathematical study by Brighton [12]) is that compositionality is more stable with meaning spaces of high dimensionality.

Zuidema, 2001 [114]. A learning algorithm similar to the one of Kirby [64] is used to demonstrate that the iterated learning framework moves the lower bound result of Nowak, Komarova, and Niyogi [84], enabling stable languages with fewer exposures to the data. Zuidema also shows that iterated learning mitigates some classic learnability results such as [41].

Kirby, Smith and Brighton 2002 [69]. To explore in more detail the ways in which the structure of the meaning space can influence the relative probability that compositional versus holistic languages will emerge, many thousands of simulations with differing meaning sets and bottleneck sizes are run until a stable language results. As with earlier work by Brighton [13] a correlation between degree of compositionality and the structure of the meaning space is shown. The learners in this simulation were an extended form of simple Hebbian associative networks described in [98].

Not all computational modeling of the emergence of syntactic structure relies on the iterated learning framework. The next section will deal with models that include a component of biological evolution, but some interesting hybrid models should be mentioned here. Tonkes [106] explores which languages fit a particular learning bias by a hill-climbing approach as well as allowing those languages to emerge naturally in an IL model. Perfors [89] uses genetic programming techniques to study the conditions for the emergence of structured goal-directed communication.

Hashimoto and Ikegami [49] evolve grammars directly to understand the relationship between their formal properties and selection pressures based on *speaking*, *recognizing*, and *being recognized* (see also [113] for review). Another approach is exemplified by Hashimoto [47], who explores a "constructivist" theory of linguistic structure in which a web of statistical relationships arises out of the dynamics of interactions between agents. More recently, Hashimoto has linked these two models, which he says reflect different time scales of linguistic evolution. Ultimately, this kind of synthesis of local and global dynamics may prove the most fruitful direction for future research (see [48] for an introduction).

Some of the recent work on modeling grounded semantics has started to look at emergent signaling systems that have internal structure. Cangelosi's models of goal-directed communication in foraging agents (mentioned in Section 5) demonstrate the emergence of a signal with two components: a *proto-verb* and a *proto-noun* [19]. The former designates the relevant behavior for the forager (approaching or avoiding), and the latter identifies the class of foraged item. What seems to be happening in these simulations is that the agents evolve internal representations that enable them to forage successfully. These representations then form the basis of the subsequent cultural evolution of the proto-noun/verb distinction.

The theme of grounding structural linguistic distinctions on a prior task-oriented substrate is also taken up by Hazlehurst and Hutchins [51, 61]. In their model the environment consists of a spatially organized "scene" within which there may be a number of objects. Each agent attends to one object in the scene at any one point in time. During an interaction, there are two agents: a speaker and a hearer. Both are able to see what object the other is attending to. In other words, they are both able to follow the gaze of the other. During an interaction, the speaker attempts to direct the hearer's gaze to a particular object in a visual field. A successful interaction involves both speaker and hearer moving their gaze over the visual field step-by-step in such a way that they always share intention and end up gazing at the intended object. As well as being able to observe the speaker's direction of gaze, hearers also "hear" signals from the speaker. Over time signals emerge not only for the different types of objects, but also for different "moves" of attention across the visual field. Once again, the emergent signaling system, which Hazlehurst and Hutchins argue exhibits the hallmarks of syntactic structure, is grounded in the structure of the task rather than being directly predetermined by the experimenter.

For many, modeling the evolution of syntactic structure is an important research goal since syntax is the most strikingly unique aspect of human language. Sometimes it can seem like the emergence of syntax is the Holy Grail for ALife models of language, and the success of the growing body of work on the subject appears to be cause to celebrate the ALife approach. We should be cautious, however, in evaluating the success of the work so far. Although it is clear that properties such as compositionality and recursion are fundamental to the syntax of all natural languages, there is much else besides! As Bickerton [8] puts it in his critique of computational/mathematical models (or modelers), "they account for 'the cat sat on the mat' and then cross their fingers, confident that 'self-organization' will take care of the rest." In fact, none of the models I am aware of can account for *the cat sat on the mat*. A fundamental part of human

language is demonstrated by that sentence that has yet to be accounted for by these models: the functional/contentive dichotomy in the lexicons of all languages [23]. In the case of the sentence Bickerton gives, a more sophisticated model would account for the different status of *the* and *on* from the other words in the sentence with regard to their processing, storage, acquisition, form, distribution, and function.

There has been relatively little research using ALife methods that tackles more specifically *linguistic* features of the syntax of human languages. However, Kirby [63] looks at the link between learning and processing constraints and universal properties of language using an IL approach. The universals examined relate to word order, the formation of relative clauses, patterns of agreement (i.e., the ways in which the forms of pairs of words covary according to their syntactic and semantic relation in a sentence), and a constraint on long-distance dependencies between positions within sentences. In a similar spirit, although not employing a "full" IL model, is the work by Christiansen [27] that relates the learning biases of recurrent neural networks to some of the linguistic universals listed above. Ultimately, if ALife modeling is to succeed as an approach to syntax, there needs to be much more work along these lines, showing how the properties discovered by syntactic theorists are related to the complex dynamical systems in Figure 1.

7   The (Co)evolution of Language Acquisition

How does the language learner evolve? How does the evolution of languages through iterated learning impact on this?

So far in this review, the only models that have appealed to biological evolution have been those looking at innate simple signaling systems. More specifically linguistic behavior has been explained in relation to learning and cultural evolution (which I have suggested emerges out of *iterated learning*). However, as mentioned in Section 2, this is not the end of the story.

The IL models discussed so far essentially explore the relationship between inductive bias on the one hand, and emergent universal properties of the structure of languages on the other. However, the models say little about where this bias comes from. A common sense notion of inductive bias is that it is the knowledge that the learner brings to bear on a learning task before any of the data has been experienced. This notion of prior knowledge corresponds closely to the Chomskyan notion of linguistic innateness. What else will provide a learner its prior biases if not its genetic specification?

At this point there is much disagreement, the review of which would take several other articles, so suffice it to say that there are two main areas of controversy. There is disagreement about (a) whether the innate biases are domain specific or domain general, and (b) how these biases come about. Obviously these two questions are related. For example, one might take the stance (as Pinker and Bloom [90] do) that innate constraints[14] on learning evolve under selective pressures related to communication. The logical conclusion to this is that the particular set of constraints (or, to put it another way, the architecture of a language acquisition device) is specific to the domain of communication using language.

Ultimately, the role for ALife modeling in this contentious area should be to act as a check on the theoretical positions, ensuring that the mechanisms appealed to actually lead to the results intended. As was pointed out in Section 2, ALife techniques are particularly appropriate to cases where two or more adaptive systems interact, because

---

14 Interestingly much of linguistic theory assumes hard constraints on learning and has little to say on the matter of the more general conception of innateness in terms of *bias*.

it is often difficult to predict the results of such interactions in advance of modeling them.

In particular, the interaction of learning and evolution is a "classic" topic for ALife models [5] and is also highly pertinent to the evolution of language. Batali [2] presents an early attempt to look at how learning bias might evolve in the face of a syntactic learning task. In his simulations, recurrent neural networks are trained on a next-character prediction task where the characters are drawn from a language such as $a^n b^n$. Initial experiments showed that the networks were unable to learn the task well. In particular, they appeared to overgeneralize to the language $a^n b^m$. Batali hypothesized that the problem lay with the inductive bias of the networks. To explore this idea, the initial weight settings of the network (usually randomly assigned) were evolved by a genetic algorithm whose fitness metric was related to the ability of the networks to process strings in the language.

The combination of training through backpropagation and search for initial weights through an evolutionary algorithm resulted in a considerable improvement on the next-character prediction task for $a^n b^n$. Of course, this does not correspond well to the task facing evolving language learners, because there is not a single target language to be acquired. To test if it was possible to generalize his results to a *class* of languages, Batali repeated the experiment with a set of language problems that all had the same structural properties but differed according to the assignment of characters to underlying categories. During the "lifetime" of a single network, the language remained constant, but from generation to generation of the evolutionary search, the particular language from the class was varied.

The result was a set of initial weights that enabled a network to make successful next-character predictions for any of the languages from the class. In other words, evolution had tuned the learning bias to aid the acquisition of a particular class of language (one for which the general-purpose learner was ill equipped). Could a similar mechanism of *genetic assimilation* be operating in the evolution of language? Could the class of human languages have shaped the evolution of a language-specific learning bias in humans?

There is one obvious disanalogy between the Batali model and the real case of language evolution. Batali provides the learning task in advance of the simulation. The learners are adapting to a specific class of languages that exists prior to the evolutionary process. In reality, the language learning task is provided by other learners—this is what leads to the dynamics of the IL model discussed in the previous two sections. An obvious question is whether similar genetic assimilation will occur with the addition of this dynamic alongside learning and biological evolution.

Following up a preliminary study described by Turkel [107] (itself a modification of Hinton & Nowlan's [53] classic model), Kirby and Hurford [67] attempt to simulate the joint action of learning, culture, and evolution. The representation of a language in this model is based on an approach to language acquisition called principles and parameters [25]. In P+P theory, the language acquisition device consists of a set of fixed principles (invariant properties of language) and a set of switchable parameters (which account for linguistic variation). Learning consists of searching for parameter settings that best model the data seen (the learning procedure used in [67] is the trigger learning algorithm of [40]). In the simulation, the language acquisition device (LAD) is specified by an eight-long ternary vector made up of 1s, 0s, and ?s. The ?s are parameters that will be set after learning, whereas the 1s and 0s represent the innately fixed principles.

To assess the roles of natural selection and iterated learning, Kirby and Hurford arbitrarily specify some languages to be more functional (with respect to communication) than others. The question they ask is, under which conditions will the eventual distribution of languages reflect the functional pressures they build in?

In their experiment, Kirby and Hurford evolve LADs on the basis of the communicative success of agents with those LADs with respect to neighboring agents. The results showed partial genetic assimilation of parameters into principles, and a consequent reduction in the range of linguistic variation. Surprisingly, however, the languages that are learnable by agents at the end of the simulation tended not to be those that were particularly functional. In fact, some of the "nativization" of parameters into principles led to agents that were biased *against* functional languages.

The result changes radically if the functionality of a language influences the *cultural* transmission of that language. A second simulation adds a factor whereby it is more difficult to process (and therefore learn) less functional languages. With this extra mechanism in place, the resultant LADs show assimilation of functional principles.

This brief summary shows that there are nontrivial interactions between the three adaptive systems underlying language. Ultimately, it will not be possible to ignore one or other of these systems; a more holistic approach will be necessary. The computational challenges are great, however. The Kirby and Hurford model is hugely simplified, which means the applicability of their results to linguistic problems is not clear.

Briscoe [15] attempts to make coevolutionary models closer to those studied in mainstream linguistics. The simple vectors of [67] are replaced by Briscoe with an actual theory of syntactic variation: generalized universal categorial grammar. The dynamics that arise from Briscoe's models are far from straightforward, and their analysis far from complete. However, their usefulness for tackling linguistic problems such as the reason for particular trajectories in language change, or the mechanism behind creolization, is greater by virtue of their relative closeness to real linguistic models.

Even putting the degrees of abstractness issue to one side, there are many other challenges facing a framework that brings together language learning, language evolution through iterated learning, and the evolution of learning biases. One serious issue is whether we can draw reliable conclusions from our models if we do not know the genetic basis for the language learning mechanisms. Yamauchi [112] demonstrates, for example, that the assimilation of features into an LAD can only happen where there is a close match between genotypic and phenotypic space. Essentially, small changes in the genome (e.g., through mutation) must correspond to small changes in the resultant phenotype for genetic assimilation to occur.[15] Considering the complexity of the relationship between the genetic makeup of an individual on the one hand, and the space of possible languages on the other, it seems that it will be some time before our simulations will give uncontroversial answers to the question of the origin of the LAD.

8   Future Work

What are the big questions that remain? Where next for ALife models?

At the start of this article, I listed a number of "facts about language" that an explanatory account of language should account for. The work reviewed here leads to two conclusions:

1. Artificial life techniques are highly appropriate for the study of human language, but...

2. we have a long way to go yet.

---

15 These issues are explored further in [16].

Perhaps, rather than looking through this list and totting up a score on how well particular approaches have done in explaining these facts, it is more useful to think of them as informing us of which issues really matter. In other words, I hope that the issues of interest to linguists may act as a compass for future research directions for ALife.

In a similar way I hope this review will be useful in identifying the *gaps* in the simulation literature. It is these that may eventually prove to be the most rewarding (but, perhaps, also the most difficult) areas of study. There are two areas that I, personally, feel will be particularly fruitful:

Coevolution of syntax and semantics in the ILM. So far, the results of running IL models suggest that syntactic structure will emerge given (a) a learning bottleneck, and (b) a structured meaning space. In most simulations that show emergent compositionality, this meaning space structure is given in advance by the experimenter. Other models, however, demonstrate that meaning spaces can themselves be learned, and they vary from agent to agent. Can structure in both meanings and signals emerge in a simulation? The logic of the IL model suggests that they might, since only by matching the structure of the signal with the structure of the meaning can languages optimize their transmissibility. Hutchins and Hazlehurst's work on modeling attention following [51, 61] may be a fruitful starting point for examining these issues.

Origins of iterated learning. A combination of cultural evolution through iterated learning with biological evolution of learning biases looks like the best approach to understanding the emergence of syntax. However, it begs some fairly fundamental questions. For example, why is it that only human beings have an open-ended, syntactically structured learned communication system? This is not directly answered by the IL framework, or indeed the models of grammatical assimilation.

A possible answer to this puzzle lies in understanding what is necessary to build into an IL model for it to work. All the models that look at compositionality assume that the learner is given training data consisting of pairs of meanings and signals. The justification for this is that we must assume that at least some of the time children are able to infer the communicative intentions of the adults around them (or the other children around them in the case of creolization). But how is it that children actually do this? It is clearly not by explicit feedback (since some people with language disabilities have extremely impoverished production but score more highly on comprehension tests). Nor is language learning simply a case of association; rather, children are exquisitely attuned to the thought processes of others [10]. Perhaps, then, the uniqueness of language is down to the uniqueness of the human environment of adaptation that leads to selection for an ability and desire to *mind read*. Artificial life models that uncover interactions between the complexity of social groups, theory of mind, and observational learning may help us to uncover the *origins* of human language.

Further down the line, research will need to tackle specific features of language that linguists are interested in—for example, cross-linguistic variation in the syntax of anaphora, or constraints on the grammaticality of different word orders. Studies of the biases that arise out of serial processing and how these affect the process of linguistic transmission are relevant here [27, 28].

There are a number of outstanding questions that ALife techniques may help to answer. Which features of language should be ascribed to innate biases (arbitrary or adaptively evolved?), and which to general properties of learning? Which features are

explained by as yet poorly understood properties of cultural evolution, and which to emergent properties of the interaction of these systems?

Linguistics continues to develop more and more intricate understanding of the syntax of languages in all their aspects, universal and specific, dynamic and synchronic. This work is heavily data driven. In many ways, we have an embarrassment of information about what language is and how it behaves.

And yet to many people there is confusion and disagreement about what to do with all this knowledge. Explanatory models for syntactic phenomena are rarely uncontroversial and are usually short lived. I would contend that this is because linguistics does not have a way of tackling the complexity of the interaction of the actual processes underlying the origins and dynamics of language. Just as ALife since its conception has made sense of the array of facts about biological systems that are known by studying in miniature the processes that give rise to these systems—so too I hope it can help provide an explanatory underpinning for linguistics. We are just getting started.

Acknowledgments

References

1. Ackley, D., & Littman, M. (1994). Altruism in the evolution of communication. In R. Brooks & P. Maes (Eds.), *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (pp. 40–48). Redwood City, CA: Addison-Wesley.

2. Batali, J. (1994). Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In R. Brooks & P. Maes (Eds.), *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (pp. 160–171). Redwood City, CA: Addison-Wesley.

3. Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 405–426). Cambridge, UK: Cambridge University Press.

4. Batali, J. (in press). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge, UK: Cambridge University Press.

5. Belew, R., & Mitchell, M. (Eds.). (1996). *Adaptive individuals in evolving populations*. Reading, MA: Addison-Wesley.

6. Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.

7. Bickerton, D. (1981). *Roots of language*. Ann Arbor, MI: Karoma.

8. Bickerton, D. (in press). Rooting innateness in evolution. In M. Christiansen & S. Kirby (Eds.), *Language evolution: The states of the art*. Oxford, UK: Oxford University Press.

9. Billard, A., & Dautenhahn, K. (1999). Experiments in learning by imitation—Grounding and use of communication in robotic agents. *Adaptive Behavior, 7*, 415–438.

10. Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

11. Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge, UK: Cambridge University Press.

12. Brighton, H. (2002). Compositional syntax from cultural transmission. *Artifical Life, 8*, 25–54.

13. Brighton, H., & Kirby, S. (2001). The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In J. Kelemen & P. Sosik (Eds.), *Advances in Artificial Life (Proceedings of the 6th European Conference on Artificial Life)*. Heidelberg: Springer.

14. Briscoe, E. (2000). Evolutionary perspectives on diachronic syntax. In S. Pintzuk, G. Tsoulas, & A. Warner (Eds.), *Diachronic syntax: Models and mechanisms*. Oxford, UK: Oxford University Press.

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

16. Briscoe, E. (in press). Grammatical assimilation. In M. Christiansen & S. Kirby (Eds.), *Language evolution: The states of the art*. Oxford, UK: Oxford University Press.

17. Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

18. Bullock, S. (1997). An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In P. Husbands & I. Harvey (Eds.), *Proceedings of the Fourth European Conference on Artificial Life* (pp. 454–463). Cambridge, MA: MIT Press.

19. Cangelosi, A. (2001). Evolution of communication and language using signals, symbols, and words. *IEEE Transactions on Evolutionary Computation, 5*(2), 93–101.

20. Cangelosi, A., Greco, A., & Harnad, S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science, 12*(2), 143–162.

21. Cangelosi, A., & Parisi, D. (1998). The emergence of a 'language' in an evolving population of neural networks. *Connection Science, 10*(2), 83–97.

22. Cangelosi, A., & Parisi, D. (Eds.). (2002). *Simulating the evolution of language*. London: Springer.

23. Cann, R. (2000). Functional versus lexical: A cognitive dichotomy. In R. D. Borsley (Ed.), *The nature and function of syntactic categories*. New York: Academic Press.

24. Cheney, D., & Seyfarth, R. (1990). *How monkeys see the world: Inside the mind of another species*. Chicago: University of Chicago Press.

25. Chomsky, N. (1981). Principles and parameters in syntactic theory. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics: The logical problem of language acquisition*. London: Longman.

26. Chomsky, N., Belletti, A., & Rizzi, L. (2000). An interview on minimalism. University of Siena online publication. http://www.cicl.unisi.it/pubblicazioni.php.

27. Christiansen, M., & Devlin, J. (1997). Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the 19th Annual Cognitive Science Society Conference* (pp. 113–118). Mahwah, NJ: Erlbaum.

28. Christiansen, M., & Ellefson, M. R. (in press). Linguistic adaptation without linguistic constraints: The role of sequential learning in language evolution. In J-L Dessalles, A. Wray, & C. Knight (Eds.), *Transitions to language*. Oxford, UK: Oxford University Press.

29. Clark, R. (1992). The selection of syntactic knowledge. *Language Acquisition, 2*, 85–149.

30. Clark, R., & Roberts, I. (1993). A computational model of language learnability and language change. *Linguistic Inquiry, 24,* 299–345.

31. Clark, R. A. J. (1996). *Internal and external factors affecting language change: A computational model.* Master's thesis, University of Edinburgh.

32. Croft, W. (1995). *Typology and universals.* Cambridge, UK: Cambridge University Press.

33. Cziko, G. (1995). *Without miracles: Universal selection theory and the second Darwinian revolution.* Cambridge, MA: MIT Press.

34. Darwin, C. (1981). *The descent of man, and selection in relation to sex.* Princeton, NJ: Princeton University Press. (Original work published 1871)

35. de Boer, B. (2001). *The origins of vowel systems.* Oxford, UK: Oxford University Press.

36. Dennett, D. (1987). *The intentional stance.* Cambridge, MA: MIT Press.

37. DiPaolo, E. (1997). An investigation into the evolution of communication. *Adaptive Behavior, 6,* 285–324.

38. DiPaolo, E. (1999). A little more than kind and less than kin: The unwarranted use of kin selection in spatial models of communication. In D. Floreano, J-D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life* (pp. 504–513). Berlin: Springer.

39. Epstein, J. M., & Axtell, R. (1996). *Growing artificial societies: Social science from the bottom up.* Cambridge, MA: MIT Press.

40. Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry, 25,* 355–407.

41. Gold, E. M. (1999). Language identification in the limit. *Information and Control, 10,* 447–474.

42. Goroll, N. (1999). *(The deep blue) Nile: Neuronal influences on language evolution.* Master's thesis, University of Edinburgh.

43. Hamburger, H., & Wexler, K. (1975). A mathematical theory of learning transformational grammar. *Journal of Mathematical Psychology, 12,* 137–177.

44. Hamilton, W. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology, 7,* 1–16, 17–52. (Parts I and II)

45. Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition, 56,* 61–98.

46. Harnad, S. (1990). The symbol grounding problem. *Physica, D, 42,* 335–346.

47. Hashimoto, T. (1999). Modeling categorization dynamics through conversation by constructive approach. In D. Floreano, J-D Nicoud, & F. Mondada (Eds.), *Advances in artifical life* (pp. 730–734). Berlin: Springer.

48. Hashimoto, T. (2001). The constructive approach to the dynamical view of language. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 307–324). London: Springer.

49. Hashimoto, T., & Ikegami, T. (1996). Emergence of net-grammar in communicating agents. *Biosystems, 38,* 1–14.

50. Hauser, M. D. (1996). *The evolution of communication.* Cambridge, MA: MIT Press.

51. Hazlehurst, B., & Hutchins, E. (1998). The emergence of propositions from the co-ordination of talk and action in a shared world. *Language and Cognitive Processes, 13,* 373–424.

52. Hebb, D. (1949). *The organization of behavior.* New York: Wiley.

53. Hinton, G., & Nowlan, S. (1987). How learning can guide evolution. *Complex Systems, 1,* 495–502.

54. Hurford, J. R. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua, 77,* 187–222.

55. Hurford, J. R. (1991). The evolution of critical period for language acquisition. *Cognition, 40*, 159–201.

56. Hurford, J. R. (1999). Artificially growing a numeral system. In J. Gvozdanovic (Ed.), *Numeral types and changes worldwide* (pp. 7–41). Berlin: Mouton de Gruyter.

57. Hurford, J. R. (2000). Social transmission favours linguistic generalization. In C. Knight, M. Studdert-Kennedy, & J. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 324–352). Cambridge, UK: Cambridge University Press.

58. Hurford, J. R. (in press). Expression/induction models of language evolution: Dimensions and issues. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge, UK: Cambridge University Press.

59. Hurford, J. R., & Kirby, S. (1998). Co-evolution of language-size and the critical period. In D. Birdsong (Ed.), *New perspectives on the critical period hypothesis and second language acquisition* (pp. 39–64). Mahwah, NJ: Erlbaum.

60. Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial societies: The computer simulation of social life* (pp. 157–189). London: UCL Press.

61. Hutchins, E., & Hazlehurst, B. (2002). Auto-organization and the emergence of shared language structure. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 279–305). London: Springer.

62. Johansson, C. (1997). *A view from language*. Lund, Sweden: Lund University Press.

63. Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals*. Oxford, UK: Oxford University Press.

64. Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 303–323). Cambridge, UK: Cambridge University Press.

65. Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation, 5*(2), 102–110.

66. Kirby, S. (in press). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge, UK: Cambridge University Press.

67. Kirby, S., & Hurford, J. R. (1997). Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands & I. Harvey (Eds.), *Proceedings of the Fourth European Conference on Artificial Life* (pp. 493–502). Cambridge, MA: MIT Press.

68. Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). London: Springer.

69. Kirby, S., Smith, K., & Brighton, H. (2002). *Language evolves to aid its own survival*. Manuscript in preparation.

70. Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 380–402). Oxford, UK: Blackwell.

71. Langton, C. (Ed.). (1995). *Artificial life: An overview*. Cambridge, MA: MIT Press.

72. Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

73. Lightfoot, D. (1999). *The development of language: Acquisition, change, and evolution*. Oxford, UK: Blackwell.

74. Lindgren, K., & Nordahl, M. G. (1995). Cooperation and community structure in artificial ecosystems. In C. Langton (Ed.), *Artificial life: An overview* (pp. 39–60). Cambridge, MA: MIT Press.

75. MacLennan, B., & Burghardt, G. (1994). Synthetic ethology and the evolution of cooperative communication. *Adaptive Behavior, 2*, 161–187.

76. Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.

77. Montague, R. (1970). English as a formal language. In B. Visentini (Ed.), *Linguaggi nella Società e nella Tecnica* (pp. 189–223). Milan: Edizioni di Comunità.

78. Nagel, K., & Rasmussen, S. (1994). Traffic at the edge of chaos. In R. A. Brooks and P. Maes (Eds.), *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (p. 222). Cambridge, MA: MIT Press.

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

80. Newmeyer, F. J. (1992). Iconicity and generative grammar. *Language, 68*, 756–796.

81. Niyogi, P., & Berwick, R. (1995). *The logical problem of language change*. (Tech. Rep. AIM-1516). MIT AI Lab, Cambridge, MA.

82. Noble, J. (1998). Evolved signals: Expensive hype vs. conspiratorial whispers. In C. Adami, R. Belew, H. Kitano, & C. Taylor (Eds.), *Artificial Life 6: Proceedings of the Sixth International Conference on Artificial Life* (pp. 358–367). Cambridge, MA: MIT Press.

83. Noble, J. (1999). Cooperation, conflict and the evolution of communication. *Adaptive Behavior, 7*, 349–370.

84. Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science, 291*, 114–117.

85. Oliphant, M. (1997). *Formal approaches to innate and learned communication: Laying the foundation for language*. Unpublished doctoral dissertation, University of California, San Diego.

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

87. Oliphant, M. (in press). Learned systems of arbitrary reference: The foundation of human linguistic uniqueness. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge, UK: Cambridge University Press.

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

89. Perfors, A. (2000). *Simulated evolution of communication: The emergence of meaning*. Master's thesis, Department of Linguistics, Stanford University, Stanford, CA.

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

91. Prusinkiewicz, P. (1995). Visual models of morphogenesis. In C. Langton (Ed.), *Artificial life: An overview* (pp. 61–74). Cambridge, MA: MIT Press.

92. Quinn, M. (2001). Evolving communication without dedicated communication channels. In J. Kelemen & P. Sosik (Eds.), *Advances in Artificial Life: Proceedings of the 6th European Conference on Artifical Life* (pp. 357–366). Heidelberg: Springer.

93. Ragir, S. (2002). Constraints on communities with indigenous sign languages: Clues to the dynamics of language origins. In A. Wray (Ed.), *The transition to language* (pp. 272–296). Oxford, UK: Oxford University Press.

94. Savage-Rumbaugh, S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York: Wiley.

95. Schuster, P. (1995). Extended molecular evolutionary biology: Artificial life bridging the gap between chemistry and biology. In C. Langton (Ed.), *Artificial life: An overview* (pp. 39–60). Cambridge, MA: MIT Press.

96. Smith, A. D. M. (2001). Establishing communication systems without explicit meaning transmission. In J. Kelemen & P. Sosik (Eds.), *Advances in Artificial Life: Proceedings of the 6th European Conference on Artifical Life* (pp. 381–390). Heidelberg: Springer.

97. Smith, K. (2001). The importance of rapid cultural convergence in the evolution of learned symbolic communication. In J. Kelemen & P. Sosik (Eds.), *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life* (pp. 381–390). Heidelberg: Springer.

98. Smith, K. (2002, in press). The cultural evolution of communication in a population of neural networks. *Connection Science.*

99. Smith, K. (in press). Natural selection and cultural selection in the evolution of communication. *Adaptive Behavior.*

100. Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication, 1*, 1–34.

101. Steels, L. (1999). *The talking heads experiment: Volume I. Words and meanings.* Antwerpen: Laboratorium. (Special pre-edition)

102. Steels, L., & Kaplan, F. (1999). Collective learning and semiotic dynamics. In D. Floreano, J.-D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life* (pp. 679–688). Berlin: Springer.

103. Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language* (pp. 252–271). Oxford, UK: Oxford University Press.

104. Steels, L., & Vogt, P. (1997). Grounding adaptive language games in robotic agents. In P. Husbands & I. Harvey (Eds.), *Proceedings of the Fourth European Conference on Artificial Life* (pp. 474–482). Cambridge, MA: MIT Press.

105. Teal, T., & Taylor, C. (1999). Compression and adaptation. In D. Floreano, J. D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life* (pp. 709–719). (No. 1674 in Lecture Notes in Computer Science). Berlin: Springer.

106. Tonkes, B. (2002). *On the origins of linguistic structure: Computational models of the evolution of language.* Unpublished doctoral dissertation, University of Queensland, Australia.

107. Turkel, W. J. (in press). The learning guided evolution of natural language. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models.* Cambridge, UK: Cambridge University Press.

108. Vogt, P. (2000). *Lexicon grounding on mobile robots.* Unpublished doctoral dissertation, Vrije Universiteit Brussel.

109. Werner, G., & Dyer, M. (1992). Evolution of communication in artificial organisms. In C. Langton, C. Taylor, J. Farmer, & S. Rasmussen (Eds.), *Artificial Life 2* (pp. 659–687). Redwood City, CA: Addison-Wesley.

110. Wilson, E. (1975). *Sociobiology: The new synthesis.* Cambridge, MA: Harvard University Press.

111. Wonnacot, E., & Kirby, S. (2002). *Emergent linguistic dynamics from simple mechanisms: A microscopic-modeling approach to logistic change.* Manuscript in preparation.

112. Yamauchi, H. (2001). The difficulty of the Baldwinian account of linguistic innateness. In J. Kelemen & P. Sosik (Eds.), *Advances in Artificial Life: Proceedings of the 6th European Conference on Artificial Life* (pp. 391–400). Heidelberg: Springer.

113. Zuidema, W. H. (2000). *Evolution of syntax in groups of agents.* Master's thesis, Theoretical Biology, Utrecht University.

114. Zuidema, W. H. (2001). Emergent syntax: The unremitting value of computational modeling for understanding the origins of complex language. In J. Kelemen & P. Sosík (Eds.), *Advances in Artificial Life* (pp. 641–644). (No. 2159 in Lecture Notes in Artificial Intelligence). Berlin: Springer.