Natural Language from Artificial Life

Simon Kirby
Language Evolution and Computation Research Unit
Department of Theoretical and Applied Linguistics
University of Edinburgh
http://www.ling.ed.ac.uk/~ simon

Time-stamp: "2002-02-07 16:26:54 simon"

1 Introduction

What can artifical life offer linguistics? and why is language a particularly appropriate target for A-life methods?

In this paper 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 paper 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, de Boer, 2001), critical periods for language acquisition (Hurford, 1991; Hurford and Kirby, 1998), or language change (Briscoe, 2000b; Niyogi and Berwick, 1995; Hare and Elman, 1995; Wonnacot and Kirby, 2002; Clark, 1992; Clark and Roberts, 1993; Clark, 1996; Johansson, 1997). Nor will the extensive literature on models of animal communication be the main target of this review (although some of the papers will be discussed in section 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 modelling will be the core of any future research framework for an explantory 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 signalling systems; cultural evolution of simple signalling systems that are learned rather than innate; the grounding of signals and the modelling of flexible meaning spaces; the emergence of complex structured languages both through "negociation" 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.

---

[1] At least one other excellent reviews of the literature on computational modelling of the evolution of language exists (Steels, 1997) which differs in its goal and coverage from this article.

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 (Langton, 1995), there are models of molecular evolution (Schuster, 1995) of cooperation in ecosystems (Lindgren and Nordahl, 1995), and of morphogenesis (Prusinkiewicz, 1995). There have also been attempts to use alife techniques in social science domains (Epstein and Axtell, 1996) and even to model traffic jams (Nagel and Rasmussen, 1994). One rationale for the A-life approach in these domains is that they all involve complex synergistic interactions which, when taken together, lead to emergent behaviours which are hard to predict. In complex dynamical systems, verbal theorising often leads to incorrect predictions because our intuitions about the links between local interactions and global behaviour are notoriously unreliable. Furthermore, the classical alternative — analytic mathematical modelling — may require the kinds of idealisation which will necessitate the removal of the very network of interactions that give rise to the target of explanation.

A-life provides the answer to the methodological problem of linking theory and prediction for complex dynamical systems involving multiple interacting components. Micro simulation 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 a-life models.

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

1. It is a non-trivial 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 six years old without being taught (in the sense of deliberate instruction), and without being given any reliable feedback about their communicative performance (Brown and Hanlon, 1970).

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 (Lenneberg, 1967).

6. Our eventual ability at communicating 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 (Hauser, 1996).
9. The most able non-human 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 (Savage-Rumbaugh and Lewin, 1994).
10. There is no unequivocal evidence that any now-exinct hominid was in possession of human-like communication. This suggests that language may be as little as 5000 generations old.
11. There are around 6000 different languages on Earth. Many of these show discernable similarities due to a common history. Others show similarities which 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 (Croft, 1990).
12. Languages change over a relatively short historical timescale. No language is completely static.
13. 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.
14. Despite its complexity, humans can manage fairly well without fully syntactic language. Many people are able to cope with little disadvantage in communities which employ a language they are not proficient in, for example.
15. 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) (Ragir, 2002; Bickerton, 1981).
16. Many aspects of language appear to be localised in particular areas of the cortex, and may be damaged by, for example, stroke.
17. 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. Context-free languages are not learnable in the limit in the most general case (Gold, 1967).
18. Native speakers are still able to make "grammaticality judgements" about sentences which 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 non-linguistic human behaviour.
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 (Lightfoot, 1999).

In order to come up with a truly *explanatory* linguistics, we would need to a) understand the origins of this unique behaviour, and b) derive the properties of this behaviour from the properties of non-linguistic 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 behaviours 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 analyse 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 view 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 timescale, 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.

On top of language learning on an ontogenetic timescale, and language evolution on a glossogenetic (i.e. historical) timescale, we need to consider the manner in which the brains of language users are shaped by biological evolution on a phylogenetic timescale. 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 our brains were evolving, the genes that controlled its development were under selection pressure. It is possible that this pressure changed as the very behaviours that the brain supported emerged.

To summarise: 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

---

[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 generalised to predict the grammaticality of unseen utterances and which can be employed for the description of any language as an explantory goal. In Chomsky, Belletti, and Rizzi (2000), Chomsky makes a comparison to the discovery of the period table. The periodic table was an elegant description that lead to developments in physical theory to account for its existence. In a sense, physics was required to explain the period table. Contrast this with the alternative possibility: that the periodic table had been derived from physical principles and then had been compared with the 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.

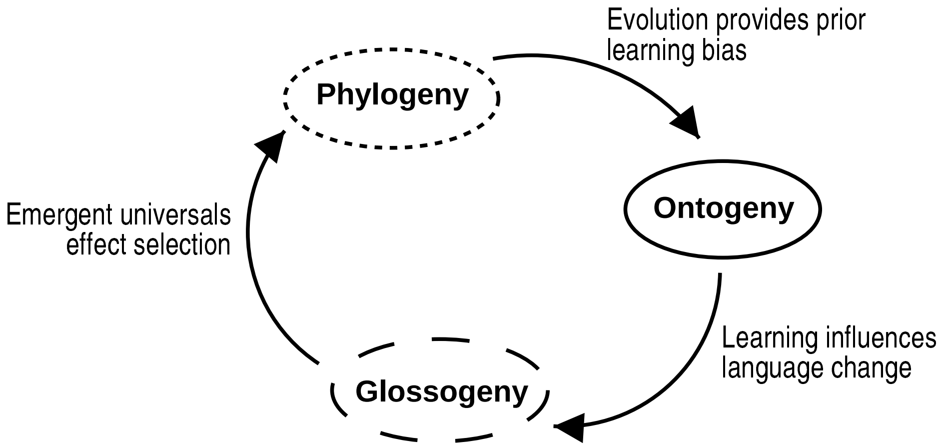

Figure 1: Language is the result of an interaction between three complex adaptive systems that operate on different timescales: the timescale of biological evolution (phylogeny), the timescale of individual learning (ontogeny), and the timescale of language change (glossogeny).

effecting how languages are learned (figure 1).

English, therefore, is the result of a stageringly 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 A-life models. In the next few sections, I will summarise a line of research that leads from simple animal signalling systems to the coevolution of languages and the language acquisition device in order to show the sort of work that can be done.

3 Innate signalling systems

What are the simplest systems of communication? and what does A-life tell us about their evolution?

The vast majority of communication that we are aware of is the result of purely innately coded behaviours. 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 (1997) reviews the work carried out within the framework of artifical life on innate communication, and draws-out the main themes in this research within his own idealised simulation model. In this section, I will summarise 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., Dennet, 1987) and ethologists looking at animal communication (e.g., Wilson, 1975). Oliphant 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.” (Oliphant, 1997, 14)

Exploitation in this context refers to cases where the behaviour 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 behaviour has been observed. If I were to avoid entering a colleagues office because I heard her sneezing, this could be considered an exploitative interaction. The sneezing is a behaviour I recognise 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 behaviour of an individual is determined to some extent by the expectation that another individual will respond in a particular way to that behaviour. 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 colour 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 signalling. In other words, the two behaviours are the result of genetically hard-wired responses to situations.

Several artificial life models have been proposed to explore what is required for such a system to emerge. Werner and Dyer (1992) used a simulation which evolved agents that were divided into two types: males and females. The females in their model were immobile but sighted, whereas the males were mobile but blind. The females were able to send signals to the males, and a fitness payoff was awarded for a male/female encounter.

In the work of Maclennan and Burghardt (1994), 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 agents’ 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 (1994) 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 which we will return to later.

Rather than model specific ecological conditions, Oliphant (1997) sets out a generalised approach to innate signalling systems.[3] We assume a set S of low-cost behaviours (signals)

---

[3] Oliphant is not unique in taking this approach. Di Paolo (1997), Noble (1998), and Bullock (1997) also use generalised models of innate signalling in order to explore the conditions under which innate communication evolves.

which are observable and can be distinguished be 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? In order 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, is:

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

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 optimally 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 a meaning 1 and the signal b for meaning 0, but interpreted both signals as meaning 0, then this agents genome would be: $ba00$. (Incedentally, 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 communication "games". In each game a random meaning would be chosen, one of the agents would produce the signal associated with that meaning (according to their genetically encoded transmission function), and the other agent would interpret that signal according to their 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 equalling 1? There are two key ways in which the general simulation model can be parameterised:

---

[4] See, for example, (Noble, 1999) for discussion of what difference it makes if we change assumptions about the cost of signalling.

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, there are two alternatives I wish to consider[5]:

Mutual benefit Both speaker and hearer benefit from successful communication.

Altruistic behaviour Only the hearer benefits from successful communication.

From this perspective, the simulation of Ackley and Littman (1994) is different from the other two described earlier in that they modelled an altruistic situation as opposed to a mutualitic 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. To understand when a perfect communicative accuracy score *can* be acheived 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 organisation The population has *spatial structure*, and communicative partners are chosen that are close together.

In the generalised model with random game-partner choice, altruistic communication does not emerge. However, with a spatially organised population, Oliphant shows that an optimal $ca(S, R)$ can be acheived even in cases where the speaker does not benefit (as in the case of Ackley and Littman, 1994). He argues that this is because spatial organisation is equivalent to *kin selection*, a mechanism known to promote the evolution of altruistic behaviour (Hamilton, 1964). This is because, when you are more likely to talk to your neighbours, 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 (1999) points out that Oliphant does not actually formally test the equivalence of spatial organisation and kin selection. He also shows that, in some circumstances, there are alternative mechanisms that are not related to kin selection by which spatial organisation can promote the evolution of optimal communication.

The use of artificial life modelling techniques for studying the evolution of innate communication has a relatively mature literature now. Most researchers would probably agree that evolution by natural selection can tune a simple communication system under reasonable ecological assumptions. In this case, it could be argued that we can consider further work in this area to be superfluous. 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,

---

[5] For a more sophisticated analysis of the space of possibilities see (Noble, 1999) which considers varying the cost of signalling and issues such as honesty.

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 (2001) 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 in order to maximise 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 is their proximity sensors. Neither are there any obvious "meanings" in the simulation. However, communication evolves in Quinn's simulation. The typical behaviour of two evolved agents is as follows:

1. Both agents rotate anti-clockwise until it faces the other robot.
2. The first one to face the other moves forward to get closer to the other.
3. Once it is in range, it starts moving backwards and forwards 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 signalling system consisting of one signal — the oscillating movement 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 signalling — the sort of communication that is common in the natural world. It is striking, however, that human language is predominately very *unlike* one of these communication systems (even a cursory glance back at the list of features given in section 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 signalling 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 paper, 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.[6] 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 timescale) using genetic algorithms (Clark and Roberts, 1993; Hurford, 1999). Biological evolution is simply optimising the meaning-signal mapping to a communicative function (under the assumption that communication offers a selective advantage), and surely individual learning is performing an identical optimisation?

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 together (as pointed out in section 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 behaviour. 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 behaviour 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 (Brown and Hanlon, 1970).

Turning to cultural evolution, there have been many attempts to treat this and biological evolution as specific instantiations of the same mechanism (e.g., Cziko, 1995) ever since Darwin mentioned the parallels in the *Decent of Man* (Darwin, 1871). 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 Kirby, 1999 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, Smith, and Brighton (2002) point at, language change as an evolutionary process violates the central dogma of molecular biology (that there is no reverse translation from soma to germ line).

---

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

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 signalling 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 behaviours. 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 (Batali, in press; Batali, 1998). 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 modelling learning, rather surprisingly there is very little looking at this type of *iterated 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 facinating 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 (2001), Kirby and Hurford (2001), Brighton and Kirby (2001), and Brighton (2002) set out a generalised *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 in order to build a model of the population behaviour. 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 (modelling death) and the introduction of a new learner (modelling 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 behaviour is found in the natural world (since most, if not all, non-human signalling systems are innate). However, see Hurford (1989) and Hutchins and Hazelhurst (1995) for early examples.

One of the key features that differentiates the various artificial life simulations that broadly fall within the iterated learning 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 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 (1999) sets out a taxonomy of observational learning strategies based on how they change the language being transmitted by iterated learning:

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

Maintainance A learning strategy is capable of maintainance 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, maintainance, and construction underlines the importance of looking at iterated learning. On first looking at modelling learning organisms, the a-life researcher 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 maintainance 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.

By far the most thorough study of what is needed to be a constructor is Smith (2002). 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.[7] 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 retreive 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.

---

[7] Actually, it is common in IL models for reception to play very little 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. Modelling reception is still important, however, because without it the researcher cannot calculate the $ca(S, R)$ score for the population.

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 $4^3 = 81$ possible learning rules in total. These can be represented as a four-long vector:

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

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

By putting a homogenous population of agents with a particular learning rule in an iterated learning framework with a random initial language, Smith categorises 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 non-learners). 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 (Hebb, 1949). In a further study, Smith (2001b) also models a heterogeneous population of the various learning rules which is under biological selection pressure. This is an example of a *coevolutionary* model,[8] which are discussed in the context of syntactic systems in section 7. The main conclusion we can conclude from Smith's (2001b) research is that the interactions between learning, culture, and evolution are non-trivial. In particular, genetic drift plays a major role in ensuring that enough constructors dominate the population at some point in order 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 (1989) work in noting that construction is facilitated by learning algorithms which 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 (1997). Obverter can be thought of as a *bayesian* strategy for communication which is trying to maximise 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 $\sigma \in S$ which maximises $r(\sigma, \mu)$. In other words, obverters produce the signal which, 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 (1997) 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 behaviour 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 (although,

---

[8] In fact, the model can be seen as a combination of models of learning from Smith (2002) and cultural transmission from (Smith, in press).

we will return to this point later in this paper). 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, 2002; Steels, 1999) 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.[9] 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 papers, Stevan Harnard (e.g., Harnad, 1990) has focussed 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." (Harnad, 1990)

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.[10]

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., Steels and Vogt, 1997; Vogt, 2000; Billard and Dautenhahn, 1999), 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 (Cangelosi, Greco, and Harnad, 2000).

Cangelosi and Parisi (1998), 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 which 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.

---

[9] We have already seen an approach that does not assume a predefined set of meanings (or indeed, signals) in the work of Quinn (2001). However, his approach currently only works for evolving innate communication systems.

[10] 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 and solve everything at once. A sensible a-life 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 effect emergent languages in iterated learning) does not mention grounding, then the model need not solve the grounding problem.

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 recognise 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) which 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.[11] That said, the perceptual systems 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 (2001) 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?

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 (Steels, 1999), two robot controlled cameras try and communicate about a scene in front of them. The scene consists of a set of coloured 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, colour etc.) along which it may make discriminations amongst 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, i.e., 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

---

[11] 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 organised population. Oliphant (1997) 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 favours the emergence of an optimal system in this case.

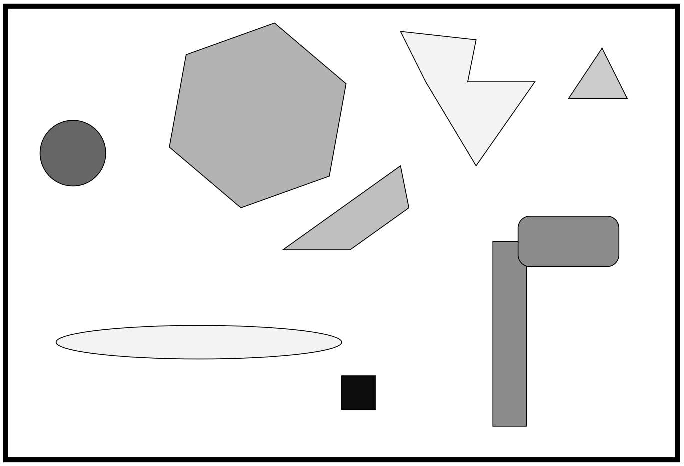

Figure 2: A schematic of a typical scene for the talking heads experiment. Note that the real scene would be subject to environmental conditions such as reflections and variation in light etc.

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 agents 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 in order to improve the chances of future successful communication. This can involve refining their discrimination trees and/or refining their lexica. For more details of the exact mechanisms, refer to the original literature.

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 realise that by not pre-defining 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 agent's to communicate perfectly about this topic but have quite different lexica. 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 colour, let alone have the colour *black* paired with this word, but despite this, communication is successful. Eventually, however, it is likely to fail (see Smith (2001a) for further analysis).

In Steels and Kaplan (1999) 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 apparantly 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? and 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 (2001) 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. Smith (2001a) 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 (Markman, 1989) — 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 colour-term universals by Berlin and Kay, 1969).

Whilst 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), none of the simulations described so far have dealt with the most strikingly unique aspects 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 characterise 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 characterisation of the general properties of human language syntax (see, for example Pinker & Bloom's (1990) list of features of language, or Newmeyer's (1992) explanation for polystratality of syntax), or by picking a principle from a particular model of syntax and studying that

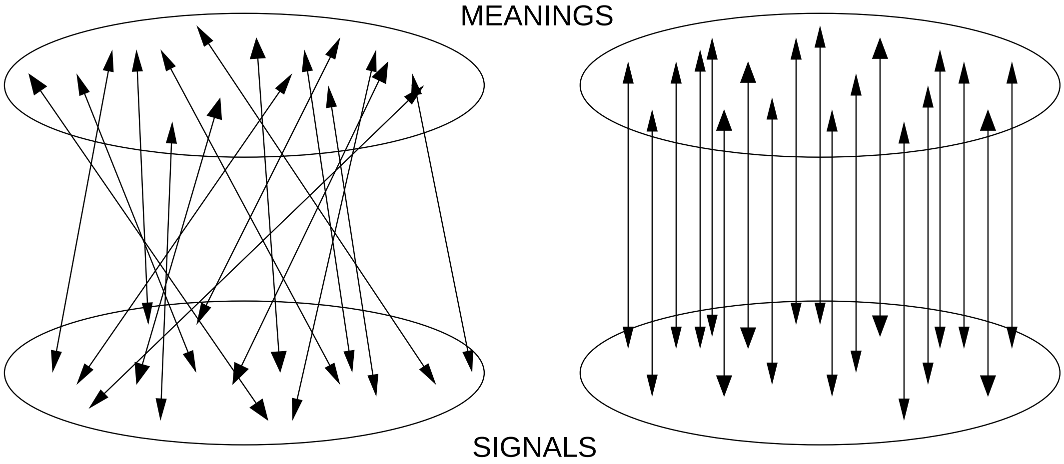

Figure 3: Two types of mapping between a signal space and a meaning space: holistic on the left, and compositional on the right.

in more detail (for example, both Newmeyer (1991) and Christiansen and Ellefson (in press) look at the Subjacency Principle from generative grammar).

In this section, I will take the former approach and attempt to characterise one of the most fundamental properties of human language syntax: compositionality. In linguistic semantics (e.g. Montague, 1970), 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.

This table shows the difference between compositional English sentences on the left with *holistic* equivalents on the right with similar meanings[12] (the last one is the Vervet alarm call for leopards — see, e.g., Cheney and Seyfarth, 1990).

Compositional	Holistic
walked	went
I greet you	Hi
died	bought the farm
I thought I saw a pussy cat	*bark*

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).

---

[12] Of course, in some circumstances, the phrase "bought the farm" is more compositional (i.e., when refering to someone buying a farm). Idioms generally appear to have compositional syntax, but in fact have a non-compositional 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.

A simple example of this kind of model is given in Kirby and Hurford (2001). 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 feedforward, 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 paper can be used. In order to produce a signal s from a meaning m, a network finds the signal that would maximise the likelihood of correct reception if the agent was talking to another agent identical to itself.

$$\begin{aligned}
&(1) & s_{\text{desired}} &= \operatorname*{argmax}_s P(s|m) \\
&(2) & &= \operatorname*{argmax}_s \frac{P(m|s)P(s)}{P(m)} \\
&(3) & &= \operatorname*{argmax}_s P(m|s) \\
&(4) & &= \operatorname*{argmax}_s C(m|s)
\end{aligned}$$

Where $C(m|s)$ is the confidence that the network has in the mapping $s \to m$. In other words, find the signal that maximises the network's confidence in the given meaning. In order to calculate $C(m|s)$, we treat the real-numbered network outputs $o[1 \ldots 8]$ as a measure of confidence in the the meaning bits $m[1 \ldots 8]$.

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

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

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

1. An initial population is setup consisting of two randomly initialised 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.[13]
5. The speaker is removed, the hearer is designated a speaker, and a new hearer is added (with randomly initialised weights).

---

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

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 behaviour with this model:

1. With low b, the language is unstable and unexpressive.
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 generaliseable (see Hurford, 2000). For the network, it is trivial to form generalisations 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 generalisable decreases, and the network can tolerate languages that involve a degree of memorisation.

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 being due to languages adapting to the learning task. However, there are an increasing number of models in the literature which replicate this basic result (see Hurford, 2002, for a deeper review):

Batali, 1998: 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,[14] but the agents converge on a system of strings that are made up of substrings that reflect the meaning-space structure. The result is remarkably reminiscent of the morphology of natural languages. These results are replicated in Goroll (1999).

Kirby, 2000: 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

---

[14] Tonkes (2002) classifies this simulation and Batali's other work as a *negociation model* in contrast with the *iterated learning model.* This difference in terminology reflects the fact that Batali has no generational turnover.

network. Once again, the language evolves such that utterances are made by combining words into sentences. In order 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, in press: 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, in press: 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 generalises through analogy from structures the agent has memorised. Some interesting structures emerge that are reminisent 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: 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: In all the previous papers, every meaning is equally probable. This paper demonstrates that a skewed distribution over meanings results in a different result. If speakers prefer short strings, then languages emerge which 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: This paper employs the MDL approach of Teal and Taylor (1999) but with a modified form of finite state machine that is able to represent meanings. These are used to predict the cirumstances under which compositionality will be stable. The result (confirmed in the mathematical study of Brighton (2002)) is that compositionality is more stable with meaning spaces of high dimensionalty.

Zuidema, 2001: A learning algorithm similar to the one of Kirby (2000) is used to demonstrate that the iterated learning framework moves the lower bound result of Nowak, Komarova, and Niyogi (2001) enabling stable languages with fewer exposures to the data. Zuidema (2001) also shows that iterated learning mitigates some classic learnability results such as Gold (1967).

Kirby, Smith, and Brighton, 2002: In order 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 and Kirby (2001) 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 Smith (2002).

Not all computational modelling 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 (2002) explores which languages fit a particular learning bias by a hill-climbing approach as well as allowing those languages to emerge naturally in an iterated learning model. Perfors (2000) uses genetic programming techniques in order to study the conditions for the emergence of structured goal-directed communication.

Hashimoto and Ikegami (1996) evolve grammars directly to understand the relationship between their formal properties and selection pressures based on *speaking*, *recognizing* and *being recognised* (see also, Zuidema, 2000 for review). Another approach is exemplified by Hashimoto (1999) which explores a "constructivist" theory of to linguistic structure in which a web of statistical relationships arise out of the dynamics of interactions between agents. More recently, Hashimoto has linked these two models, which he says reflect different timescales of linguistic evolution. Ultimately, this kind of synthesis of local and global dynamics may prove the most fruitful direction for future research (see Hashimoto, 2001 for an introduction).

For many, modelling 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 artificial life 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 (in press) puts it in his critique of computational/mathematical models (or modellers), "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 lexica of all languages (Cann, 2000). 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 (1999) looks at the link between learning and processing constraints and universal properties of language using an iterated learning approach. The universals examined relate to: word-order, the formation of relative clauses, patterns of agreement (that is, 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 similar spirit, although not employing a "full" iterated learning model, is the work by Christiansen and Devlin (1997) that relates the learning biases of recurrent neural networks to some of the linguistic universals listed above. Ultimately, if artificial life modelling 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? and 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 signalling systems. More specifically linguistic behaviour, 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 iterated learning 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, suffice to say that there are two main areas of controversy. Firstly, there is disagreement about whether the innate biases are domain specific or domain general, and secondly, about how these biases come about. Obviously these two questions are related. For example, one might take the stance (as Pinker and Bloom (1990) do) that innate constraints[15] 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 modelling 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 it is often hard to predict the results of such interactions in advance of modelling them.

In particular, the interaction of learning and evolution is a "classic" topic for artificial life models (Belew and Mitchell, 1996), and is also highly pertinent to the evolution of language. Batali (1994) is 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^nb^n. Initial experiments showed that the networks were unable to learn the task well. In particular, they appeared to overgeneralise to the language a^nb^m. Batali hypothesised 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 backpropogation and search for initial weights through an evolutionary algorithm resulted in correct learning of the class a^nb^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 generalise his results to a *class* of languages, Batali repeated the experiment with a set of language problems that all had the

---

[15] 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*.

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 successfully learn 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 dysanalogy 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 iterated learning 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 by Turkel (in press) (itself a modification of Hinton & Nowlan's (1987) classic model), Kirby and Hurford (1997) 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 (Chomsky, 1981). 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 which best model the data seen (the learning procedure used in Kirby and Hurford (1997) is the trigger learning algorithm of Gibson and Wexler (1994)). In the simulation, the 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.

In order 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 neighbouring agents. The results showed partial genetic assimilation of parameters into principles, and a concequent 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 "nativisation" of paramters into principles lead 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 harder 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 non-trivial 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 (2000a) attempts to make coevolutionary models closer to those studied in mainstream linguistics. The simple vectors of (Kirby and Hurford, 1997) are replaced by Briscoe with an actual theory of syntactic variation: generalised 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 creolisation, 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 (2001) 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.[16] 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 paper, 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.

One way in which I hope this review will be useful is to direct future research using simulation techniques. In other words, it is the *gaps* in the literature that may eventually prove to be the most rewarding (but perhaps also, the most difficult) areas of study. There are two areas which I feel will be particularly fruitful:

Coevolution of syntax and semantics in the ILM So far, the results of running iterated learning models suggests 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 learnt, and vary from agent to agent. Can structure in meanings and signals both emerge in a simulation? The logic of the ILM suggests that they might, since only by matching the structure of the signal with the structure of the meaning, can languages optimise their transmissability.

---

[16] These issues are explored further in Briscoe (in press).

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 iterated learning framework, or indeed the models of grammatical assimilation.

A possible answer to this puzzle lies in understanding what it is necessary to build into an ILM in order 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 creolisation). 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 exquisitly attuned to the thought processes of others (Bloom, 2000). 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 – e.g., universal properties of 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 (Christiansen and Devlin, 1997; Christiansen and Ellefson, in press).

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 embarrasment 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.

Acknowledgements

Much of the work reported here was carried out by the various researchers at the Language Evolution and Computation Research Unit in the Department of Theoretical and Applied Linguistics, University of Edinburgh, namely: Henry Brighton, Takashi Hashimoto, James Hurford, Joseph Poulshock, Andrew Smith, Kenny Smith, and Hajime Yamauchi. I am indebted to them for helping me to understand their varied perspectives, and putting up with my constant nagging for new results and new simulations. I hope I have not misrepresented their work (and the work of everyone else) too badly.

References

Ackley, D. and M. Littman (1994). Altruism in the evolution of communication. In R. Brooks and 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.

Batali, J. (1994). Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In R. Brooks and 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.

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

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: Cambridge University Press.

Belew, R. and M. Mitchell (Eds.) (1996). *Adaptive Individuals in Evolving Populations*. Addison-Wesley.

Berlin, B. and P. Kay (1969). *Basic color terms: their universality and evolution*. University of California Press.

Bickerton, D. (1981). *Roots of Language*. Karoma.

Bickerton, D. (in press). Rooting innateness in evolution. In M. Christiansen and S. Kirby (Eds.), *Language Evolution: the States of the Art*. Oxford University Press.

Billard, A. and K. Dautenhahn (1999). Experiments in learning by imitation - grounding and use of communication in robotic agents. *Adaptive Behavior 7*(3/4), 415–438.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. MIT Press.

Bowerman, M. and S. Choi (2001). Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman and S. C. Levinson (Eds.), *Language Acquisition and Conceptual Development*, pp. 475–511. Cambridge University Press.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artifical Life 8*(1).

Brighton, H. and S. Kirby (2001). The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In J. Kelemen and P. Sosik (Eds.), *Advances in*

Artificial Life (Proceedings of the 6th European Conference on Artificial Life). Heidelberg: Springer-Verlag.

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

Briscoe, E. (in press). Grammatical assimilation. In M. Christiansen and S. Kirby (Eds.), *Language Evolution: the states of the art*. Oxford University Press.

Briscoe, T. (2000b). Evolutionary perspectives on diachronic syntax. In S. Pintzuk, G. Tsoulas, and A. Warner (Eds.), *Diachronic Syntax: Models and Mechanisms*. Oxford: Oxford University Press.

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

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

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

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

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

Cann, R. (2000). Functional versus lexical: a cognitive dichotomy. In R. D. Borsley (Ed.), *The Nature and Function of Syntactic Categories*. Academic Press.

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

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

Chomsky, N., A. Belletti, and L. Rizzi (2000). An interview on minimalism. University of Siena Online Publication. http://www.cicl.unisi.it/pubblicazioni.php.

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

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

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

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

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

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

Cziko, G. (1995). *Without Miracles: universal selection theory and the second Darwinian revolution.* MIT Press.

Darwin, C. (1871). *The Descent of Man, and selection in relation to sex.* London: John Murray. Reprinted in 1981 by Princeton University Press.

de Boer, B. (2001). *The Origins of Vowel Systems.* Oxford: Oxford University Press.

Dennet, D. (1987). *The Intentional Stance.* MIT Press.

Di Paolo, E. (1997). An investigation into the evolution of communication. *Adaptive Behaviour 6*, 285–324.

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

Epstein, J. M. and R. Axtell (1996). *Growing Artificial Societies: Social science from the bottom up.* MIT Press.

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

Gold, E. M. (1967). Language identification in the limit. *Information and Control 10*, 447–474.

Goroll, N. (1999). (the deep blue) nile: Neuronal influences on language evolution. Master's thesis, University of Edinburgh.

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

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

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

Hashimoto, T. (1999). Modeling categorization dynamics through conversation by constructive approach. In D. Floreano, J.-D. Nicoud, and F. Mondada (Eds.), *Advances in Artifical Life*, pp. 730–734. Springer.

Hashimoto, T. (2001). The constructive approach to the dynamical view of language. In A. Cangelosi and D. Parisi (Eds.), *Simulating the Evolution of Language*, Chapter 14, pp. 307–324. London: Springer Verlag.

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

Hauser, M. D. (1996). *The Evolution of Communication.* Cambridge, MA: MIT Press.

Hebb, D. (1949). *The Organization of Behavior.* Wiley.

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

Hurford, J. (1999). Artificially growing a numeral system. In J. Gvozdanovic (Ed.), *Numeral Types and Changes Worldwide*, pp. 7–41. Mouton de Gruyter.

Hurford, J. and S. Kirby (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.* Lawrence Erlbaum.

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

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

Hurford, J. R. (2000). Social transmission favours linguistic generalization. In C. Knight, M. Studdert-Kennedy, and J. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 324–352. Cambridge: Cambridge University Press.

Hurford, J. R. (2002). Expression/induction models of language evolution: dimensions and issues. In T. Briscoe (Ed.), *Linguistic Evolution Through Language Acquisition: Formal and Computational Models.* Cambridge: Cambridge University Press.

Hutchins, E. and B. Hazelhurst (1995). How to invent a lexicon: the development of shared symbols in interaction. In N. Gilbert and R. Conte (Eds.), *Artificial societies: the computer simulation of social life.* london: UCL Press.

Johansson, C. (1997). *A View From Language.* Lund University Press.

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

Kirby, S. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy, and J. R. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 303–323. Cambridge: Cambridge University Press.

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.

Kirby, S. (in press). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge: Cambridge University Press.

Kirby, S. and J. Hurford (2001). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi (Eds.), *Simulating the Evolution of Language.* Springer Verlag.

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

Kirby, S., K. Smith, and H. Brighton (2002). Language evolves to aid its own survival. In preparation.

Langton, C. (Ed.) (1995). *Artificial Life: an overview.* MIT Press.

Lenneberg, E. (1967). *Biological Foundations of Language.* Wiley.

Lightfoot, D. (1999). *The Development of Language: Acquisition, Change, and Evolution.* Oxford: Blackwell.

Lindgren, K. and M. G. Nordahl (1995). Cooperation and community structure in artificial ecosystems. See Langton (1995), pp. 39–60.

Maclennan, B. and G. Burghardt (1994). Synthetic ethology and the evolution of cooperative communication. *Adaptive Behaviour 2*, 161–187.

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

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.

Nagel, K. and S. Rasmussen (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*, pp. 222. MIT Press.

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

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

Niyogi, P. and R. Berwick (1995). The logical problem of language change. Technical Report AIM-1516, MIT AI Lab.

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

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

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

Oliphant, M. (1997). *Formal approaches to innate and learned communication: Laying the foundation for language.* Ph. D. thesis, UCSD.

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior 7*(3/4), 371–384.

Oliphant, M. (in press). Learned systems of arbitrary reference: the foundation of human linguistic uniqueness. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge: Cambridge University Press.

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

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

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

Prusinkiewicz, P. (1995). Visual models of morphogenesis. See Langton (1995), pp. 61–74.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In J. Kelemen and P. Sosik (Eds.), *Advances in Artificial Life: Proceedings of the* 6^{th} *European Conference on Artifical Life, Prague, September 2001.* Springer-Verlag.

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.* Oxford: Oxford University Press.

Savage-Rumbaugh, S. and R. Lewin (1994). *Kanzi: the ape at the brink of the human mind.* Wiley.

Schuster, P. (1995). Extended molecular evolutionary biology: Artificial life bridging the gap between chemistry and biology. See Langton (1995), pp. 39–60.

Smith, A. D. M. (2001a). Establishing communication systems without explicit meaning transmission. In J. Kelemen and P. Sosik (Eds.), *Advances in Artificial Life: Proceedings of the 6^{th} European Conference on Artifical Life, Prague, September 2001.* Springer-Verlag.

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

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. Submitted.

Smith, K. (in press). Learners are losers: Natural selection and learning in the evolution of communication. *Adaptive Behaviour*.

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

Steels, L. (1999). *The Talking Heads Experiment*, Volume I. Words and Meanings. Antwerpen: Laboratorium. Special pre-edition.

Steels, L. and F. Kaplan (1999). Collective learning and semiotic dynamics. In D. Floreano, J.-D. Nicoud, and F. Mondada (Eds.), *ECAL99*, pp. 679–688. Springer-Verlag.

Steels, L., F. Kaplan, A. McIntyre, and J. Van Looveren (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The Transition to Language.* Oxford, UK: Oxford University Press.

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

Teal, T. and C. Taylor (1999). Compression and adaptation. In D. Floreano, J. D. Nicoud, and F. Mondada (Eds.), *Advances in Artificial Life*, Number 1674 in Lecture Notes in Computer Science. Springer.

Tonkes, B. (2002). *On the Origins of Linguistic Structure: Computational models of the evolution of language.* Ph. D. thesis, University of Queensland.

Turkel, W. J. (in press). The learning guided evolution of natural language. In T. J. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models.* Cambridge: Cambridge University Press.

Vogt, P. (2000). *Lexicon Grounding on Mobile Robots.* Ph. D. thesis, Vrije Universiteit Brussel.

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

Wilson, E. (1975). *Sociobiology: the new synthesis.* Harvard University Press.

Wonnacot, E. and S. Kirby (2002). Emergent linguistic dynamics from simple mechanisms: A microscopic-modeling approach to logistic change. In preparation.

Yamauchi, H. (2001). The difficulty of the baldwinian account of linguistic innateness. In J. Kelemen and P. Sosik (Eds.), *Advances in Artificial Life (Proceedings of the 6th European Conference on Artificial Life)*. Heidelberg: Springer-Verlag.

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

Zuidema, W. (2001). Emergent syntax: The unremitting value of computational modeling for understanding the origins of complex language. In J. Kelemen and P. Sosík (Eds.), *Advances in Artificial Life*, Number 2159 in Lecture Notes in Artificial Intelligence, pp. 641–644. Berlin: Springer.