The Role of Anticipation in the Emergence of Language

Samarth Swarup[1] and Les Gasser[1,2]
[1]Department of Computer Science,
[2]Graduate School of Library and Information Science,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA
{swarup,gasser}@uiuc.edu

Abstract

We review some of the main theories about how language emerged. We suggest that including the study of the emergence of artificial languages, in simulation settings, allows us to ask a more general question, namely, *what are the minimal initial conditions for the emergence of language*? This is a very important question from a technological viewpoint, because it is very closely tied to questions of intelligence and autonomy. We identify anticipation as being a key underlying computational principle in the emergence of language. We suggest that this is in fact present implicitly in many of the theories in contention today. Focused simulations that address precise questions are necessary to isolate the roles of the minimal initial conditions for the emergence of language.

1 What is the problem of language emergence?

It is very hard to imagine what life would be like without language. Before some point in our evolutionary history, however, our ancestors did not have language. How did language (and the capacity for it) evolve? This is the problem of language emergence.

The emergence of language is considered to be the last major transition in evolution [1]. It is one of the clearest distinctions between humans and other animals, and speculations about the origins of language go back to Plato's *Cratylus* dialogue, which discusses the connection between names and things. What makes the subject particularly difficult is the lack of data about the earliest languages. Despite this lack, the publication of Darwin's works on evolution lead to a great deal of speculation on possible scenarios for the evolution of language. This led the *Société de Linguistique de Paris*, when it was formed in 1865, to declare in its bylaws that it would not accept any communications dealing with the origin of language. A similar statement was made by the Philological Society of London in 1873 [2].

In the last fifty years or so, however, the question has again gained scientific validity due to relevant discoveries in archaeology, anthropology, and neuroscience. A lot more is now known about the biology, environment, and lifestyles of the early *homo* species. This has lead to a renewed spate of theories about the origins of language. It is the aim of this article to review some of the main contenders, and to address in particular the role of anticipation in the emergence of various aspects of language.

We start by describing the problem of the emergence of human language. Then we suggest that by expanding this question to ask what the minimal initial conditions for the emergence of language are, we can build a more general theory which will provide us with a better understanding of how to design systems that can create their own language. We lay out a space of communication systems, and analyze how the notion of anticipation can be used to build a framework to study the movement from simple to progressively more complex communication systems. After that we examine some of the main theories of the emergence of human language, and some of the work in artificial language evolution (through simulation). We find that these have been addressing different regions of the communication systems space. However, we can use these to infer some basic conditions for the emergence of various kinds of communication systems already, and by building an anticipatory framework we can provide the scaffolding for further simulations that will deal with more complex forms of language.

1.1 What form does an answer to this problem take?

Theories of the origins of language address two questions: how language evolved, and why. Broadly, the answers to how language evolved consist of speculations on the mechanisms, or *preadaptations*, that made language possible, and the stages that lie between animal-like signaling and modern human language. The answers to why language evolved consist of speculations on the functional properties of language, environmental conditions, and selection pressures that gave language an adaptive advantage.

There are a couple of important points to remember here. First, the various proposals for why language evolved are not mutually exclusive. Indeed it is likely most of these contributed to the selection pressure for the evolution of language. Box 1 summarizes the main ideas about why language evolved.

Second, any postulated preadaptations for language must be selected for in their own right. This means that we cannot suppose that some preadaptation emerged in order to make language possible. Evolution does not proceed according to some pre-specified program, and therefore such a suggestion would violate causality.

An example of a preadaptation is the change in the shape and robustness of the jaw which made possible, as a side effect, the production of the range of speech sounds we enjoy today. This happened when *homo ergaster* moved from the arboreal habitats occupied by the australopithecines to a more open savannah habitat. This led to a change in diet from being predominantly vegetarian to incorporating more animal-based products. This in turn led to the change in the shape and robustness of the jaw [3].

Box 1: Functional Scenarios for the Evolution of Language
Johansson provides a nice overview of the various scenarios that have been proposed to have provided the selection pressure for the emergence of language [4]. We list them here.

1. Hunting, which leads to a pressure for a language for cooperation.
2. Tool-making which, arguably, lead to an increase in intelligence, and provided the mental capabilities required for language (such as combinatoriality).
3. Sexual selection, such as a preference for more articulate mates, or sexual conflict as a driving force, or because better communicative ability can lead to social/political power.
4. Child-care and teaching, which leads to a pressure for a language for teaching.
5. Social relations in groups and tribes:
    - Predation, perhaps for coordination for group defense.
    - Inter-group competition.
    - Intra-group competition for resources.
    - Mating opportunities.
    - Intra-group aggression and politics, such as alliance-formation, negotiation, etc.
6. Children at play, where language may have appeared through mimicry, for example.
7. Music.
8. Story-telling.
9. Art.

A *complete* theory of the emergence of human language would need to answer at least the following questions:

- Why have only humans developed language?
- Is it due to a difference in degree, or a difference in kind?
- How much of language is innate, and how did it become so?
- Did language emerge gradually, and if so what did earlier forms of language

look like?

1.1.1 Why have only humans developed language?

Szathmáry has suggested that there can be two possible reasons for the uniqueness of an adaptation: it might be *variation-limited* or *selection-limited*. Being variation-limited means that the necessary mutations occur extremely rarely. Being selection-limited means that they only confer a selective advantage in extremely rare conditions [4, quoted]. Hurford has pointed out, however, that just because other species haven't developed language doesn't mean they won't [5]. Language has only emerged in the last 100,000 to 500,000 years [3], which is a short while for evolution. Every major evolutionary transition must have had a vanguard - a species that was the first to achieve it, and solely enjoyed its benefits until the other species caught up.

1.1.2 Is it due to a difference in degree, or a difference in kind?

A counter-argument to many of the scenarios listed in box 1 is that other species do them too. Why haven't they developed language? Hunting, for example, is a very common activity in the animal kingdom. Even cooperative hunting, which is proposed to have provided the selective pressure for communication, is quite common. The question, then, is, are these viable propositions? Is a difference in the degree to which we engage in some activity, for example our increased period of childhood, or our increased social group size, sufficient to explain why language evolved? Or is there a different *kind* of activity we engage in, that other species don't, that led to the evolution of language? The same question holds for our cognitive capabilities. Does the emergence of language require some special cognitive capability that other animals lack, or is it that we are just better at (some aspects of) cognition?

1.1.3 How much of language is innate, and how did it become so?

It is hard to argue that there aren't at least some aspects of language which are innate. The capacity for symbolization is probably innate. Further, children can acquire a grammatical language even if the linguistic input they receive is not grammatical, as in the emergence of creole languages from pidgins, and in the famous example of the Nicaraguan Sign Language, where a community of deaf children in a school in Nicaragua invented a grammatical sign language based on the pidgin-like *Lenguaje de Signos Nicaragüense* that they were exposed to at home [6]. This does not necessarily mean that grammar is innate, however. For one thing, the development of a creole seems to depend on the size of the community. If the community is not large enough, a grammatical language does not emerge.

The idea that we might have an innate language acquisition device (or a universal grammar), which appeared by means other than natural selection, was first proposed by Chomsky [7]. It has been extremely controversial [8], and in one of his most recent articles, he (with Hauser and Fitch) proposes that the only aspect of grammar that is innate is the ability to do recursion [9]. This proposal, also, has generated debate [10].

1.1.4 Did language emerge gradually, and if so, what did earlier forms of language look like?

An idea that seems to find general agreement is Bickerton's proposal of a *protolanguage* [11]. A protolanguage is basically modern language without the rich syntax. It is compositional, that is, it consists of words that are strung together into sentences, but it doesn't have properties like tense and aspect. It is also supposed to have a closed (that is, fixed) vocabulary. Bickerton has proposed that modern language was preceded by protolanguage, which may have existed for as many as a million years before modern language appeared, and further, that protolanguage still makes its appearance in pidgins, and in some aspects of language acquisition (a twist on "ontogeny recapitulates phylogeny"). Jackendoff has expanded on the idea of a protolanguage, suggesting several different stages. These are summarized in box 2. Johansson provides a nice summary of all the "protos" that make up protolanguage: proto-speech, proto-gestures, proto-semantics, and proto-syntax [12].

Box 2: Proto-language

Jackendoff has postulated the following stages in the evolution of the language capacity [13]. Bickerton's proposed protolanguage [11] is subsumed in this sequence.

1. The use of symbols in a non-situation-specific fashion.
2. An open, unlimited class of symbols.
3. A generative system for single symbols: proto-phonology.
4. Concatenation of symbols to build larger utterances.
5. Using linear position to signal semantic relationships.
6. Phrase structure.
7. Vocabulary for relational concepts.
8. Beyond phrase structure: inflection and further syntax.

Computer scientists have only recently become interested in the question of language emergence, partly because we believe that some of these issues can be addressed through agent-based simulation. However, we believe that in this case the appropriate question is slightly different, and more general.

2 A modified question: What are the minimal initial conditions for the emergence of language?

Part of the problem with trying to explain the emergence of language is that language is unique. No other species has evolved language, and so any explanation is going to be a “just so” story. However, when we include *artificial* language evolution[1] in the mix, we can ask the more general question, *what are the minimal initial conditions for the emergence of language?* One important thing to keep in mind is that the minimal conditions are not just cognitive, but also environmental. Another way of asking the same question is, what mechanisms/conditions do we need to design/provide to enable the emergence of language in a population of machines?

This is an important question because, besides being an important scientific problem, the study of the emergence of language is also very important from a technological perspective. Multi-agent systems are becoming increasingly widespread, being used in widely differing contexts such as spacecraft control, military mission scheduling, auctions, agent-based models of social networks and organizations, etc. The general approach to communication and coordination in multi-agent systems is to pre-impose a designed language. However, such predefined languages are often found to be inadequate, especially as multi-agent systems increase in size and complexity, as they reflect the designer’s viewpoint rather than the agents’, and are unable to adapt to changing environmental conditions and task definitions. It is much more desirable for the agents to be able to create and maintain their own language.

The last decade has seen increasing application of computational and mathematical methods to the study of language evolution (see [14] for a recent review). This has led to important advances on questions such as how a shared language is established in a population [15], [16], [17], the emergence of syntax [18], and symbol grounding [19], [20]. However, we are still far from a general theory.

2.1 What form does an answer to the modified question take?

In order to construct such a general theory, we need to map out the space of possible communication systems, and analyze the factors that lead to the emergence of these. Figure 1 shows one possible way to lay out this space, and where some commonly considered communication systems in the language evolution literature would lie in this space.

Continuous and discrete communication systems are distinguished along the x-axis. A continuous communication system is one that uses the magnitude

---

[1] Note that by artificial languages, we do not mean those constructed by humans, such as Esperanto and Klingon. Rather, we are referring to attempts to evolve a language in a population of (simulated or real) agents.

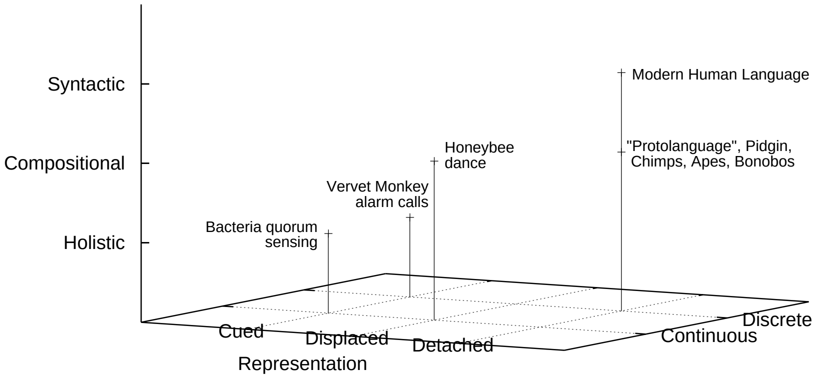

Figure 1: The space of communication systems.

of some quantity to communicate information. For example, scout honeybees communicate the quality of a nest-site they have discovered by the vigor of their waggle dance [21]. Bacteria do quorum sensing by producing molecules called autoinducers [22], [23]. Quorum sensing is the control of gene expression in response to cell density. This means, for example, that bacteria will often not express virulence factor until their colony is big enough to have a high probability of successfully infecting the host.

The y-axis distinguishes the kind of representation used by the communication system. The notion of cued and detached representations is due to Gärdenfors [24]. In his own words, "A *cued* representation stands for something that is present in the current external situation of the representing organism. When, for example, a particular object is categorized as food, the animal will then act differently than if the same object had been categorized as a potential mate... In contrast, *detached* representations may stand for objects or events that are neither present in the current situation nor triggered by some recent situation." Since the honeybee representation of food sources or nest sites seems to fall inbetween cued and detached, we include *displaced* representations in our space of communication systems. By displaced representation, we mean representations which stand for objects or events that are not in the current situation, but have been triggered by some recent situation. The notion of displacement is one of Hockett's design features of language [25], but he uses the term to mean anything other can cued representations[2]. We think it is important to make these distinctions because they imply different computational properties of the underlying cognitive system. Deacon talks about a similar taxonomy of kinds of symbols: *iconic*, *indexical*, and *true symbols* [26]. Icons physically resemble that which they represent, for example onomatopoeic words like "pitter-patter". Indices involve correlations between the symbol and the referent, for example a symptom and a disease. True symbols, in contrast to the other types, are entirely arbitrary. For example, the word "chair" doesn't tell us anything about the (kind of) object to which it refers.

The z-axis distinguishes various levels of structure that might be present in the communication system. The simplest kind of communication in this sense is *holistic*, where every "meaning" or concept has an independent symbol associated with it. There is no relation between the symbols, and no internal structure to them. The classic example is the alarm call system of vervet monkeys [27]. Vervet monkeys have different calls for flying predators like eagles, and ground predators like snakes and leopards. These calls have no relation to each other. They don't, for example, have a common component that means "predator".

A significantly more complex form of structural organization is *compositional* language. This means that utterances are composed of meaningful parts, which combine meaningfully. For example, "green ball" means not just that there is something green and something that is a ball, but that it is the ball that is green. The language capacity of chimps, bonobos, and apes seems to be at this level [28].

Finally, the most complex form of structural organization we know is modern human language with its rich syntax.

The communication space gets increasingly complex along each axis. Thus the simplest communication system is continuous, cued, and holistic, such as the quorum sensing of bacteria, and the most complex is discrete, detached, and syntactic, of which the only known example is modern human language. We will see next that this space correlates well with kinds of anticipation. This brings up the central question of this article.

3 What is the role of anticipation in the emergence of language?

Anticipation is widely considered to be a very important component of cognition. The notion of anticipation is closely related to prediction or expectation. To put it in a sentence, expectation is knowledge about the future, and anticipation is what you do with it. Robert Rosen has defined an anticipatory system to be one that has an internal model of itself and/or its environment, which it uses in planning (or action selection) [29]. The "model", of course, can be of varying degrees of complexity, from simple stimulus-response to systems with complex

---

[2] Note that what we call displaced, Gärdenfors would probably include under cued, and what Gärdenfors calls detached, Hockett would include under displaced.

internal states. The most famous experimental demonstration of anticipatory behavior is Pavlov's dog, which learned to anticipate food at the sound of a bell, and showed this by starting to salivate.

We argue below that anticipation provides a very nice framework for studying cognitive requirements for language, because it is correlated with language: the more sophisticated the anticipatory behavior exhibited by a population, the more complex is their communication system.

Butz et al. have described four kinds of anticipation in relation to adaptive behavior [30]:

- Implicit anticipatory behavior
- Payoff anticipatory behavior
- Sensory anticipatory behavior
- State anticipatory behavior

Implicit anticipation corresponds to the situation where the agent is not explicitly computing expectations, but still exhibits some anticipatory behavior. The anticipation, in this case, has been carried out by evolution (or the designer, for artificial agents), by equipping the agent with a genome that will "work well" in its environment. There is no learning beyond that done through evolution, since learning is essentially equivalent to prediction. Bacteria are among the simplest kinds of implicitly anticipatory agents, though admittedly they blur the distinction between learning and evolution through horizontal gene transfer [31].

Payoff anticipation consists of forming expectations of rewards for states of the environment, and utilizing these expected rewards during planning. The simplest kind of reinforcement learning, called model-free reinforcement learning [32], is an example of payoff anticipation because it computes a value function which is simply the expected cumulative discounted future reward for each state, and then the agent chooses actions which take it to states with high values. Honeybees could be considered to be payoff anticipatory agents, because it is unlikely that they have a predictive model of the environment in their heads that they use for planning. They also exhibit associative learning (that is, classical conditioning), which again requires payoff anticipation. They are also capable of learning some other things, however, such as landmarks and other cues that they use for navigation on their foraging trips [33]. Vervet monkeys are probably capable of more sophisticated anticipatory behavior, but the kind of anticipation required for their alarm call system is only payoff anticipation.

Sensory anticipation involves predictions that do not influence behavior directly, but only the sensory processing of the agent. It is strongly related to phenomena like priming, where a particular sensory input causes enhanced attention to a subsequent sensory input to the point that it can be hallucinated. This kind of anticipation is probably mediated by top-down connections in the sensory pathways, that carry predictions of expected sensory input. This kind

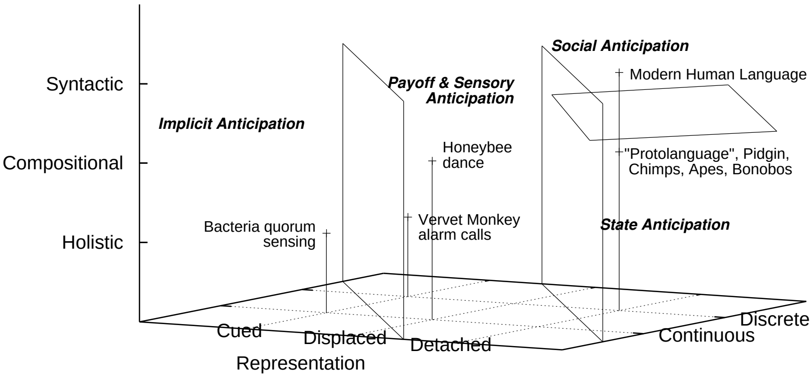

Figure 2: The space of communication systems, partitioned by kinds of anticipation required.

of neural architecture is quite widespread, and probably honeybees as well as vervet monkeys are capable of sensory anticipation. It is thought that auditory anticipation is an important part of speech comprehension in humans, and so our linguistic capacity would probably be degraded if we were not capable of sensory anticipation. However, it probably is not one of the minimal initial conditions for language.

State anticipation involves having a detailed predictive model of the environment, which directly influences decision-making. Agents which have representations of goals, and perform mental simulations of actions to come up with a plan for reaching the goal from their current state are performing state anticipatory behavior. For example, a chimp that comes across a termite nest and then goes to find a stick from which it strips off the leaves, and then returns to the termite nest and uses the stick to fish out termites to eat is clearly performing state anticipatory behavior.

A fifth kind of anticipation, social anticipation, is now considered to be distinct from the other four. We believe that social anticipation is qualitatively different from other kinds of anticipation. In other kinds of anticipatory behavior, the environment is considered to be stationary. This means that, theoretically, environmental inputs are assumed to be generated by stationary distributions. Social anticipation, however, takes place in an “environment” consisting of other agents that are also performing social anticipation. This makes the environment non-stationary, which presents a much harder learning problem. We are not suggesting, however, that social anticipation could not have evolved from the previous kinds of anticipation. Unpacking how it might have done so however, is beyond the scope of this paper. It would involve, perhaps, anticipation in environments which are mostly stationary, for example where animals learn very occasionally or very slowly. It would also involve unpacking *degrees* of social anticipation. For instance, animals such as chimps and bonobos are capable of social anticipation to a degree perhaps, because they have to vie for dominance in their groups in order to achieve better feeding and mating opportunities. However, it is not clear to what degree they have a definite *theory of mind*, that is, a general understanding of how other members of their own species behave. It has been argued by several researchers that it is this theory of mind, or social intelligence, that sets us apart from other animals, and that brought about the emergence of language [34], [35].

Figure 2 shows how kinds of anticipation partition the space of communication systems. In the figure, it should be understood that agents that are capable of a particular kind of anticipation are also capable of all the previous kinds of anticipation, but not subsequent ones. For example, honeybees are capable of implicit, payoff, and sensory anticipation, but not state or social anticipation.

Now, keeping this analysis in mind, we take a look at some of the theories and simulations of language emergence.

4 What are some of the theories of the emergence of language?

There are three basic questions:

- Why/How did the capacity for symbolization evolve?
- Why/How did structured language (that is, compositionality and syntax) evolve?
- Why/How does a population converge upon a common language?

Various hypotheses have been put forward to answer one or more of these questions. We review some of the main ones below.

4.1 Natural Language Evolution

4.1.1 Anticipatory Planning

To our knowledge, Gärdenfors and his colleagues are the only ones who have explicitly invoked anticipation in theorizing about the emergence of language [24], [36], [37].

Gärdenfors and Osvath [36] have argued that Olduwan culture led to the evolution of anticipatory cognition, which in turn led to the emergence of symbolic communication. Olduwan culture is a term used to refer to the use of stone tools by pre-historic hominins[3], roughly in the period 2.6 to 1.5 million years ago. Further, they claim that this sort of cognition is unique to humans. Once anticipatory cognition appeared, it made communication about future goals beneficial by enabling long-term planning.

Various animal species, particularly primates, are known to be capable of planning. However, they argue, the plans of other animals always address present needs, while humans are the only animals capable of planning for future needs, that is, *anticipatory* planning. For example, a chimp may go look for a stick on finding a termite nest, as mentioned earlier, however a chimp won't spend its time putting together a collection of sticks to carry around in anticipation of finding a termite nest.

The reason that Olduwan culture is presumed to have given rise to anticipatory planning is because it appeared at a time in our past when our hominin ancestors had to make the transition from a forest environment to a savannah. Food sources are much more scarce on the savannah, and this led to a number of changes. The dietary change and its effect on the jaw has already been mentioned. The scarcity of food also meant, however, that the hominins had to range farther in search of food. This makes it beneficial to carry along tools (for dressing meat etc.) rather than trying to bring a carcass all the way back home. It might also be a good idea to make caches of tools at various hunting locations, so that tools don't have to be carried everywhere. These kinds of behavior require anticipatory planning, though, because the hominin would have to fashion tools and carry them to caches without being cued by the need to hunt or scavenge. In other words, they would have to plan for anticipated *goals*.

This is a very sophisticated kind of planning. Butz et al. lay out their framework in the context of a partially-observable Markov decision process (POMDP). A POMDP contains a reward function, which essentially corresponds to a goal. An agent that is solving a POMDP is trying to maximize reward. The *anticipation* of goals can be interpreted in two ways.

The first approach is to think of the agent as having to solve several POMDPs, with different reward functions and possibly different state spaces. For example, one problem to solve would be finding food, another would be making tools, and so on. Over its lifetime, the agent encounters a series of POMDPs, which are related to each other. Anticipation of goals now corresponds to predicting what future reward functions will look like, based on the reward functions seen to date. This involves some kind of meta-learning or meta-cognition. In the machine learning literature, this is most commonly referred to as transfer learning, and has been gaining increasing attention [38], [39], though the focus is on using experience from past problems to solve new problems better (more quickly, accurately, and robustly), rather than on predicting future goals and doing anticipatory learning.

The second approach is to think of the agent as having one *large* POMDP to deal with, where the goal is simply survival. The agent then has to decompose

---

[3] *Hominin* refers to all the species of humans that ever evolved. The term *hominid* includes chimps and gorillas.

this POMDP into smaller POMDPs by discovering sub-goals. This is also a very hard problem and has been getting a lot of attention in the machine learning community [40], [41], [42].

It is not clear to what extent animals other than humans are capable of such sophisticated cognitive processes. It may be the case that other animals, such as the great apes, are capable of this *kind* of cognition, but to a lesser *degree*. They are, for example, capable of decomposing a relatively simple problem into a hierarchy of subgoals, such as pushing over a chair to a spot underneath a banana hanging from the ceiling so that they can climb up and reach the banana. But they certainly don't have the capabilities of humans, who can plan their entire careers and lives.

Gärdenfors et al. argue that once anticipatory planning appeared, it led to a selective pressure for evolving a means for cooperation about future goals, and that this led to the emergence of symbols. This is a very interesting hypothesis, which can be examined closely through simulations of agents with varying levels of anticipatory planning capabilities, perhaps by using some of the techniques cited above.

4.1.2 Social Intelligence

The key question for the emergence of structured language is, what cognitive preadaptation could provide the computational machinery for generating and processing highly structured language?

Several people have talked about social intelligence or "Machiavellian" intelligence as being the key factor in the emergence of structured language [35], [43], [44], [45].

Cheney and Seyfarth [44] point out that nonhuman primates do not seem to have a theory of mind. Their vocalizations may be intended to modify audience behavior, but are not intended to modify audience beliefs. They seem to be incapable of distinguishing their own knowledge from that of another individual's. As an example, they present an analysis of baboon contact barks. Baboons generally move through wooded areas in a group. Individuals that are separated from one another produce loud barks. These barks however do not seem to be produced with the intent of informing others of their location, rather they seem to be emotional responses to the stress of being potentially lost. If they had an informative intent, we would expect individuals that are securely in the center of the group to respond to contact barks in order to inform the others of the location of the group. However, playback experiments have shown this not to be the case. They produce answering barks only if they themselves are separated from the group.

Despite this, nonhuman primates have a very structured understanding of social interactions, which they glean from both direct observation and through listening to vocalizations. Cheney and Seyfarth provide the following characterization of nonhuman primates' social knowledge.

- It is representational.

- It has discrete values.
- It is hierarchically structured.
- It is rule-governed and open-ended.
- It is propositional.
- It is independent of sensory modality.

They point out that these are very similar to the structural properties of human language, though they do not claim that all of the syntactic properties of human language are represented here. Recursion, for example, is not present in this list.

We hypothesize that, since nonhuman primates do not have a theory of mind, they are incapable of social anticipation. Instead, they are using state anticipation to keep track of social structure. This is computationally expensive, since it requires keeping track of each individuals responses to various kinds of cues, and therefore it sets an upper limit on the number of individuals and the number of different cues they can keep track of. When *homo ergaster* moved from the forest to the savannah, it would have faced a selection pressure for increasing group size, because in a more open landscape larger groups are more effective at repelling predators and protecting food. This probably led to adaptations such as a larger neo-cortex, and language, that can help maintain the cohesiveness of a larger group. This is also known as the social grooming theory of language [45]. Developing a theory of mind (that is, social anticipation) helps by allowing generalization, which reduces the computational burden of maintaining separate models of different individuals, and thus allows a larger group size. Further it requires the ability to do recursion, at least to a limited depth, because an individual must model another, who is in turn modeling him, and so on. These abilities, and the existing social knowledge system, probably got exapted (or recruited [46]) for the linguistic system.

4.1.3 The Mirror System Hypothesis

Arbib has suggested an alternate preadaptation for structured language: the mirror neuron system [47]. Mirror neurons were first discovered in the premotor area F5 of macaque monkeys [48]. Mirror neurons are observed to be active when the monkey executes a goal-directed arm movement, like picking up some food. They are also active when the monkey observes someone else (the experimenter, for example) perform the same movement. This observation has generated a lot of interest as to their functional role, and it has been suggested that they might form the precursor of a “mental simulation” system used to model internal states of conspecifics [49].

Interestingly, the homolog in the human brain of area F5 is Broca’s area, which is critically involved in language production and comprehension. These findings led to the mirror system hypothesis of Arbib and Rizzolatti:

"The *parity requirement* for language in humans - that what counts for the speaker must count approximately the same for the hearer - is met because Broca's area evolved atop the mirror system for grasping, with its capacity to generate and recognize a set of actions." [50]

Arbib has developed a fairly detailed account of the evolutionary stages, starting with motor control (grasping), and proceeding through the development of the mirror system, that might have led to the emergence of language. His hypothesized stages are listed in box 3. He suggests that proto-language consisted of two stages, proto-sign and proto-speech. Proto-sign emerged first, by exaptation from a system for imitation of movements. This means that language was initially gestural. The compositional nature of movements provided the right computational machinery for developing structured language.

Note that there is quite a gap between having a mirror system, and having a (compositional) language. A mirror system does not even provide the ability to do imitation learning. Gallese and Goldman point out that imitation behavior has never been observed with mirror neuron activity [49]. Arbib also points out that "further evolution of the brain was required for the mirror system for grasping to become an imitation system for grasping." He says, therefore, that stages S1 through S3 (see box 2) are pre-hominid. We believe that these stages do not require more than payoff and sensory anticipation.

Moving from simple imitation to complex imitation probably requires state anticipation, because it involves building a model of the behavior of conspecifics. It is not known, however, to what extent the mirror system is built (that is, learned) or inbuilt (that is, innate).

The crucial step for the development of language is the next one: from complex imitation to proto-sign. It is not clear how exactly this might have happened. Arbib points out that it must involve some neurological change. In fact he believes that proto-sign was preceded by pantomime, which is also qualitatively different from imitation. The key difference is intentionality. Imitation is performed for the purpose of reproducing a movement, whereas pantomime is performed with the intention of getting the other to think about what is being represented. Zlatev et al., similarly, posit "bodily mimesis" as the key transitionary phase from imitation to communication [51]. We believe that going from imitation to signing involves at least a rudimentary form of social anticipation, because it requires knowing that others have mental states that are distinct from one's own, and that they can be manipulated through one's actions. Going from mimesis to symbolicity might have resulted from a combination of goal-anticipation, as discussed in section 4.1.1, and social anticipation. In other words, it may be more a difference of degree than kind, as evidenced by the fact that bonobos and chimpanzees are capable of understanding simple symbolic language, though after much training.

Box 3: The Mirror System Hypothesis for Language Evolution: Arbib suggests the following stages in the evolution of the language capacity [47].

- S1: Grasping.
- S2: A mirror system for grasping shared with the common ancestor of human and monkey.
- S3: A simple imitation system for object-directed grasping through much-repeated exposure. This is shared with the common ancestor of human and chimpanzee.
- S4: A complex imitation system for grasping - the ability to recognize another's performance as a set of familiar actions and then repeat them, or to recognize that such a performance combines novel actions which can be approximated by variants of actions already in the repertoire.
- S5: *Protosign*, a manual-based communication system, breaking through the fixed repertoire of primate vocalizations to yield an open repertoire.
- S6: *Protospeech*, resulting from the ability of control mechanisms evolved for protosign coming to control the vocal apparatus with increasing flexibility.
- S7: *Language*, the change from action-object frames to verb-argument structures to syntax and semantics; the co-evolution of cognitive and linguistic capacity.

His criteria for language readiness are,

- LR1: *Complex imitation.*
- LR2: *Symbolization.*
- LR3: *Parity (mirror property).* What counts for the speaker must count for the listener.
- LR4: *Intended communication.*
- LR5: *From hierarchical structuring to temporal ordering.*
- LR6: *Beyond the here-and-now.*
- LR7: *Paedomorphy and sociality.* Paedomorphy is the prolonged period of infant dependency, which is especially pronounced in humans.

We look next at the attempts to explore the space of communication systems through computer simulation.

4.2 Artificial Language Evolution

4.2.1 Emergence of Signaling

Werner and Dyer did one of the earliest simulations of the emergence of language in a population of artificial organisms [52]. They simulated a population of "male" and "female" agents on a gridworld. Females stayed fixed in position (they were scattered over the grid), and males could move about. Further, females could "see" males (upto a certain distance) and produce a "sound", whereas males could "hear" but not "see" females. The controllers for the agents were small recurrent neural networks that were updated by means of a genetic algorithm. Whenever a male succeeded in finding a female, their genomes were combined using crossover and mutation to produce two new individuals which were placed in random locations on the grid. The parents were removed to conserve the size of the population.

They showed that this simple setup was sufficient for the emergence of a communication protocol, which led to an increase in successful mating over time. The exceedingly simple cognitive architecture of the agents (and the evolutionary procedure) does not allow explicit model building, since the recurrent neural networks did not have their weights updated other than by the genetic algorithm. So this is an example of implicit anticipation. It should also be noted that females would have to give directions more or less continuously to the males. They did not develop any sense of compositionality. A female could not say, for example, "go straight and then turn left".

Tuci et al. have done some recent experiments on the emergence of signaling in very simple robots [53]. They had two kinds of robots, with different sensors, in a C-shaped maze. One type of robot could see the location of a goal (a light), using an ambient light sensor, and the other type could sense walls using infrared sensors. Further, both kinds could produce and detect sounds. The goal was to navigate to the light without any collisions. Neither kind of robot could achieve this goal by itself, so they had to evolve a system of communication to help them to cooperate. The controllers for the robots were, again, small recurrent neural networks which were updated using a genetic algorithm, and weights were not updated other than through evolution. They showed that the robots were able to evolve a communication protocol for achieving the goal. The communication system was continuous (in time), because the robots were not capable of temporal abstraction and thus had to be constantly informing each other of their state. This, just like the Werner and Dyer simulation, is an example of implicit anticipation.

These two experiments show how easy it is to develop a signaling system, even in extremely simple cognitive agents, without any explicit anticipatory mechanism or any clear notion of symbols or language. In the space of communication systems, these systems would lie in the same location as bacteria. There have been several other simulations along these lines, some of which result in discrete communication systems, but none seem to go beyond cued representation. See [14] for a good overview of these as well as simulations of structured communication systems.

4.2.2 Emergence of Lexicons

Work on the emergence of shared symbol systems has focused more on how these systems come to be shared in a population (our third basic question), than the emergence of the capacity for symbolization.

The most famous of these experiments has been the series of Talking Heads experiments carried out by Steels et al. See [54] for a review of these and other experiments based on language games. A typical game consisted of a speaker and a hearer agent (robotic heads) that were presented with a collection of colored geometric shapes on a screen. This collection comprised the context for communication. The speaker agent would select one shape, and use its internal vocabulary to produce a symbol to communicate to the hearer which shape it had selected. A game was deemed a success if the hearer agent correctly picked out the shape based on the symbol it had heard. If it failed, it would be told what the correct shape was. There were two aspects to this game scenario. One was to show that a shared symbol system could emerge in the population of agents through these simple interactions. The other was to show that a shared conceptual space would emerge as well. The agents start without the ability to discriminate shapes, and learn to do it by building discrimination trees as needed. For example if the speaker picked the blue triangle, and the hearer had not yet learned to discriminate blue from other colors, it would extend its discrimination tree in order to do so.

It is difficult to put these experiments into our space of communication systems because they are not ecological. By this we mean that the agents are engaged in communication as their primary task. This makes it impossible to use these to conclude anything about the emergence of the communication *capacity*. However, for our purposes, they do make the important point that once the capacity for symbolization appears, the emergence of a shared symbolic system can happen through population dynamics alone, without the need for more complex cognitive mechanisms.

An example of an ecological simulation is the “mushroom-world” of Cangelosi and Parisi [55]. This world consists of two types of mushrooms: edible and poisonous. An agent, which is a feed-forward neural network, can learn to discriminate between these two types on the basis of the sensory impression they generate (differences in color, shape, size, etc.) Agents and mushrooms are distributed in a grid-world, and the agents move around and eat mushrooms. For each edible mushroom that they consume, they get 10 “energy points”, and for each poisonous mushroom they consume, they get -11 energy points. At the end of a fixed lifetime, the 20 most energetic agents would be selected to form the next generation by replication and mutation. The weights of the neural networks only changed by mutation. They showed that when the agents were allowed to label mushrooms for each other (that is, communicate), eventually a shared stable communication system, consisting of just two symbols corresponding to “edible” and “poisonous”, emerged. This is very similar to an alarm call system. Here again, once *capacity* to communicate was provided, a stable conventional symbol system eventually evolved.

4.2.3 Emergence of Structure

There have been very few ecological simulations of the emergence of structured language. Cangelosi extended the mushroom-world model to allow the production of two-word utterances [56]. In this version, there were six kinds of mushrooms, of which three were edible and three were poisonous. The edible ones needed to be "approached" in the right way in order to be eaten successfully. The right way to approach was determined by the color of the mushroom. The neural networks representing the agents had two clusters of linguistics units, which correspond to a two-word utterance. The agents were not forced to use two words, and in some cases, did not. However, in ten of eighteen simulations, the populations evolved compositional utterances, and in seven of those, the evolved language could be interpreted as having a verb-object structure. This is because one word in one cluster was consistently used to refer to the poisonous mushrooms ("avoid"), and another to the edible mushrooms ("approach"), and the words in the other cluster were used to distinguish types of edible mushrooms.

Interestingly, the simulation was carried out in two stages. In the first stage, the population was evolved to learn foraging, that is, how to distinguish the mushrooms and how to approach the edible ones. In the second stage, the population would consist of 20 parents from the previous generation along with 80 children from the new generation, and the children would learn from the parents by using the back-propagation algorithm. In other words, the children are forming a payoff and sensory anticipatory model. The communication was still cued, however. It is also not clear if compositional language emerged from anything other than the vagaries of the learning process.

Smith et al. have suggested that the emergence of grammar might, in fact, have more to do with the learning process than with ecological conditions [57]. Their Iterated Learning Model consisted of parents teaching language to children in a succession of generations. In such a situation, if the environment has structure, they showed that compositional language is more easily learnable, and therefore can be correctly passed through the *transmission bottleneck*. The transmission bottleneck refers to the fact that children have to learn the language from a finite sample, and therefore may not see all valid sentences. A rule-based language allows them to generalize correctly to unseen instances. Their model is quite abstract, however, and the agents do not have a cognitive architecture, so we can't put it into our communication space.

5 What are some of the sufficient conditions for the emergence of language?

Based on the above examination of theories and examples of language emergence, we can start to infer some of the minimal conditions for the emergence of various kinds of language.

- **Adaptive value**: This is the most basic condition for the emergence of communication. If communication doesn't have an adaptive value, it will not evolve.

- **Memory**: Memory is a very large and complex part of cognition. Humans, e.g., have working memory, long-term memory, propositional memory, episodic memory, muscle memory, etc. We don't mean that all of these are necessary for language. What we mean is that the cognitive system must not be purely Markovian. This can be achieved by using a hierarchical planning system like a semi-Markov decision process, for example. This is necessary because our language is discrete. If the agents have no memory, then they would have to communicate continuously in order to cooperate in the achievement of some goal. See the experiments of Tuci et al. [53] about the evolution of signaling. Further, if the agents don't have memory, communication is forced to be cued.

- **Symbol generation**: This refers to the ability to generate new symbols when required. This is also known as having an *open* symbol system. In the absence of this ability, we cannot have detached representations.

- **Planning in non-stationary environments**: If social intelligence is one of the keys to developing rich syntactic language, then the agents must be capable of planning while taking into account that other agents are also planning. This makes the learning environment non-stationary, as we have observed earlier. It may have led to the recursion that is observed in modern language, such as center-embedding of clauses, e.g. "The cat *the dog chased* ran up a tree.". In a recent article, Hauser et al. hypothesize that this may be the only innate aspect of grammar [9].

Anticipation provides a framework for analyzing the computational properties of cognitive systems. As we have seen, there has been very little simulation work examining how communication systems that have displaced or detached representations might have emerged. This means that most of the space of communication systems has not really been investigated through simulation. Theories of language emergence, however, deal mostly with this area of the space of communication systems. These theories are now acquiring sufficient detail that the door to simulation studies has been opened. What we need are focused simulations that ask precise questions, for example, can a population that is capable of complex imitation develop the ability to communicate through mimesis? What exactly is required to make this transition? How complex does the cognitive system have to be to make the transition from compositional to syntactic language? and so on.

We are heading into an exciting period in the study of language evolution as we are beginning to see the emergence of a detailed understanding of the minimal conditions required for the emergence of language. When we succeed in answering these questions, not only will we have solved a very difficult scientific problem [58], but the technological possibilities will be tremendous. We will be able to design truly autonomous *populations* of agents which will be able to collectively accomplish feats that are currently beyond human or machine capabilities.

Acknowledgements

We thank Martin Butz, Kiran Lakkaraju, participants in the UIUC language evolution seminar in the fall semester of 2006, and an anonymous reviewer, for their feedback, which helped to improve the paper.

References

[1] Smith, J.M., Szathmáry, E.: The Major Transitions in Evolution. Oxford University Press (1995)

[2] Kendon, A.: Some considerations for a theory of language origins. Man, New Series **26**(2) (1991) 199–221

[3] Aiello, L.C.: The foundations of human language. In Jablonski, N.G., Aiello, L.C., eds.: The Origin and Diversification of Language. Wattis Symposium Series in Anthropology, Memoirs of the California Academy of Sciences, Number 24. University of California Press (1998)

[4] Johansson, S.: Why Did Language Evolve? In: Origins of Language: Constraints on Hypotheses. John Benjamins Publishing Company, Amsterdam/Philadelphia (2005)

[5] Hurford, J.: The evolution of language and languages. In Dunbar, R., Knight, C., Power, C., eds.: The Evolution of Culture. Edinburgh University Press (1999) 173–193

[6] Comrie, B.: From potential to realisation: An episode in the origin of language. In Trabant, J., Ward, S., eds.: New Essays on the Origin of Language. Number 133 in Trends in Linguistics, Studies and Monographs. Mouton de Gruyter, Berlin - New York (2001)

[7] Chomsky, N.: Rules and Representations. Columbia University Press, New York (1980)

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

[9] Hauser, M.D., Chomsky, N., Fitch, W.T.: The faculty of language: What is it, who has it, and how did it evolve? Science **298** (2002) 1569–1579

[10] Pinker, S., Jackendoff, R.: The faculty of language: What's special about it? Cognition **95** (2005) 201–236

[11] Bickerton, D.: Language and Species. University of Chicago Press (1990)

[12] Johansson, S.: Protolanguage. In: Origins of Language: Constraints on Hypotheses. John Benjamins Publishing Company, Amsterdam/Philadelphia (2005)

[13] Jackendoff, R.: Possible stages in the evolution of the language capacity. Trends in Cognitive Science **3**(7) (1999) 272–279

[14] Wagner, K., Reggia, J., Uriagereka, J., Wilkinson, G.: Progress in the simulation of emergent communication and language. Adaptive Behavior **11**(1) (2003) 37–69

[15] Cucker, F., Smale, S., Zhou, D.X.: Modelling Language Evolution. In: Language Acquisition, Change and Emergence. (2003)

[16] Komarova, N.L.: Replicator-mutator equation, universality property and population dynamics of learning. Journal of Theoretical Biology **230** (2004) 227–239

[17] Steels, L., Kaplan, F.: Collective learning and semiotic dynamics. In Floreano, D., Nicoud, J.D., Mondada, F., eds.: Proceedings of the European Conference on Artificial Life (ECAL), Springer-Verlag (1999) 679–688

[18] Nowak, M.A., Plotkin, J.B., Jansen, V.A.A.: The evolution of syntactic commmunication. Nature **404** (2000) 495–498

[19] Roy, D.: Grounding words in perception and action: Computational insights. Trends in Cognitive Sciences **9**(8) (2005) 389–396

[20] Taddeo, M., Floridi, L.: Solving the symbol grounding problem: a critical review of fifteen years of research. Journal of Experimental and Theoretical Artificial Intelligence **17**(4) (2005) 419–445

[21] Franks, N.R., Pratt, S.C., Mallon, E.B., Britton, N.F., Sumpter, D.J.T.: Information flow, opinion polling and collective intelligence in house-hunting social insects. Phil. Trans. R. Soc. Lond. B **357** (2002) 1567–1584

[22] Bassler, B.L.: How bacteria talk to each other: Regulation of gene expression by quorum sensing. Current Opinion in Microbiology **2**(6) (1999) 582–587

[23] Reading, N.C., Sperandio, V.: Quorum sensing: The many languages of bacteria. FEMS Microbiology Letters **254**(1) (2006) 1–11

[24] Gärdenfors, P.: Cooperation and the evolution of symbolic communication. In Oller, K., Griebel, U., eds.: Evolution of Communicative Systems: A Comparative Approach. MIT Press, Cambridge, MA, USA (2004) 237–256

[25] Hockett, C.F.: The origin of speech. Scientific American **203**(3) (1960) 88–96

[26] Deacon, T.: The Symbolic Species: The Co-Evolution of Language and the Brain. W. W. Norton and Co. (1997)

[27] Seyfarth, R.M., Cheney, D.L.: Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. Science **210** (1980) 801–803

[28] Savage-Rumbaugh, S., Rumbaugh, D.M., McDonald, K.: Language learning in two species of apes. Neuroscience and Biobehavioral Reviews **9**(4) (1985) 653–665

[29] Rosen, R.: Anticipatory Systems. Pergamon Press (1985)

[30] Butz, M.V., Sigaud, O., Gérard, P.: Internal models and anticipations in adaptive learning systems. In Butz, M.V., Sigaud, O., Gérard, P., eds.: Anticipatory Behavior in Adaptive Learning Systems: Foundations, Theories, and Systems. Volume 2684 of LNCS. Springer-Verlag, Berlin (2003) 86–109

[31] Davidson, J.: Genetic exchange between bacteria in the environment. Plasmid **42**(2) (1999) 73–91

[32] Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA (1998)

[33] Hammer, M., Menzel, R.: Learning and memory in the honeybee. The Journal of Neuroscience **15**(3) (1995) 1617–1630

[34] Aiello, L.C., Dunbar, R.I.M.: Neocortex size, group size, and the evolution of language. Current Anthropology **34**(2) (1993) 184–193

[35] Worden, R.: The evolution of language from social intelligence. In Hurford, J.R., Studdert-Kennedy, M., Knight, C., eds.: Approaches to the Evolution of Language: Social and Cognitive Bases. Cambridge University Press, Cambridge (1998)

[36] Gärdenfors, P., Osvath, M.: The evolution of anticipatory cognition as a precursor to symbolic communication. In: Proceedings of the Morris Symposium on the Evolution of Language, Stony Brook, NY, USA (2005)

[37] Balkenius, C., Gärdenfors, P., Hall, L.: The origin of symbols in the brain. In: Proceedings of the Conference on the Evolution of Language, Paris (2000)

[38] Vilalta, R., Drissi, Y.: A perspective view and survey of meta-learning. Artificial Intelligence Review **18**(2) (2002) 77–95

[39] Swarup, S., Mahmud, M.M.H., Lakkaraju, K., Ray, S.R.: Cumulative learning: Towards designing cognitive architectures for artificial agents that have a lifetime. Technical Report UIUCDCS-R-2005-2514 (2005)

[40] McGovern, A., Barto, A.G.: Automatic discovery of subgoals in reinforcement learning using diverse density. Proceedings of the International Conference on Machine Learning (2001)

[41] Sutton, R., Precup, D., Singh, S.: Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. Artificial Intelligence **112**(1-2) (1999) 181–211

[42] Drummond, C.: Accelerating reinforcement learning by composing solutions of automatically identified subtasks. Journal of Artificial Intelligence **16** (2002) 59–104

[43] Byrne, R., Whiten, A., eds.: Machiavellian Intelligence. Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans. Clarendon Press, Oxford (1988)

[44] Cheney, D.L., Seyfarth, R.M.: Constraints and preadaptations in the earliest stages of language evolution. Linguistic Review **22** (2005) 135–159

[45] Dunbar, R.I.M.: Grooming, Gossip, and the Evolution of Language. Faber and Faber (1997)

[46] Steels, L.: The recruitment theory of language origins. In: Proceedings of the Alice V. and David H. Morris International Symposium on Language and Cognition: The Evolution of Language, SUNY Stony Brook (2005)

[47] Arbib, M.A.: From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. Behavioral and Brain Sciences **28** (2005) 105–167

[48] Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Lupino, G., Matelli, M.: Functional organization of inferior area 6 in the macaque monkey. ii. area f5 and the control of distal movements. Experimental Brain Research **71**(3) (1988) 491–507

[49] Gallese, V., Goldman, A.: Mirror neurons and the simulation theory of mind-reading. Trends in Cognitive Sciences **2**(12) (1998) 493–501

[50] Rizzolatti, G., Arbib, M.A.: Language within our grasp. Trends in Neurosciences **21**(5) (1998) 188–194

[51] Zlatev, J., Persson, T., Gärdenfors, P.: Bodily mimesis as "the missing link" in human cognitive evolution. Technical Report LUCS121, Lund University Cognitive Science, Lund, Sweden (2005)

[52] Werner, G.M., Dyer, M.G.: Evolution of communication in artificial organisms. In Langton, C.G., Taylor, C., Farmer, J.D., Rasmussen, S., eds.: Artificial Life II, Reading, Mass., Addison-Wesley (1991)

[53] Tuci, E., Ampatzis, C., Vicentini, F., Dorigo, M.: Operational aspects of the evolved signalling behavior in a group of cooperating and communicating robots. In: Proceedings of the 3rd International Symposium on the Emergence and Evolution of Linguistic Communication (EELCIII); at the Conference on the Simulation of Adaptive Behavior (SAB06). Volume 4211 of LNAI., Rome, Italy (2006) 113–127

[54] Vogt, P.: Language evolution and robotics: Issues in symbol grounding and language acquisition. In: Artificial Cognition Systems. Idea Group (2006)

[55] Cangelosi, A., Parisi, D.: The emergence of a "language" in an evolving population of neural networks. Connection Science **10** (1998) 83–97

[56] Cangelosi, A.: Modeling the evolution of communication: From stimulus associations to grounded symbolic associations. In D. Floreano, J.N., Mondada, F., eds.: Proceedings of the European Conference on Artificial Life, Berlin, Springer-Verlag (1999) 654–663

[57] Smith, K., Kirby, S., Brighton, H.: Iterated learning: A framework for the emergence of language. Artificial Life **9**(4) (2003) 371–386

[58] Christiansen, M.H., Kirby, S.: Language evolution: The hardest problem in science? In Christiansen, M., Kirby, S., eds.: Language Evolution: The States of the Art. Oxford University Press (2003)