

Modelling Language-Physiology Coevolution

Daniel Livingstone and Colin Fyfe
Applied Computational Intelligence Research Unit
University of Paisley
livi-ci0@paisley.ac.uk

Introduction

A feature of current computational models of language evolution is that the individuals in later populations are not structurally, 'physiologically', different from those in the first. Evolution may be working on the language itself, as learned by agents which do not evolve, or on an innate communication scheme. A number of models specifically demonstrate self-organisation of communication schemes and grammars in populations that are already capable of language.

Such models do not show communities evolving from those capable of some simple proto-language towards those capable of some fuller language. In contrast, in human evolution, vocalisations and speech provided a selective advantage that led to the exaptation and adaptation of aspects of human physiology to support improved language capacity (Deacon 1992; Lieberman 1992). This led to a process of language-physiology coevolution. From the coevolution of physiology and language, hominids developed differences from other primates, such as increased brain size and a supralaryngeal vocal tract.

The coevolution of speech and physiology in humans was also not without cost. The larger brain costs more energy to maintain and requires a longer infancy to allow brain growth to complete. The dropped epiglottis allows greater clarity and distinctiveness in speech, but increases risk of choking.

While some vocalisations are evolved responses - crying, laughter, etc. - speech is learned afresh with every individual. Learning allows quicker adaptation to changes in the environment and faster solutions to environmental problems. The Baldwin effect (Baldwin 1896) relates how learning can influence evolution: individuals most capable of successfully adapting to their environment will be more likely to contribute to future generations. Thus an increase in language capabilities can occur over a population due to the higher fitness of individuals with greatest language capabilities, where the use of language provides a selective advantage.

Continuous evolution from a species capable of only a basic proto-language to one capable of full language is then possible. Wray (this volume) describes the evolution of language from a simple proto-language. The start point is a learned holistic proto-language, with capabilities similar to the verbal communication of modern primates, with a small set of words to convey general meanings. The start point of language evolution in our model is the ability to learn and use very limited proto-language consisting of only a pair of signals.

We develop our model to show evolution in language agent physiology to support communication. In this model, the number of unique signals that an agent can produce defines its language ability. An agent's language ability may be greater or less than required for the agent to be able to communicate all information of interest. Where additional language capabilities may provide benefit to individuals, such evolution is dependent on the benefit being greater than the costs involved. We show that language negotiation can succeed in heterogeneous populations, and that it is possible to model the evolution of language ability. Finally we discuss the results with respect to the evolution of human language.

Computational Models of Language Learning Populations

A number of recent computational models demonstrate the evolution of innate communication schemes (for example, Oliphant 1996, Cangelosi & Parisi 1996, Di Paolo, 1997). Other models demonstrate the self-organisation of lexicons, grammars and sound systems in populations of language agents without evolution (see Kirby, this volume, de Boer, this volume, Steels, 1996, Batali 1997).

Batali (1994) combines evolution and learning in an ANN (artificial neural network) model in which recurrent neural networks attempt to learn context-free grammars in an investigation of innate language biases and critical periods. The language agents have a fixed structure, and a predetermined number of inputs and outputs. Evolution determines initial weight values for the networks, selecting appropriate values for the class of languages on which the population is trained. This model demonstrates how evolution can tune innate learning mechanisms towards certain grammars, once the mechanisms for language have developed.

In Fyfe and Livingstone (1997) a population was modelled in which the individuals learn to identify stochastic sources in an environment and then learn a common language to communicate about the sources present in any given environment. The agents were implemented as neural networks with three layers of neurons - an environment input layer, a hidden internal state layer and a language output layer. The language agents first learned to identify different sources in the environment before language negotiation began.

Experiments were performed with communities of agents with differing representational capabilities, i.e. different hidden layer sizes. Communicative success was seen to improve in populations with a common representational capability, and evolution towards homogenous representation capability was observed. A weakness in this model was that the production of language was compared, but the ability to interpret signals was not tested. Thus, by producing the same word for the same environment two agents are assumed to be successfully communicating. No pressure to produce different signals for different environments, or to be able to decode signals, existed.

A model using ANN agents learning to produce and interpret signals, with observational rather than reinforcement based learning, is presented in Oliphant (1997, VI.E). In Oliphant's model, ANN agents relate meanings to signals and vice-versa with winner-take-all on the produced vector, with meanings and signals represented by binary vectors with only one active value. Using a form of Hebbian learning (learning which increases the strength of a weight when the neurons which are connected by it both fire simultaneously), Oliphant shows successful negotiation to a common optimal language, with a different signal being used for each meaning.

Another model where a community of ANN negotiates a shared lexicon is presented in Hutchins and Hazelhurst (1995). The agents within this model are similar to the ones presented in this paper, although their agents are more complicated with an additional layer (used only in learning). The authors limit their investigation to the development of a shared lexicon.

Modelling Language-Physiology Coevolution

Our model comprises of a population of simple language agents capable of sending and receiving messages to share information. In our model, agents learn to map messages that are sent by other agents to internal states or 'meanings'. The agents use these messages to coordinate their internal states. By learning from each other a coordinated communication system is developed by a community of agents. An artificial neural network architecture is used for the production of messages from each arbitrary meaning, and for the reverse mapping.

We add to the model by making the expressive capability of the language dependent on hereditary genes. The genes determine the number of language nodes possessed by agents, which determines the range of signals that can be produced. This forms the basis for investigating the evolution of language ability within our model. A fitness cost according to the number of language nodes possessed by an agent is added to represent the expense of language adaptations.

Our population is composed of two layer fully connected ANN with N inputs and M outputs. The meaning is modelled as a bipolar (± 1) vector of length N, which is presented at the inputs of a language

agent. This can be fed forward through the NN to determine an agent's 'word' for that meaning, each output being thresholded to a bipolar value (± 1), as given below

$$y_j = \sum_{i=0}^{N-1} x_i w_{ij} \quad (1)$$

then $Y_j = 1$ if $y_j \geq 0$, $Y_j = -1$ if $y_j < 0$

where the vector \mathbf{y} , $y_1 \dots y_M$, is the word generated for meaning vector \mathbf{x} . In all cases described, a sparse 'grandmother' coding of the meaning is used with only one value in the vector of +1, all others being -1. To interpret a signal vector, the signal can be fed back to generate a meaning vector. Competition can then be applied to set one bit of the vector to +1, and the remaining bits to -1.

One significant difference in this model from that in Fyfe and Livingstone (1997) is that the successful interpretation of the environment is taken as given. By doing away with the interpretation of a complex environment, we can concentrate on the processes of language negotiation and evolution, adding complexity when basic phenomena are better understood.

The population training algorithm for language negotiation is described in Figure 1.

- | |
|--|
| <ol style="list-style-type: none"> 1. For t training rounds 2. pick random environment 3. for each agent (picked in random order) 4. pick another agent to be teacher 5. generate training signal from teacher 6. train pupil and teacher on signal |
|--|

Figure 1. Population training algorithm

A standard training algorithm can be used for individual agent training during population training. The training signal would be presented at the pupils outputs, the corresponding meaning at the inputs and the training algorithm used to update the weights. Language would be successfully negotiated, but with no pressure on an agent being able to interpret the correct meaning from a presented word the language could use a single word for more than one meaning. It would even be possible for a language to be negotiated which used only one word for all meanings. To be useful a communication scheme must carry information. A negotiated communication scheme should optimise the amount of information transmittable.

A learning algorithm which ensures the emergence of an optimal communication scheme - with a different signal shared by the whole population for each meaning state - is desired. The algorithm used in Oliphant (1997, VI.E) works in the case where individuals in a population are removed and replaced by new learners, but does not work in a generation based population training algorithm. The algorithm in Figure 1 is used as we wish to study evolution acting on language capability over generations of the population.

The results of Oliphant suggest that optimal learning is performed by an algorithm using the transmission behaviour of the population to train language reception and the reception behaviour to train language production. In this way an agent is most likely to succeed in communicating with other members of the population. This indicates inverse learning. An inverse, top-down, learning approach is presented in a number of generative models for ANN learning presented by Hinton and Ghahramani (1997). In these models there are feedback generative weights and feedforward recognition weights, and the problem of recognition is posed as which hidden units could be responsible for generating the input pattern (Figure 2).

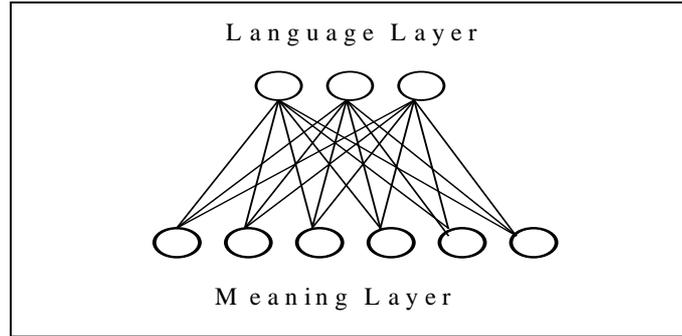


Figure 2. A language agent neural network. Recognition weights map language layer patterns to meanings, generative weights map meanings to patterns at the language layer

An additional concern is that we wish to model the evolution of language capability. Thus our population must be able to negotiate useful languages even where they are ‘physically’ incapable of learning a communication scheme capable of representing all environmental states. Similarly, the earliest proto-language users, physiologically incapable of learning a modern language, must have been able to learn useful, if restricted, languages.

In our model, language agents are represented as two layer ANN with a meaning layer and a language layer. Meanings are represented as sparse bipolar vectors, with only one bit set to +1. Signals are represented at the language layer as arbitrary bipolar vectors. Thus, for N language neurons, there are 2^N possible signals or words in the language, and for M meaning neurons there are M possible meanings. Competition exists between neurons in the meaning layer, such that any signal fed back from the language layer only has one corresponding meaning. When a signal is fed-back during interpretation, it is likely that several meaning neurons will fire at different strengths, the competition allowing a single meaning to be chosen unambiguously.

The output word from one NN is presented at the output layer of another agent ANN and fed back to produce a generated environment (2) - a ‘meaning’ vector - for the presented word. The error between the actual environment, x_i , and the generated environment, x'_i , is used for learning by the receiver agent, (3). The error is multiplied by a learning rate, η , to determine the correction to be applied to the weights, w , connecting the layers.

$$x_i = \sum_{j=0}^M y_j w_{ij} \quad (2)$$

$$\Delta w_{ij} = \eta(x_i - x'_i)y_j \quad (3)$$

We adapted the learning algorithm such that learning is only performed when a word is misclassified. When a word is correctly classified the receiving agent performs no learning. The learning algorithm is otherwise unchanged from (3).

More details of the model, including notes on implementation issues, can be found in Livingstone and Fyfe (1998).

Experiment 1: Language Negotiation in Populations of Homogenous Language Capability

For our first experiment we test whether populations of networks are able to negotiate language with the learning algorithm given above. Language agents have seven possible states, and agent populations are homogenous with from zero to seven language neurons.

We use spatially arranged populations, arranged in rings, and all communication occurs within a neighbourhood of the currently chosen agent. So, a teacher will be picked randomly according to the location of the current learner. The area of the neighbourhood is defined by a normal distribution centred on the learner. In all experiments described the standard deviation used is 0.6, placing a strong preference

on immediate and very close neighbours. A ring arrangement is not necessary, but means that all agents have the same number of neighbours with the same chances of selection.

With neighbourhood limited communication the results in Figure 3 were produced for 300 and 1000 training rounds for populations of 40 agents, averaged over 10 runs for each language capability. Fitness is measured by computing the number of signals out of 100 received which are correctly interpreted in a separate fitness phase after completing learning. The same neighbourhoods used in language negotiation are used for communication during fitness evaluation.

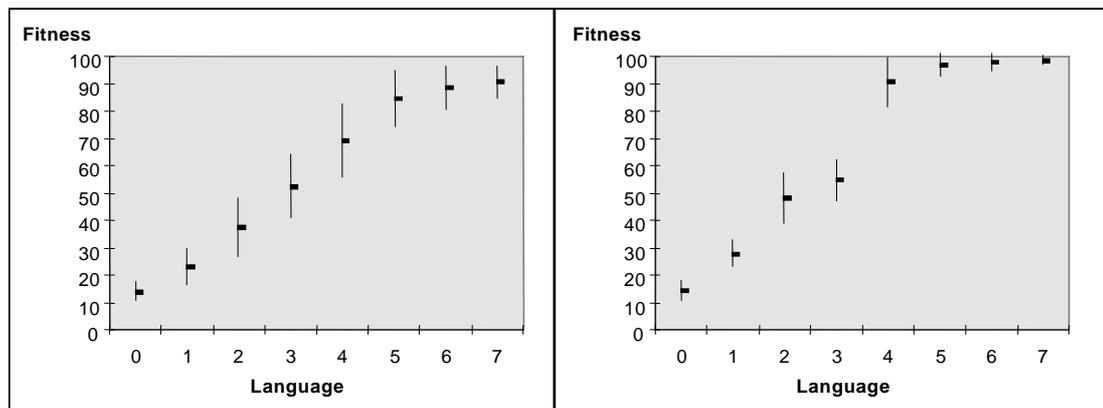


Figure 3. Average fitness, and standard deviation, with homogenous populations of 0-7 language neurons measured over ten experiment runs each case. With 300 (left) and 1000 (right) negotiation rounds.

With no language neurons, the performance of the agents is at chance level - on average agents guess the correct meaning one in seven times. As language capability increases, so does fitness showing that agents are successfully sharing information about their internal state. With more neurons, average success increases and with two or more language neurons increased training time increases fitness. If training time is extended sufficiently, populations with four or more neurons will consistently negotiate a language capable of allowing all information to be shared.

The results show that using our learning algorithm homogenous populations are able to negotiate useful language, even where the capability for an optimal language does not exist. As the language capability increases the success rate of communication increases. With four or more language units, the expressive power of the language exceeds the communication requirements of the environment. This has the potential for allowing multiple signals for the one meaning to be recognised correctly.

To test this, a homogenous language population with four language units is trained for 5,000 rounds. This allows a very high degree of coordination amongst the agents but, due to the local communication, does not negotiate a common language over the whole population. It is observed that large neighbourhoods negotiate a common signal for a given meaning, but distant agents may have significant differences in communication schemes used. At the boundaries between neighbourhoods, agents may exist which interpret signals from different schemes correctly. This is shown in Figure 4. The three agents included each attained a maximal fitness score, interpreting all signals correctly. A degree of “bilingualism” is possible in the agent communication schemes of due to the redundancy in their representation capacity.

agents in its neighbourhood with no benefit for itself. Another selection mechanism is required. (Oliphant 1994) shows that spatially organising a population can enable the evolution of altruistic behaviour through the ‘selfish gene’ effect.

The same neighbourhood functions used for communication between agents are also used when selecting parents for successive generations. After fitness has been evaluated, the new generation is formed. A parent is picked randomly according to relative fitnesses of all agents. A mate is picked for this parent according to the neighbourhood distribution. Two agents are then placed into the next generation, their language ability determined by applying crossover and mutation operators to the parent genes. A mutation rate of 0.005 per bit is used. Offspring are placed in the next generation in positions similar to those occupied by their parents in the previous generation.

This approximates kin selection, by placing close kin in similar neighbourhoods and enabling kin selection. Additionally, good signallers receive a more direct benefit. By increasing fitness of a neighbour, a good signaller increases the likelihood that it will be selected as the second partner for mating. The issue of spatial constraints on the evolution of communication is also discussed in Di Paolo (1998), based on a model in which agents can increase chance of reward from social partner selection by occupying space in a dense cluster of agents.

A number of runs were performed, all starting with homogenous populations with only one language production neuron. Two typical results can be seen in Figure 6. In these experiments there is a relatively rapid increase in the language capability of the population. In all runs, the language capability evolves to an extent that the language is able to express all meanings. The populations do not evolve homogeneously.

As well as the spatial selection providing a secondary benefit to a good signaller, the kinship rules make survival harder for defectors – agents which provide poor signals. Individual defectors can enjoy a free ride from good signallers but groups of defectors form, providing each other with poor signals limiting their survival and reproduction.

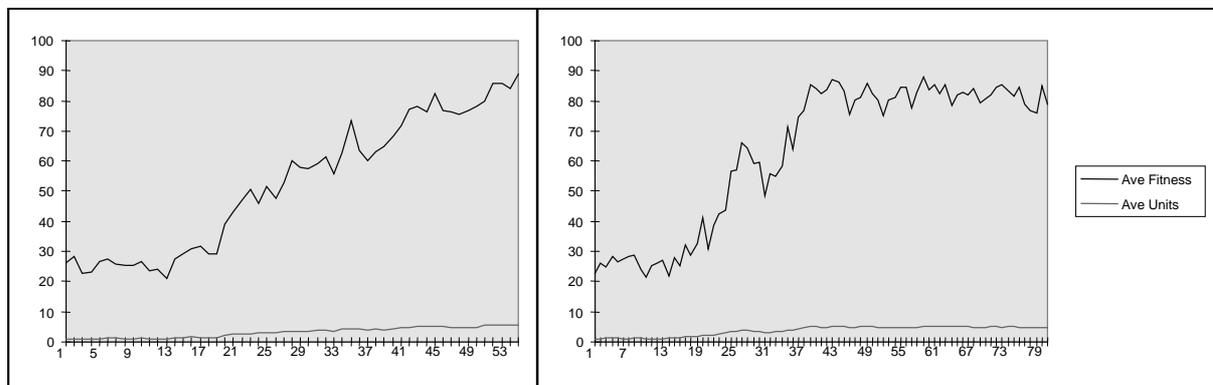


Figure 6. Average Fitness & Language Neurons against Generation. Two runs showing coevolution of language with language production capability. 300 training rounds per generation.

Experiment 3: Costly Language-Physiology Coevolution

With our model we aim to show how language may be adopted by a population despite a cost of possessing such an ability. For this purpose a fitness penalty is added to the model. By applying a penalty according to the number of language units, language capable agents are penalised and less capable agents get a free ride. For the results that follow, each agent is penalised by an arbitrary four fitness points per active language node they possess. In tests, other penalties have been used without qualitatively altering results.

With spatial selection, as previously described, the population still evolves to a more language capable one. Results of two experiments are shown in Figure 7.

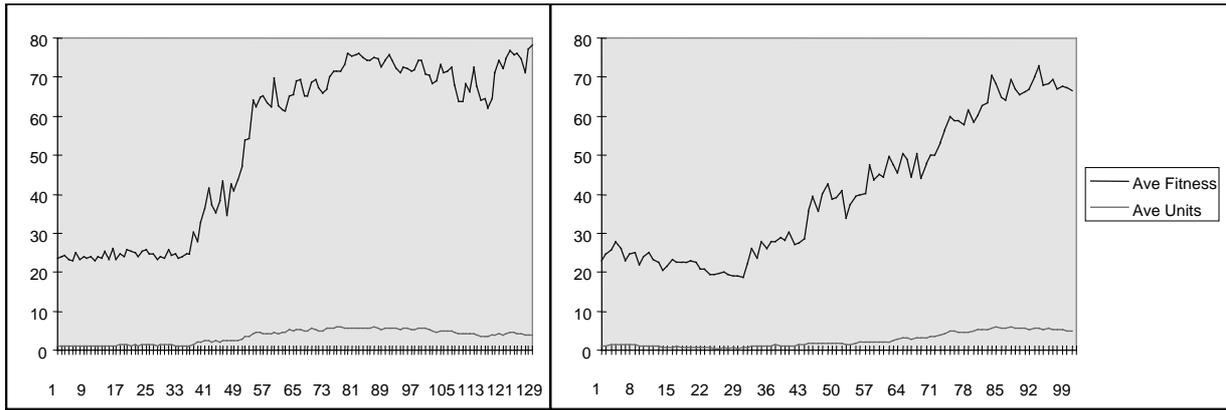


Figure 7. Average Fitness & Language Neurons against Generation. Two runs showing costly coevolution of language with language production capability. 300 language training rounds per generation.

In both experiments, a number of generations pass before increased language ability evolves. The language ability may then evolve slowly, in gradual increments, or rapidly, in a smaller number of larger increments.

As seen in the results of Experiment 1 (Figure 3), at least four language neurons are required to achieve maximum fitness. This would incur a 16 point fitness penalty in these experiments, resulting in a maximum possible fitness of 84 points. In both runs, near maximum fitness is achieved.

Experiment 4: Negotiation and Evolution without Spatial Organisation

We have stated the importance of spatial organisation on the previous sets of results, and in this experiment demonstrate this importance by reducing the strength of spatial selection and observing the results. We find our results in accord with Oliphant (1994).

The radius of neighbourhoods can be changed, larger neighbourhoods weakening the effects of spatial organisation, and the effects on language negotiation and evolution observed. Larger neighbourhoods weaken the effect of kin selection as agents will usually not be communicating with their ‘kin’ and reduces the likelihood that a good signaller agent will be selected as the partner of another agent to which good signals have been provided..

This is less serious where there is no cost of communicating - a very wide neighbourhood is reached before evolution of language fails. Where there is a cost even small increases in neighbourhood size prevents the evolution of language. This is demonstrated in the Figure 8. These graphs show that some form of spatial organisation is necessary for the evolution of an altruistic language ability.

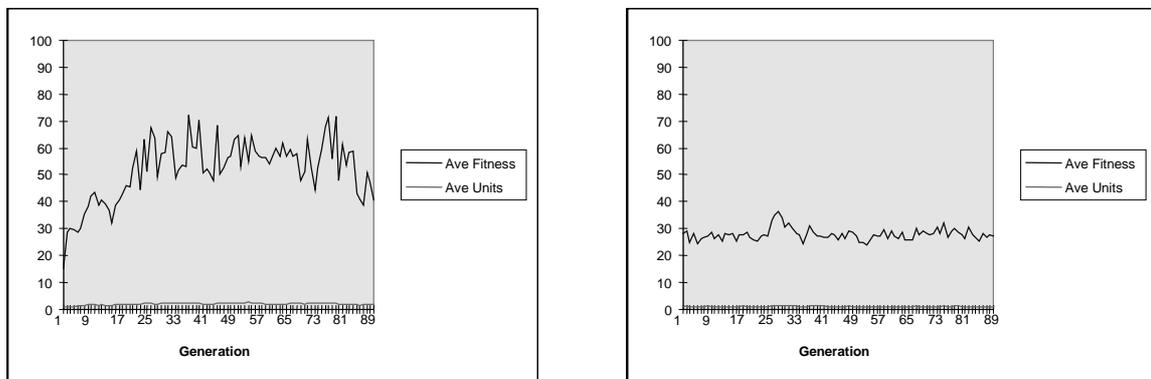


Figure 8. Evolution of language with reduced pressure from kin selection. Left, standard deviation 12, non-costly language. Right, standard deviation 3, costly language.

Spatial organisation also has an effect on language negotiation in this model. With small neighbourhoods, language forms in tight clusters. Close neighbours share negotiated languages, while two agents selected from distant points may be unable to communicate with each other. With larger neighbourhoods, the clusters are larger and ultimately global co-ordination of language can occur.

Discussion

The experiments above show evolution in the language ability of populations of simple language agents. The relevance of this to the evolution of language in hominids is now considered.

Our artificial language lacks almost all of the properties of a real language - in particular there is no *grammar*, no word morphology or sentence syntax. All communication consists of single words for single meanings. At the same time our artificial language also differs from human involuntary non-verbal vocalisations, such as laughter, crying or screaming, in that the communication scheme is learned rather than inherited.

Our model can be considered an abstract representation of the evolution of language - language provides an unspecified selective benefit to groups of kin who are able to share information. A number of different ideas of the 'original' selective benefits of language exist, it being widely agreed, however, that language does indeed have a selective benefit. However, language is clearly used to serve a great many functions (Dunbar 1997). Embodied experiments on the evolution of language focus on very specific uses of language, and some work has been carried out on evaluating the adaptive value of specific features or uses of language (for example, Cangelosi & Harnard 1998). 'Disembodied' experiments, such as presented in this paper, simply accept that language provides a benefit and directly reward agents for successful communication.

Although not grounded in an environment, the language ability of the agents in our model can be considered as their ability to learn useful languages - with inevitable cost and potential benefit. Our model clearly shows that despite physiological costs, such abilities can evolve. This process takes a number of generations before succeeding in a population. Individuals of different abilities frequently coexist within populations. To allow language agents of differing abilities to communicate we have given all agents a full capability to receive and learn signals, even those above their own production ability. This detail is also important for the evolution of language. For the evolution of language to occur over a period of time, it is necessary that language is learnable and useful at every stage of phylogenetic evolution.

Burling (this volume) argues that the ability to comprehend language has to be greater than that at producing language for the evolution of language to occur, that it is the act of interpreting an action that turns it into a signal. Lyn and Savage-Rumbaugh (1998) provide evidence on the language learning abilities of pygmy chimpanzees, which by far exceed their ability to produce language. The ability to learn to interpret signals similarly exceeds the ability to produce signals in our model. Within our model the ability to comprehend language is fixed throughout although it is clear that the ability to understand, as well as to produce, language has evolved.

To conclude, some models have highlighted selection mechanisms required for evolution of altruistic language users or the self-organisation of lexicons and grammars. Further models supporting evolutionary theories of language are required that demonstrate evolving populations in which agents are able to negotiate and learn useful languages at all stages of their evolution.

This has been achieved, with a high level of abstraction, in the model presented here. We have shown communities of varying abilities negotiating and using language successfully. Language ability has been seen to evolve gradually in communities of agents, and agents of differing abilities have been able to coexist within those communities with kin selection favouring individuals more language capable. Prospects for further work include the evolution of agents capable of rudimentary grammars from grammar incapable agents and information theoretic studies of evolving communication schemes.

Acknowledgements

Thanks are due to Mike Oliphant for his helpful comments on an earlier draft of this chapter.

References

- Baldwin, J. M. (1896). A New Factor in Evolution. *The American Naturalist* (30): 441-451.
- Batali, J. (1994). Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In Brooks, R. & Maes, P. (eds.), *Proceedings of the Fourth artificial life workshop*, Cambridge, MA, MIT Press.
- Batali, J. (1998). Computational simulations of the emergence of grammar. In Hurford, J., Studdert-Kennedy, M. & Knight, C. (eds.) *Approaches to the Evolution of Language*. Cambridge University Press.
- Cangelosi, A. & Parisi, D. (1996). The emergence of a "language" in an evolving population of neural networks. 18th Conference of the Cognitive Science Society, San Diego.
- Cangelosi, A. & Harnard, S. (1998). Adaptive advantages of "Hearsay" over direct sensorimotor experience. Paper read at the Second International Conference on the Evolution of Language, University of East London, April 6th-9th.
- Deacon, T. W. (1992). Brain-Language Coevolution. In Hawkins, J. A. & Gell-Mann, M. (eds.), *The Evolution of Human Languages*. Addison-Wesley: 273-303.
- Di Paolo, E. A. (1997). An Investigation into the Evolution of Communication. *Adaptive Behavior* 6(2): 285-324.
- (1998). Spatio-temporal and structural constraints in the evolution of communication. Paper read at the Second International Conference on the Evolution of Language, University of East London, April 6th-9th.
- Dunbar, R. (1997). *Grooming, gossip and the evolution of language*, Faber & Faber.
- Fyfe, C. & Livingstone, D. (1997). Developing a Community Language. Poster presented at the Fourth European Conference on Artificial Life, ECAL 97, Brighton, July 28th-31st.
- Hinton, G. E. & Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations. *Philosophical transactions of the Royal Society B*.
- Hutchins, E. & Hazelhurst, B. (1995). How to invent a lexicon: the development of shared symbols in interaction. In Gilbert, N. & Conte, R. (eds.) *Artificial Societies: The computer simulation of social life*. UCL Press.
- Lieberman, P. (1992). On The Evolution of Human Language. In Hawkins, J. A. & Gell-Mann M. (eds.) *The Evolution of Human Languages*, Addison-Wesley: 21-47.
- Livingstone, D. and Fyfe, C. (1998) A Computational Model of Language-Physiology Coevolution. *Computing and Information Systems, CIS Departmental Journal* 5(2), University of Paisley
- Lyn, H. & Savage-Runbaugh, E. S. (1998). Observational word learning in bonobos. Paper read at the Second International Conference on the Evolution of Language, University of East London, April 6th-9th.
- Oliphant, M. (1994). Evolving cooperation in the non-iterated prisoner's dilemma: The importance of spatial organization. In Brooks, R. & Maes, P. (eds.), *Proceedings of the Fourth artificial life workshop*, Cambridge M.A., MIT Press.
- (1996). The Dilemma of Saussurean Communication. *BioSystems* 37(1-2): 31-38.
- (1997). Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language, PhD thesis, University of California, San Diego.
- Pinker, S. (1994). *The Language Instinct*, Penguin.

Steels, L. (1996). Emergent adaptive lexicons. In Maes, P., Mataric, M., Meyer, J.-A., Pollack, J. & Wilson, S. W. (eds.) *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour*, MIT Press, Cambridge, MA.