IJCAI 2005

Tutorial

Modelling language origins and evolution

31 July 2005, Edinburgh, Scotland

http://www.ling.ed.ac.uk/~paulv/tutorial.html

Paul Vogt[1,2]   Bart de Boer[3]   Tony Belpaeme[4]

[1] University of Edinburgh
School of Philosophy, Psychology and Language Sciences
Language Evolution and Computation Research Unit
40 George Square, Edinburgh EH8 9LL, United Kingdom

[2] Tilburg University
Computational Linguistics
Induction of Linguistic Knowledge
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

[3] University of Groningen
Kunstmatige Intelligentie
Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

[4] University of Plymouth
School of Computing, Communications and Electronics
Adaptive Behaviour and Cognition Research Group
A318 Portland Square, Plymouth PL4 8AA, United Kingdom

paulv@ling.ed.ac.uk (Paul Vogt)
b.de.boer@ai.rug.nl (Bart de Boer)
tony.belpaeme@plymouth.ac.uk (Tony Belpaeme)

Contents

1 Introduction

In recent years computer scientists have been called upon by linguists, psychologists and cognitive scientists to help them find answers to some of the oldest and most profound questions in science: what are the origins of language and how does language evolve?

Humans are the only species having the capacity of language. True, other animals communicate: birds twitter, dolphins chatter and monkeys call, but no other animal has a communication system with the same complexity and versatility as human language. We are the first species to have evolved such a magnificent communication system, and it has affected our evolution and existence ever since the earliest forms of language were spoken by our ancestors. However, as we are the first to have language we are left with a number of questions about the origins and the evolution of language. Language does leave fossils[1], which makes it hard if not impossible to reconstruct how language evolved.

This tutorial is about how techniques from artificial intelligence can serve to build computer models with which we can shed light on the mystery of language origins and evolution.

Modelling language evolution can be organised in three areas:

Origins of language. This area investigates how linguistic capabilities could evolve. These may include drives to communicate, the evolution of the physical apparatus underlying communication, the evolution of cognitive capacities that allow language production, interpretation and acquisition.

Emergence of language. In this area, the underlying mechanisms for language production, interpretation and acquisition are assumed. From these mechanisms, the quest is to find mechanisms that allow specific aspects of language to emerge. Examples are the emergence of sound systems, meaning, vocabularies and grammar.

Evolution of language. Here the aim is to investigate how language evolves over time across multiple generations, given an emergence mechanism. Studies investigate, for instance, a transition from holistic communication systems to compositional communication systems, or other aspects of language change and diversity.

[1]This is not exactly true. At least two instances of "fossilised language" exist. One in the form of imprints our ancestor's brain and nervous system made in the skull, studying these dents in the skull reveals whether our ancestor's brains were adapted for language (e.g. Lieberman, 1998). The other fossil evidence comes in the form of fossilised linguistic expressions, which are observed when a new language is created out of two or more existing ones (for example when two cultures speaking different languages are forced to communicate; this process is called creolisation). Since the development of writing we have a record of language change, but the capacity of language evolved long before writing was invented.

An overview on language evolution and grounding can be found in (Vogt, in press). An overview on evolution of grammars in (Briscoe, 2002). Cangelosi and Parisi (2001) edited one of the first collections of papers on modelling the origins and evolution of language, and is after four years still one the best entry points into the field. (Kirby, 2002) contains an overview of the field and includes some future challenges.

2 Research topics

The computational modelling research can be split up in a number of research topics: the evolution of speech, the evolution of communication, grounding, the evolution of lexicons, grammaticalisation and language change and diversity. Each of these topics will be handled in detail in the next sections, together with a practical example.

2.1 Evolution of speech

2.1.1 A doo-doo-doo, a da-da-da

Evolution of speech is an area of language evolution with promising possibilities for computational modeling. Speech is a physical signal, and therefore its properties and its constraints are better known than for other aspects of language. Much is known about how speech is produced and what the possible sounds are that the human vocal tract can produce. Much is also known about the way speech sounds are perceived and (to a smaller extent) how they are processed by the brain. The universal tendencies of the repertoires of speech sounds - the number and kinds of sounds that are used, the shape and number of syllables, the way in which speech sounds change over time - are also relatively well-known. There are also possible equivalents to speech sounds in our evolutionary closest relatives, the great apes. Finally, some information about speech can be deduced from fossils. All these properties facilitate the construction of realistic computational models of evolution of speech, as well as to facilitate the comparison of data from computational models to that of real language.

Midsagittal section of the human head showing the brain and the vocal tract.

In order to use linguistic data and in order to compare computational data, some basic knowledge of phonetics, phonology and phonotactics is required. *Phonetics* is the study of the physical speech signal and looks at the production and perception of speech sounds. This field is perhaps most relevant for computational modelers. A good starting point is (Laver, 1994) while (Ladefoged and Maddieson, 1996) gives a good overview of the variety of speech sounds that is used in human language. *Phonology* is the study of the role of speech sounds in a language. It therefore tends to be more colored by underlying grammatical theories. Any textbook can provide a good introduction, but one should be careful to keep in mind from what theoretical perspective the textbook is written. *Phonotactics* is the study of how speech sounds are combined into syllables and words. For some reason this is a less-studied area of the study of speech. Phonotactics is also often heavily influenced by underlying theories of syllable structure. A good, and relatively theory-free introduction of the different possible syllables in human language can be found in (Vennemann, 1988). Another source of linguistic information is the study of how speech can change over time. Most of traditional historical linguistics has focused on change of speech sounds. Any introductory text on historical linguistics will provide the necessary theory as well as interesting examples, but (Crowley, 1994) is an example of a book that is perhaps not so much focused on Indo-European languages as most historical linguistics textbooks are. Equipped with this basic knowledge, a computer modeler can find interesting research topics, and communicate the results to an audience of linguists.

There are many perspectives from which the evolution of speech can be modeled. Most early work has been done on the explanation of phonological universals: why certain speech sounds occur more often than others do, and why certain combinations occur more frequently than others do. Liljencrants and Lindblom (1972) have modeled the universals of vowel systems, while Lindblom *et al.* (1984) looked at consonant-vowel syllables. This work was based on optimization, but later models have used agent-based models, either based on functional pressure (Berrah and Laboissière, 1999; de Boer, 2000) or on properties of the underlying neural systems (Oudeyer, 2005). There has also been work on syllable structure, using agent-based models (Oudeyer, 2005) and genetic algorithms (Redford *et al.*, 1998). Finally, there has been work on emergence of linguistic diversity (Livingstone and Fyfe, 1999a) and on emergence of tone systems (Ke *et al.*, 2003), both using agent-based models. More mathematical studies of properties of sound systems have also been made, e.g. (Wang and Minett, 2005).

2.1.2 An example: *emergence of vowel systems*

The example is a model that investigates how the universal tendencies of vowel systems can emerge in a population of agents. Liljencrants and Lindblom (1972) have shown that optimization results in realistic vowel systems, but it remains unclear how the optimization takes place. The hypothesis that was investigated with the example model was that optimization is the result of self-organization in a population communicating under constraints of perception and production.

The population of artificial agents can produce, perceive and remember speech sounds in a human-like way. Each agent is equipped with an articulatory synthesizer, a model of human perception for calculating the distances between different signals, and an associative memory for storing vowel prototypes. Also, each agent can interact with other agents (following a fixed pattern) by imitating them. These interactions are called imitation games, and are in the same line as "language games" (see Steels 2003). The agents can update their vowel repertoires depending on the outcome of the interactions in such a way that the expectation of future imitation success is maximized. The agents' goal in life is to imitate the other agents as well as possible with a repertoire of vowels that is as large as possible. However in doing this, the agents only use local information, and do not carry out any explicit optimization.

The repertoire of speech signals that humans can produce is limited by the physics and the physiology of the vocal tract. The repertoire of sounds that they can distinguish is limited by the properties of the auditory apparatus. In this example the focus is on vowels, and therefore only the articulatory and perceptual properties of vowels are integrated. The acoustic signal of a vowel can be described relatively straightforwardly by the resonances of the vocal tract. These resonances cause prominent peaks in the frequency spectrum of a vowel signal. The frequencies at which these peaks occur are called the formant frequencies. For different vowels, these peaks occur in different places. Vowels can therefore be uniquely characterized by their formant frequencies. In practice, only the three or four resonances at the lowest frequencies are relevant.

In the simulations discussed here, no acoustic signals are generated, although it would be straightforward to add the extra complexity. Vowels are represented by their first four formant frequencies. If an agent wants to generate a given vowel, it synthesizes these formants using an articulatory synthesizer. This synthesizer takes as inputs the three major vowel parameters: tongue position, tongue height and lip rounding (Ladefoged and Maddieson, 1996, Ch. 9). The outputs of the synthesizer are the first four formant frequencies of the corresponding vowel. The inputs are modeled as real values in the range 0 to 1. For tongue position, 0 means most to the front and 1 means most to the back. For tongue height, 0 means lowest and 1 means highest. For lip rounding, 0 means least rounded and 1 means most rounded. Thus (0, 0, 0) corresponds to the vowel [a] and to the formant frequencies (708, 1517, 2427, 3678) Hertz. The parameters (1, 1, 1) correspond to the vowel [u] and to the formant frequencies (276, 740, 2177, 3506) Hertz. This synthesizer is able to generate all possible basic vowels.

Humans also perceive vowels based on their formant frequencies. These can be used to calculate a perceptual distance between vowels.

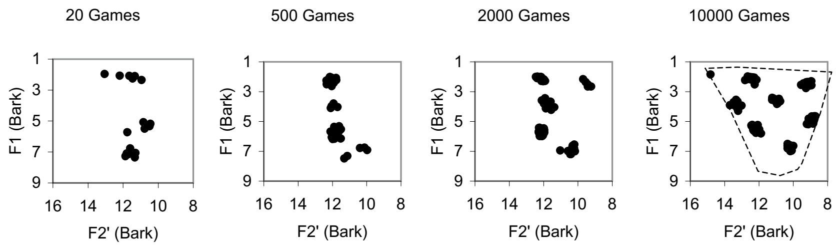

Figure 1: Emergence of a vowel system in a population of twenty agents.

Unfortunately, the distance calculation is not as simple as calculating a Euclidean distance between two formant vectors. As the bandwidth of the sound receptors at higher frequencies is greater than those at lower frequencies, blurring of spectral detail takes place at higher frequencies. This means that the formant peaks at higher frequencies can generally be compressed into one broader peak. The center frequency of this peak is called the effective second formant. The first peak in the spectrum is usually perceived as remaining in the same place, and is still referred to as the first formant.

While there are several ways to calculate the effective second formant, the one adopted in the research described here has been developed by Schwartz *et al.* (1997a). It is a non-linear weighted average of the 2nd, 3rd and 4th formant frequencies. In order to calculate distances between vowel signals, the first and effective second formant are expressed in the Bark frequency scale. The Bark scale is a perceptually inspired frequency scale that can be considered to be logarithmic for the frequencies that are relevant to formants. Equal frequency differences in Bark are perceived as equal intervals, in contrast to the way frequencies in Hertz are perceived. Weighted Euclidean distances in the space of the first and second effective formants is then used to determine which vowel in an agent's repertoire is recognized. This is done by calculating the distance between a perceived signal and all the vowels in an agent's repertoire. The vowel that is closest to the perceived signal is considered to be the one that the agent heard. The agents all start out with an empty vowel repertoire. By playing imitation games with each other, the agents have to develop a vowel system that is as large as possible, that allows for successful communication and that should be realistic if self-organization and embodiment are really factors in explaining the structure of human vowel systems.

The emergence of a vowel system in a population of twenty agents under 10% acoustic noise is shown in figure 1. In each of the frames of the figure, the acoustic aspects of the prototypes of the agents' vowels in the population are plotted in acoustic space. In this particular acoustic space (based on the first and effective second formant frequencies of a vowel signal), equal distances between points correspond to equal perceptual distances. Each vowel of each agent in the population is represented by a dot. Note that due to articulatory constraints, only a roughly triangular area of the acoustic space is available to the agents. This is indicated in the fourth frame of figure 1.

From the figure it is clear that after the first 20 games the agents still only have very few vowels. The vowels that exist are more or less randomly dispersed through the acoustic space, although some of them already show a tendency to cluster. This is caused by the fact that all agents start out with an empty vowel repertoire. In order to get the imitation games started, random vowels are inserted. However, the imitating agents in the games try to make imitations that are as close as possible and add these to their vowel repertoires. This accounts for the clustering. After some 500 imitation games, shown in the second frame, the clustering has become more pronounced. The most important process at this moment is the compacting of the clusters due to the fact that the agents move their vowel prototypes closer to the signals they perceive. However, there is still sufficient room in the auditory space for extra vowels, so the random addition of new vowels also plays a role. After 2000 games, the available vowel space becomes filled more evenly with vowels and the shape of the vowel system becomes more realistic. After 10 000 imitation games, the available acoustic space has become more or less filled up with vowels and the vowel system has become realistically symmetric and dispersed. After this has happened, the vowel system remains stable. However, it is not static. The vowel prototypes of agents (and therefore the clusters) tend to move, and it is even possible that they merge or that new clusters are formed (if they do not interfere with other clusters).

The emerging systems are realistic. Most of them conform to the universals that (Crothers, 1978) found for human vowel systems. When the percentages with which the different emerged systems occur are compared to the percentages with which human vowel systems occur, a good match is found as well. Schwartz *et al.* (1997b) have measured the occurrence of different vowel system types in the different languages in the UCLA Phonological Segment Inventory Database (UPSID, a database based on speech sound data of 451 languages, Maddieson 1984; Maddieson and Precoda 1990). They find 60 vowel systems with 6 vowels. Although their classification is not exactly the same as the classification shown in figure 4, there is good agreement. Of the systems they found in UPSID, 43% is of type A, 20% is of type B, 5% is of type C, 7% is of type D and 20% is of type E. No systems of type F were found, and two of the systems from UPSID (3%) cannot easily be fitted into the classification used here, but are probably of type A. Equally good agreement was found for systems of 4 and 5 vowels and reasonable agreement was found for systems of 7 and 8 vowels (de Boer, 1999a, Ch. 6). It seems that the simulation is capable of not only predicting the most frequently found vowel systems in human language (as was already possible with systems that optimize acoustic distinctiveness), but also of predicting the less frequently occurring vowel systems and, approximately, their relative abundance.

2.2 Evolution of communication

2.2.1 How and why did communication evolve?

In order for language to be of any use, there needs to be communication. Although some scholars have argued that language evolved for other purposes than communication, such as structuring our own thoughts or —as Dunbar (1996) claims— for defining social relations to others, a majority of researchers agree that language evolved for the purposes of communicating with others. But how and why did communication evolve?

Bees communicate the distance and heading of food sources through performing a complex dance.

Biologists define communication as behaviour of one animal which has an immediate or future effect on another animal, with both animals not necessarily being of the same species[2]. Examples abound: bees dance for each other to communicate the distance and heading of a food source, poisonous frogs display bright colours to warn and scare predators, and human males drive expensive cars to woo fertile females. What is important is that communication, to biologists at least, does not need a conscious effort. Even more, communication is often an "always on" behaviour: the male peacock cannot choose to stop displaying its bright colours (although it can stress it radiant plumage by opening its tail) or the wasp cannot choose to stop displaying its yellow and black stripes. Also note that communication is no necessarily beneficiary for both parties: when an angler fish lures prey using its bioluminescent protrusion there sure is communication, but without much benefit for the prey. Some scholars insist on a tighter definition of communication involving two-way interaction. To them the wasp's stripes or the angler fish's lure are just examples of display. True communication should, according to this view, involve an exchange of information between two parties. An excellent and comprehensive introduction to the evolution of communication is (Hauser, 1996).

[2]This definition is in some way too lax: if a rabbit rustling in the bushes thereby alerts a predator, this will have an effect on the rabbit, but we would be hard pressed to call this communication.

Many kinds of animal communication exist, some serving to impress opponents (agonistic communication), some to appease opponents (affiliative communication). Many kinds of communication are related to reproduction. Bird song typically serves on the one hand to mark territory and on the other hand to attract females. Courtship display relies on communication, involving not only visual, but also auditory and olfactory signals. The collection and distribution of food in many species involves a large amount of communication. Nestlings, for example, make a call specifically related to food begging when being fed by their parents. Probably the most well-known and well-studied example of food-related communication is the dance of honeybees (von Frisch, 1969). Communication also serves to warn conspecifics to possible danger. One of the classic examples being the vervet monkey alarm calls. It was long time believed that the vervet monkey's scream when it saw a snake was a purely emotional response, until it was observed that all monkeys in the troop stood on their hind legs and attentively watched the ground after hearing the scream. Furthermore, the vervet monkeys not only had a call to signal that a snake had been seen, it had two other distinct calls: one to signal the presence of a leopard and one to signal the sighting of an eagle. Each call elicited an appropriate response: after the leopard call the troop made it for the trees, and the eagle call resulted in the troop seeking shelter in dense bushes or near the trunk of the tree. The study of Seyfarth *et al.* (1980) made clear that the vervet alarm calls contained semantic reference: the snake call evidently referred to a snake, and was not, as previously believed, an emotional call after which one monkey stood on its hind legs and all the other monkeys in the troop imitated its behaviour. Seyfarth *et al.* observed that playing a recording of the alarm calls resulted in the same behaviour of the troop, showing that the vervets were not just imitating others, but that the *calls were associated with meaning*. Vervet communication, in other words, carries *information*.

Animal communication all seems to gravitate towards sexual reproduction and it is therefore not surprising that many computational models take an evolutionary perspective on animal communication. Computational models have been useful for studying three important issues of animal communication.

- What selective pressures can lead to the evolution of a communication system?
- How can a non-communicating species gradually evolve towards a communicating species, i.e. what route does evolution take?
- How can existing behaviour and embodiment be recruited by evolution for communication (this process is called *exaptation*)?

The first issue is probably the most hairy. It takes two to tango: if only one individual evolves a communication system, this will not benefit him if there is not a party which has a similar communication system (Wilson, 1975). This conundrum is especially valid if one assumes that communication systems, or even language, appear during one lucky evolutionary mutation (Chomsky (1980) for example has argued this is the case for the human language faculty). However, the it-takes-two-to-communicate problem can be easily ducked if one assumes gradual evolution of communication. If a certain behaviour involving a precursor to communication gives an animal a reproductive advantage, this will have a chance to propagate in the population. This leaves us with the selective pressures. Darwin believed that communication served to provide the receiver with information on the emotional and motivational state of the sender. Much work on this is based on the idea of group selection: communication evolved because it benefited the group as a whole. However, Hamilton (1963) and (Maynard Smith, 1964) theoretically showed that group selection is not correct: genes do not propagate because they are beneficial for the group, genes are —in Dawkins' (1976) words— selfish.

Tinbergen (1952) was one of the first to explore the second and third issue: on the route which evolution takes towards communication by first relying on non-signal behaviour. He suggested that communicative acts evolved from behaviour which was goal-oriented, but not communicative in nature. As soon as the behaviour was recognised as being predictive for future events, it became a ritualised communicative signal. Mostly the focus is on how signals can evolve, but note that signals can as well be learnt.

The literature contains many different views on the evolution of communication, both inspired by field work and theoretical work — too many to even just sum up in this short section. A good starting point for the novice is (Hauser, 1996) or (Hauser and Konishi, 1999). A more indepth overview can be found in (Oller and Griebel, 2004).

2.2.2 An example: *evolving communicative behaviour on robots*

A large body of research into the evolution of communication exists. For a long time computers were not the tool of choice for modellers of communication, instead the solid mathematics of game theory were used to predict whether evolutionary stable strategies (ESS) for communicative behaviour could be found. An evolutionary stable strategy is a strategy (or behaviour if you will) which over the course of evolutionary time can not elbowed out by another strategy (e.g. Maynard Smith, 1979). Recently models rely no so much on theoretical assumptions and proof, but rather on computer simulations (e.g. Buzing *et al.*, 2005; de Jong, 1999b; Gardenfors, 2004; Marocco and Nolfi, ress; Di Paolo, 1998; Noble, 2000b,a) and recently on experiment with physical robots (e.g. Yanco and Stein, 1992).

As an example we will look at a project of Matt Quin (Quinn, 2001), where he smartly combines artificial intelligence and robotics to study an aspect of the evolution of communication. As Quinn observes, many studies presuppose a communication channel to be present and then continue to study which type of signals (e.g. honest or dishonest signals) will evolve under certain circumstances. Little work has been done on the evolution of the communication channel itself (see for example, Levin 1995). How can communicative behaviour evolve from non-communicative behaviour? Quinn describes an experiment in which a number of Kephera robots perform a non-trivial coordinated movement task, which seems to involve a primitive form of communication.

The robots are equipped with infrared sensors and two motor-driven wheels, and are controlled by an evolved neural network. Quinn was careful not incorporate communication channel in the configuration of the robots or in their initial controllers. The task for robots consists of moving pairwise for at least 25 cm in 10 seconds without colliding with one another. The task is not easy: sensor data is noisy and thus unreliable, and the robot have no information on the position and orientation of other robots. Solving the task successfully involves coordinating the behaviour of two robots: the robots have to make sure they move together in the same direction without losing each other.

Quinn reports how the robots manage to evolve a controller for solving the task. Over the 30 runs, the behaviour of the robots exhibit one striking similarity: the robots take on a particular role, one robot acting as the leader, the other as the follower. The role-taking was not pre-programmed and evolved time and time again.

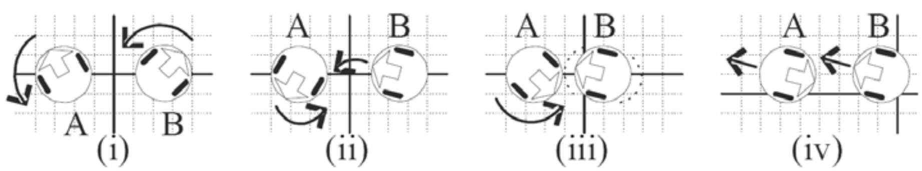

Figure 2: An example of a single interaction: both agents rotate (i) until one agent is aligned with the other (ii) at which moment he moves forward (iii). Agent B now oscillates back and forward until agent A is aligned as well, after which they simultaneously move in the same direction.

Figure 2 show how two agents align with eachother before travelling a distance together. The alignment of the first agent serves as a signal to the second agent, and seems to be a form of communication that is necesarry for succesful completion of the task. The robots evolved a behaviour which functions as “signal and response”. Quinn with his experiments provides a proof-of-concept for the evolution of communicative behaviour in a system without any dedicated communication channels. A mere selective advantage for a task executed by two robots is enough to evolve behaviour which closely resembles proto-communication.

2.3 Grounding

2.3.1 Plugging language into the world

Most computational studies on the evolution of language incorporates a multi-agent system that can learn, or evolve, a communication system of varying complexity that allows the system to communicate about a predefined set of meanings. However, as human communication is about the real world, understanding the underlying principles of language requires an understanding of the mechanisms with which the languages' meanings are rooted in reality. Models based on predefined meanings therefore face what is often referred to as the symbol grounding problem (Harnad, 1990); i.e. how does an individual attribute meaningful content to the symbols he or she is using? In models where the semantics are predefined, it is the designer and/or the observer who attributes the meaning to the language. However, when an agent is allowed to construct its own symbolic representations from its interaction with its environment, it is able to interpret these symbols meaningfully.

Robots need to "ground" their representations in order to truly understand what things mean.

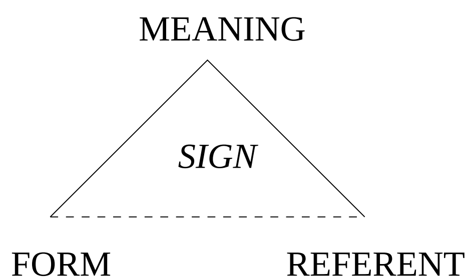

Figure 3: The semiotic triangle illustrates the relations that constitute a sign. When the form is either arbitrary or conventionalized, the sign can be interpreted as a symbol. Adapted from (Ogden and Richards, 1923).

It has been argued that - philosophically speaking - the symbol grounding problem is based on using a wrong definition of symbols (Clancey 1997; Vogt 2002b). Traditionally, following de Saussure (1974), symbols are defined in terms of a *label* (signifier) and a *meaning* (signified). The label (or *form*) is for instance a word or signal, and the meaning is a representation of something. C. S. Peirce (1931-1958), however, has defined symbols a triadic relation between label, meaning and referent, see Fig. 1. According to Pfeifer and Scheier (1999), the symbol grounding problem can only be solved by agents who are both *situated* - i.e. their behaviour are based on their interactions with their environment - and *embodied* - i.e. their behaviours are based on (a history of) their bodily experiences. Vogt (2002) has argued that Peirce's symbols are both situated and embodied. Moreover, they are meaningful by definition. Hence, the symbol grounding problem as posed by Harnad (1990) is philosophically no longer relevant. However, there still remains the hard problem of constructing the triadic relation; a problem that has been coined the *physical symbol grounding problem* (Vogt, 2002). Peirce (1958)

Few studies have tried to tackle this problem in modelling language origins and evolution; both in simulation studies (Cangelosi and Parisi, 1998; de Jong, 1999a) and in robotic models (Marocco *et al.*, 2003; Vogt, 2001; Steels and Vogt, 1997; Steels *et al.*, 2002). In the remainder of this section, a model is presented that allows agents to construct *perceptually grounded* symbols.

2.3.2 An example: *meaning creation*

The first problem that needs to be solved is the physical symbol grounding problem (i.e. creating the semiotic triangle). The problem can be decomposed into three parts: (1) Sensing and pre-processing of raw sensorimotor images, (2) categorisation or fmeaning construction, and (3) labelling. The labelling problem is either trivial (in case of using arbitrary forms) or it is based on learning conventions through language. In this subsection, I will assume the trivial solution and focus on the sensing, pre-processing and meaning formation. Learning conventions will be discussed later in section 2.4.

The discussion of how symbols can be constructed is presented for individual robots. Constructing a semiotic symbol usually starts with a sensori(motor) stimulation based on the robot's interaction with the real world (when embedded in communication, construction can also start upon 'hearing' an expression). Sensorimotor stimulation can be based on a scene acquired by a camera, which may be static as in the Talking Heads experiment (Steels *et al.*, 2002), or dynamic as in more recent experiments of Steels (2004); the activation of infrared, sonar or simple light sensors (Vogt, 2002); the flow of sensorimotor activity (Billard and Dautenhahn, 2000); or the activation of a sensorimotor coupling (Vogt, 2003a). Often, the raw data is pre-processed to reduce the huge amount of data. Typically regions of interest are identified and some feature extraction algorithm is used to describe such regions in terms of feature vectors. How this is done can be quite complex and is not discussed further in this chapter, for details consult the individual papers. When the sensing is based on the activation of sensorimotor couplings, pre-processing - or even categorisation as such - may not be required (Vogt, 2003a). Furthermore, in simulations, the image is often more abstract, such as a bitstring representing mushrooms (Cangelosi and Parisi, 1998) or just random vectors (Smith, 2003), which do not require any more pre-processing.

At the heart of creating semiotic symbols lies - technically - an agent's ability to categorise the perceptual data. Once these categories are in place, the agent can simply associate a label (or form) to this category, thus constructing the symbol (whether this symbol is useful or functional is another question, which will not be dealt with in this paper, but see (Ziemke and Sharkey, 2001; Vogt, 2005b) for discussions.) A number of techniques have been developed that allow a robot to construct categories from scratch with which it is able to recognise or discriminate one experience from another. These techniques usually rely on techniques that have been present in the AI field for quite some time, such as pattern recognition and neural networks. Some researchers use neural networks to associate (pre-processed) sensorimotor images with forms, (e.g. Marocco *et al.*, 2003; Cangelosi and Parisi, 1998; Billard and Dautenhahn, 2000), which —although they work well— makes it hard to analyse how the meanings are represented. Moreover, these techniques are often inflexible with respect to the openness of the system, because typically, the number of nodes in a neural network are fixed. Another technique that is frequently used in grounded models of language evolution is the *discrimination game* (Steels, 1996a).

The aim of the discrimination game is to categorise a sensorimotor experience such that this category distinguishes this experience from other experiences. If such a *distinctive category* (or "meaning") is found, the game is considered a success. If it fails, a new category is formed based on the experience that is categorised, such that discrimination can succeed in a future situation. This allows the agent to construct a repertoire of categories from scratch, as illustrated in figure 4.

The discrimination game in figure 4 has successfully been implemented in the Talking Heads simulation THSim[3] (Vogt, 2003b). Experiments have shown that the discrimination game is typically a fast learning mechanism and is very robust in using different representations for categories. The original implementation used binary trees (Steels, 1996a), which was used in various robotic experiments (Steels and Vogt, 1997; Steels *et al.*, 2002) and simulations (Smith, 2003). Other representations that were used include binary subspaces (de Jong, 1999a), radial basis function networks (Belpaeme, 2001b; Steels and Belpaeme, 2005), neural networks (Berthouze and Tijsseling, 2002a), predicate logic (Sierra-Santibáñez, 2001) and different variants of the

[3]THSim is freely downloadable from `http://www.ling.ed.ac.uk/~paulv/thsim.html`.

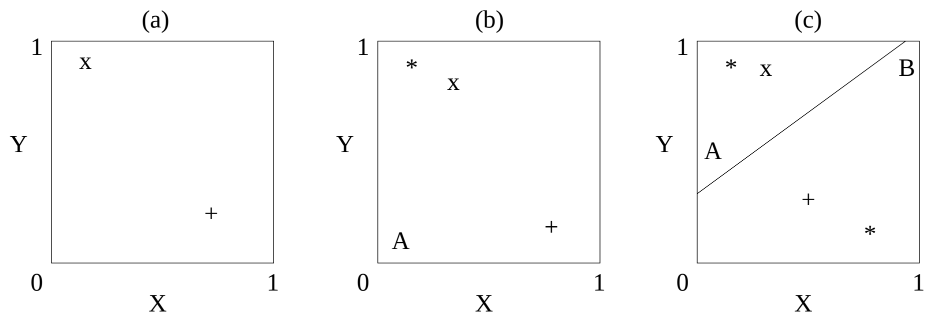

Figure 4: An illustration of three subsequent discrimination games using a prototype representation. The figure shows three instances of a combined feature space and a conceptual space. The x and + are feature vectors of observed objects (e.g., the location of an object in a 2D plane), the * denotes a prototype, while A and B are categories. Each individual plot represents one discrimination game. In game (a), the robot observed 2 objects (x and +), but has not yet formed any categories. Consequently, the game fails and a new category (A) is added to the conceptual space of which the feature vector of the target object serves as an exemplar for its prototype - the * in figure (b). In the second situation (b), again two objects are observed, which are now both categorised with category A. In this case, no distinction can be made. Suppose the + was the target (a.k.a. the topic) of the discrimination game, then a new category is formed by adding the feature vector + to the conceptual space as the new prototype of category B in figure (c). Note that this alters the initial category A. In the third game (c), both objects can be categorised distinctively. Irrespective of which one is the topic, the discrimination game succeeds. Typically, when a discrimination game succeeds, the prototype is moved slightly in the direction of the topic's feature vector.

prototype representation (Vogt, 2005a).

The discrimination game is context dependent: the robot always contrasts the topic with respect to other objects in the context. This has the consequence that the game may succeed, even if the observed feature vector has a relatively large distance to the category's prototype, leading to an overgeneralisation of symbols. However, after a while, the categories become finer grained, thus allowing the agents to resolve overgeneralisation. It is a well known fact, however, that young children also tend to overgeneralise during early word-learning (Bloom, 2000): by way of speaking, to young children everything having four legs is a doggie.

When the sensing typically yields only one region of interest (i.e. there is only one object or action), the discrimination game can only be applied in contrast to some sensorimotor images that are in the robot's memory. In such cases different models can be used as well. The *classification game* was used in experiments with Sony's AIBO where whole (segmented) images of the camera were stored as exemplars (Steels and Kaplan, 2002). The *identification game* was used to categorise the motor flow of robots following each other (Vogt, 2000). The latter used a pre-processing of the raw sensorimotor flow based on constructing delay vectors from time series (Rosenstein and Cohen, 1998). The identification game is very similar to the discrimination game in that the delay vector (or feature vector) is categorised with the nearest prototype, provided its distance was within a certain threshold. If not, the delay vector is added to the ontology as an exemplar.

As mentioned, once a category, which is a representation of the meaning in a semiotic symbol, is in place, the category can be associated with a form. This form may be arbitrary, but in language they need to be conventionalised. In language evolution models, this is often modelled by interactions called *language games* (Steels, 1996b) which will be explained hereafter.

2.4 Evolution of lexicons

2.4.1 How do word and meaning find each other?

Words need to have meaning, otherwise communication is senseless. But how do words get their meaning? How is it that English speakers refer to a chair with the word "chair"? Who came up with the word in the first place? And who decides which words to use for what objects? Does the meaning of words change? How do infants deduce the meaning of thousands of words so quickly? And, could a chair be equally well be named "table"? These and many other questions are what drive research into the *evolution of lexicons*.

Who decides that a rose should be called a "rose"?

Psycholinguists have since long studied how children acquire language (Bloom, 2000). As words are intricately linked to meanings or representations, studying how children learn the meaning of words starts with studying how children represent the world. Even at an early age, words seem to play an important role in learning object representations. Xu (2002) reports experiments where nine-month old babies were tested in keeping track of objects. Objects (a ball or a duck) were shown to the infants, together with a description of the objects ("Look Maggie a ball" or "Look a toy") and where then placed behind a screen. When the screen was lifted and the situation that was unveiled did not match the description just given, the babies showed surprise (measured by the time they looked at the objects). Seemingly the infants used the linguistic descriptions of the objects to keep track of what could be expected behind the screen. While doing so they were able to generalise across categories: they seemed to know that “toy” comprised both “ball” and “duck”. These results suggest that language plays an important role in the acquisition of object concepts.

Studies show that children start to conceptualise by first relying on perceptual cues. From the age of nine months on they rely more on functional information and on linguistic labels to learn and refine exisiting concepts (Booth and Waxman, 2002). How children exactly learn concepts using lexical labels in not quite clear. What is known, is that children start by overgeneralising linguistic concepts (for example by using the word “dogie” for both dogs and cats) after which they steadily refine the concept using negative and positive examples.

There are many computational models which try to capture how words (which in the literature are often also called *signals* or *forms*) get associated with meanings (also called *concepts*, *objects* or *representations*; pick your favourite). All models share a number of properties:

Agent-based. The models are based on building an *agent*: an abstract implementation of an individual acquiring words and meanings. Some models use only one agent, but most have tens or even thousands of agents communicating with each other.

Lexicon. An agent has a lexicon, containing words. Sometimes the lexicon is fixed, for example the agent starts with ten words. Sometimes the agents starts with an empty lexicon and acquires new words as it goes along.

Meanings. An agent needs to have meanings to associate words with. Again, in some models the meanings are given and fixed, but in the most interesting models the agent acquires new meanings and adapts existing meanings as it goes along.

Associations between words and meanings. Each agent needs to have some way of associating words with meanings. In the most basic form this is a graph connecting words and meanings. However, recently much work has relied on using *association matrices*, in which words and meanings are connected to each other with a *strength value*. Lower values signify that the connection between a particular word and meaning is not strong, and vice versa: high values signify that the connection between a word and meaning is particularly strong.

$$
\begin{array}{c}
\\
c_1 \\
c_2 \\
\vdots \\
c_n
\end{array}
\begin{array}{c}
\begin{array}{cccc} t_1 & t_2 & ... & t_m \end{array} \\
\begin{pmatrix}
0.1 & 0.6 & ... & 0.0 \\
0.0 & 0.1 & ... & 1.0 \\
\vdots & \vdots & s_{ij} & \vdots \\
0.1 & 0.0 & ... & 0.0
\end{pmatrix}
\end{array}
$$

This matrix is used to both *interpret* a word (looking up the meaning of a word) or to *produce* a word (looking up the word associated with a meaning). When using only one matrix, there is no difference between the behaviour of the agent while interpreting and producing. Sometimes it is desirable to make this difference. Humans for example actively use less words than they can understand: their performance for interpreting and producing words is asymmetrical. This is modelled by using two matrices: one production matrix and one interpretation matrix (e.g. Oliphant, 1996).

Other representations can be used as well to model to association between words and meanings, for example a neural network might do the job just as well (Berthouze and Tijsseling, 2002b). Notice how the matrix representation nicely captures *synonyms* (a meaning having more than one word) and *homonyms* (one word having several meanings).

Learning rules. The agent needs a set of rules with which it can set the associations between words and meanings. These rules have to be so that the agents' lexicons, meanings and associations are all tuned to let the agent communicate with other agents. The rules which agents use are often simple, but subtle changes to the rules often have a profound influence on the behaviour of the agents and on the outcome of the simulation (Smith, 2004).

A huge body of work exists which studies the evolution of lexicons and meanings. Topics include the dynamics of acquiring a lexicon (among others Belpaeme, 2001a; de Jong, 1999a; Hurford, 1989; Nolfi, 2005; Oliphant, 1996; Smith, 2001, 2004; Steels and Kaplan, 1998, 1999), the interaction between words and meanings (Belpaeme and Bleys, 2005b; Cangelosi, 2001), the acquisition of hierarchical meanings (Belpaeme *et al.*, 1998; Steels, 2003) and the emergence of "dialects" in spatially distributed populations.

The relation between words and meanings is so complex that a seperate scientific field exists for it called *semiotics*. Semioticians ponder on how words, objects and the objects' perception are linked together (also see section 2.3). (For an introduction see Chandler, 2001).

2.4.2 An example: *colour categories*

The example here uses computer modelling to study how a lexicon can influence the agent's internal representations, and vice versa, how internal representations influence the agent's lexicon. This phenomena is known as *linguistic relativity*, and has received renewed interest from linguists and cognitive scientists in the last few years (Gumperz and Levinson, 1996)[4]. The example focuses on colour: on how colour categories and colour words evolve (Belpaeme and Bleys, 2005b,a; Steels and Belpaeme, 2005).

Human colour categories are very remarkable: everyone, no matter where you live or what language you speak, seems to have similar colour categories. If you're a native living in the jungles of West Papua or if you've been raised on the snowfields in Alaska, chances are high that you will have colour categories for colours that we know by the English terms "red", "blue", "yellow", "green", etcetera (Berlin and Kay, 1969; Kay and Regier, 2003). This is very remarkable and has been explained by many as being a genetical trait common to all humans: the number of colour categories and the kind of colour categories is genetically determined. Just as having ten fingers or 32 teeth is genetically determined. However, there are some exceptions to this universal pattern. For example, anthropologists have studied tribes having colour categories which do not comply to this universal scheme (Davidoff *et al.*, 1999; Roberson *et al.*, 2000). Some scholars believe that a different view on the acquisition of colour categories is needed. Some believe that colour categories are learnt by every individual and are to a large extent determined by the environment we live in, others believe that colour categories are learnt from interaction with others. This latter view states that you learn your colour categories from your parents and peers mostly through linguistic interactions. You learned what red exactly is because you have heard the word "red" being used in situations where the colour red was present.

The model described here tries to unveil the mechanism through which learning colour categories could happen, and demonstrates how computational modelling can be used to inform a cross-disciplinary audience.

The model that we use is *agent based*. Each agent is a piece of computer code implementing the bare necessities for colour perception, colour categorisation and colour naming. Colour perception is implemented by taking a psychological model of human colour perception, the CIE L*a*b* model (Fairchild, 1998), which maps RGB-colour values onto three values which accurately models how humans psychologically perceive colour. Next, a model is needed for colour categories. Colour categories are perceptual categories, meaning that they are re-

[4]For a gentle introduction see "Chomsky's theory on trial: Does the language you speak control the way you think?" *New Scientist*, 30 November 2002.

lated to the senses (just like auditory categories). Perceptual categories can be represented by prototypes: a category which is sensitive to a prototype and has a gradually decreasing sensitivity to inputs further from the prototype. In our models we chose to model colour categories as points in the CIE L*a*b* space, with a Euclidean distance relation between the points to model membership to a category. Next, colour categories need to be lexicalised: each colour category should be able to get one or more colour words. This is done by associating the colour categories with random strings using an association matrix as mentioned above.

Now the agents still need to evolve their categories and words: this is done by letting the agents interact linguistically. Using a very simple protocol —dubbed the *guessing game*— agents talk to each other about colour. For this two agents are selected, one acting as the speaker and the other as the hearer. The speaker and hearer are both presented a number of colours, and the speaker chooses one of them and speaks the name of the colour to the hearer. Now the hearer guesses which colour the speaker meant.

This is a very simple interaction, and is typical for most interactions between humans where one person tries to communicate something to the other and the other has to deduce what the topic of the conversation is. There are many ways in which this interaction can go wrong: for example, the speaker might not have a word for the colour he's trying to communicate, or the hearer might not have the word he just heard in his lexicon. All these failures are opportunities to which the agents respond by extending or adapting their set of colour categories and colour words. For details on algorithms see (Belpaeme and Bleys, 2005b; Steels and Belpaeme, 2005).

Simulations typically involve 10 to 20 agents. When running the simulation we observe how, at first, the agents do not have any colour categories or words, but quickly develop into agents which are able to communicate perfectly on colour. Not totally unexpected, due to the nature of the algorithm, the colour categories of the agents start to get "tuned" to each other: a prerequisite for linguistic communication (Steels and Belpaeme, 2005). However what is remarkable is that when the simulations are run many times we observe how the agents of all these different simulations —all started with a different random seed— show a similarity which resembles the typology of colour categories in human cultures (Belpaeme and Bleys, 2005b).

The simulations show that the nature of human colour categories are possibly not fixed by our genes, but suggest that colour categories are to a large extent determined by the people we communicate with. The remarkable similarity across the colour categories of different cultures can be explained as subtle constraints, formed mostly by the nature of our colour perception, on top of a cultural acquisition process.

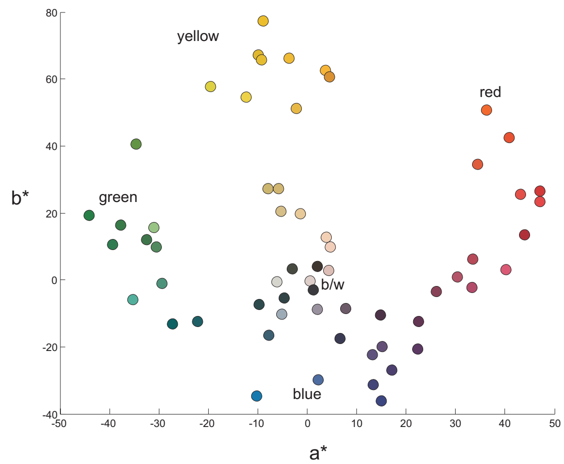

Figure 5: A plot of colour categories of 10 agents in the 2-dimensional plane.

2.5 The emergence of compositionality

2.5.1 It English what be without would?

One of the most distinctive features of human languages is the high degree of compositionality they have. This means that the utterances of human languages are highly structured: parts of the utterances map onto parts of the whole meaning of these utterances. For instance, in the phrase "orange square", the word "orange" refers to the colour orange and the word "square" to a square. In contrast, in a holistic phrase such as "kick the bucket" (an English expression referring to dying), no part of the utterance refers to a part of its meaning. One influential hypothesis on the emergence of compositionality suggests that during the course of evolution, human languages have changed into compositional languages from initially holistic protolanguages (Wray, 1998). Many models have been developed, which provide support for this idea (Brighton, 2002; Smith *et al.*, 2003; Kirby, 2001; Vogt, 2005a).

"Use the force we must". Even though Yoda speaks gibberish, his language adheres to certain rules, making it compositional.

These models have shown that learners can learn the language of an adult population while observing only a part of the language. We know that children do not hear every possible sentence of the a language, for example from English, in order to learn English. Now computer models can show that this “bottleneck” in transmitting language from a parent to a child is not problematic, and even stimulates the emergence of compositionality. Holistic languages, which have no syntax forcing you to come up with a new word for every meaning that you wish to express, can not be learnt when there is a transmission bottleneck. Only languages with a certain regularities in the form of rules, or syntax, can become communication systems which can stabely transmitted over generations of language learners. This can be understood by realising that when the learners become adults and start communicating to the next generation of learners, they have no means to produce expressions about objects / meanings they have not encountered before. Compositional languages, however, could allow a learner to produce utterances for previously unseen meanings when the learnt structures can be combined. For instance, if an agent has learnt the proper structures from the phrases “orange square”, “orange triangle” and “red square”, it would be able to produce the phrase ”red triangle”, even though it would never have encountered a red triangle before. Brighton and others have shown that —given a learning mechanism that can discover such compositional structures— a compositional language can emerge from an initially holistic language, *provided* the language is transmitted through a bottleneck. In a way, the language changes to become more learnable for future generations.

2.5.2 An example: *iterated learning model*

As mentioned, one of the crucial assumptions is that the agents are given a sophisticated induction mechanism. Typically, these mechanisms are adopted from well known machine learning techniques, such as neural networks (cfr. the study in Smith *et al.* 2003), minimum description length (e.g. Brighton, 2002), and alignment-based learning (e.g. Kirby, 2001; Vogt, 2005a). Although most of these models have been implemented in ungrounded models, Vogt’s study was implemented in a simulation of the Talking Heads experiment. Here agents evolved a language to communicate about coloured geometrical shapes. In this model the semantics is not predefined, but co-develops with the language. The agents could detect four perceptual features of objects: the three components of the RGB colour space and one feature indicating the shape of an object (note that there is no attempt here to study the emergence of human-like colour categories as was the case in Steels and Belpaeme 2005.) The semantic structures developed from a combination of the discrimination game (see section 2.3) to construct categorical features (elements in one dimension) and an inducer to discover *conceptual spaces*[5] of one or more dimensions that

[5]The term conceptual spaces (Gärdenfors, 2000) is used to denote an n-dimensional space in which categories are represented by prototypes. The conceptual space is

could serve to represent linguistic categories, such as colours or shapes (note that there was no restriction on which dimensions would constitute a conceptual space - all possible combinations were allowed). One the other hand, syntactic structures could be discovered by looking for alingments in the utterance level. Suppose, for instance, an agent has heard the following utterance-meaning pairs in which the alignments are underlined: $\underline{ab}cd - \underline{1}1\underline{0}$ and $\underline{ab}dc - \underline{1}0\underline{0}$. From such utterances, the agents could learn the following grammar:

$S \rightarrow ab/1\#0\ B$
$B \rightarrow cd/\#1\#$
$B \rightarrow dc/\#0\#$

Here S is the top sentence rule, $\#$ is a wild card and B is a non-terminal node. The sentence thus rewrites to a string "abB" where B is either "cd" or "dc" with their corresponding meanings. (Note that the actual implementation in Vogt's model is much more complex, but the example illustrates the idea.) Initially, word-meaning pairs are created at random, as in the previously seen language game models, but by chance alignments are found in the signal space. The alignments found in the meaning space are due to the regular nature of the Talking Heads environment. The model thus investigated the following twofold hypothesis:

1. The emergence of compositional linguistic structures is based on exploiting regularities in (possibly random and holistic) expressions, though constrained by semantic structures.

2. The emergence of combinatorial semantic structures is based on exploiting regularities found in the (interaction with the) world, though constrained by compositional linguistic structures.

The model combines the two most familiar approaches taken in modelling language evolution: the *iterated learning model* (Brighton, 2002; Kirby and Hurford, 2002) and the language game model. The iterated learning model (ILM) typically implements a vertical transmission of language, in which the population contains adults and learners, the learners acquire the language through interactions with adults. At some given moment the adults are replaced by the learners and new learners enter the population and the process repeats, thus providing a generational turnover. Typically in ILMs (a part of) the language is transmitted from one generation to the next in one pass; without competition, but see (Kirby, 2000) for a model with competition. The integration of the ILM with the language game allows for competition

spanned by n quality dimensions that relate to some (preprocessed) sensorimotor quality. Gärdenfors has argued that conceptual spaces can form the semantic basis for linguistic categories.

between different rules and structures, but it requires more passes through the language in order for the language to be learnt sufficiently well.

The experiments reported in Vogt (2005a,c) have revealed that compositional structures emerges rapidly under all conditions investigated, even in the absense of a bottleneck. When the language is transmitted through a bottleneck, compositionality remains stable. However, when a bottleneck is not present, the compositional structures are only stable when the learners act as speakers too (Vogt, 2005c)). In the original ILM models of Brighton (2002); Kirby (2001), only adults acted as speakers and learners as hearers. This finding can be explained that when the learners start to speak before they acquired the entire language, they already face a learning bottleneck. However, this bottleneck is not imposed by the experimenter, as was the case in the other ILM models, but is implicit to the development of the learner. This may be an important result, because it may help to explain why children are so good at learning grammar early in life. Moreover, this property may even explain why children are thought to be the driving force for the development of grammar in Nicaraguan sign language (Senghas *et al.*, 2004).

A more complex model implemented in an extended physical version of the Talking Heads experiment is being developed by Steels and his co-workers (Steels, 2004; Steels and Baillie, 2003). In this experiment the cameras do not look at a static scene pasted on the white board, but the cameras observe a dynamic scene played in front of them, such as 'pick up red ball'. The events are processed through a visual processing system, which - although advanced - is still very limited. Only slow movements can be captured and only a few objects can be recognised, but only *after* training the visual module. The pre-processed events are then matched with top down generated world knowledge, which is represented in the form of predicate calculus of which the basic building blocks are predefined (Steels and Baillie, 2003).

Using these event descriptions, the guessing game (or *description game*) proceeds. Where possible, the agents use the knowledge (lexicon, syntax and semantics) they already acquired, but when events or semantics cannot be described with the given knowledge, new parts of the language is invented, abducted or induced. New words, semantic categories, syntactic categories and hierarchical structures can be constructed using some complex techniques, which are largely based on existing techniques from computational linguistics. This way grounded construction grammars (Lakoff, 1987) can develop, as some preliminary experiments have shown (Steels, 2004).

Other models on the evolution of grammar include, for instance, language game like models (Batali, 2002; Gong *et al.*, 2004), a model of a biological evolution of a Universal Grammar (Briscoe, 2000), and models based on recurrent neural networks (Batali, 1998; Tonkes and

Wiles, 2002).

2.6 Language change and diversity

2.6.1 How fót became foot

Computer models have also been used successfully for studying language change and linguistic diversity. There is an amazing number of different languages. At present there are about 6000 different languages, and this number was probably considerably higher before the colonization of the Americas and Australia. The number of different dialects is higher still. This is in an amazing fact, as all these languages are communicatively equivalent in principle (although of course there are always differences in specialization of vocabulary) and because in other primate species, call systems tend to be innate (although there might be culturally determined differences in chimpanzee calls Crockford *et al.* 2004). The question why languages are so diverse poses itself.

Your grandmother —trying hard as she may— will use a language differently from you. Language always is in flux.

Linguistic diversity, of course, is closely related to language change. All languages change over time and this change often occurs faster than is traditionally taught in historical linguistics (Labov, 1994, e.g.). Even though there is no central control over language, and linguistic knowledge only exists in the brains of individuals, languages tend to change as a whole, so that after a while all speakers of a certain dialect end up having gone through the same language change. It is also possible that one group of speakers of a formerly uniform dialect undergoes one change, while another group undergoes a different change. Thus a language split can occur. Furthermore, even though it is assumed that words are stored independently in a speaker's brain, words that are similar tend to undergo similar changes. An example is the Dutch words “huis” (house) and “muis” (mouse). These only differ in their vowel from their English equivalents. What makes these examples particularly interesting is that there are dialects in which “huis” has undergone the change towards the standard language, whereas the word for mouse (the less frequently used of the two) remains in an older form “muus”. Not all words undergo the same change simultaneously. Apparently speakers within a population influence each other, as do words within a language.

The spread of language changes is not only determined by factors that can be considered linguistic, but also by a large number of social factors. Linguistic variants are often generated on the basis of functional factors. Sounds tend to be weakened, because this is what happens if speech is produced rapidly and casually. Morphemes tend to disappear, while frequently occurring combinations of words can be grammaticalized. It could also be true that languages tend to less “marked” (or easier to learn) constructions, determined by adaptations to language of the human brain. However, whether a speaker decides (consciously or unconsciously) to adapt a certain variant of the language depends on the social status, gender and age of the speaker, of the social status of the language variant, and of the number of other speakers that already use the variant. Social pressures can serve both to speed up language change or to slow it down. Thus there is a complex interaction between purely functional, linguistic pressures and social pressures.

The facts that language is learned, and variants are adapted by individuals, but that the acceptability of a change is determined by what is used in the population and that both linguistic and social pressures determine what changes can occur, cause the dynamics of language change to be complex and hard to predict. Therefore, agent-based computer models are an ideal means of investigating these dynamics. There are three related ways in which language change and linguistic diversity can be studied with computer simulation. One can try to model how language changes over time, how linguistic diversity can emerge and what social factors play a role in language change, -emergence and -death.

The temporal dynamics of language change have been investigated from a mathematical point of view since at least the 1970’s (see Altmann 1985 for references). As for more realistic models, the mathematics becomes intractable, computer models can provide otherwise unobtainable insights. One of the topics of research is the s-shaped curve with which language change takes place: it starts out slow, gathers speed and finally slows down again as the change is almost completed. Wang and Minett (2005) provide an overview of such research, as well as a model of their own. Dynamics of language change can also be investigated with models in the language game paradigm, by allowing agents to be added to and removed from the population (e.g. de Boer, 1999b; Steels and McIntyre, 1999).

Social factors in the emergence of linguistic diversity and language change have been investigated by different researchers. Typically these models use a spatially distributed population of agents that can adopt linguistic features from each other depending on their relative social status. An example of is the work of Nettle (1999).

Closely related to this work is the work on explanation of linguistic diversity. In this work it is investigated under what circumstances multiple languages remain stable in a population. Again, populations are usually distributed spatially in these models. An example is the work by Livingstone and Fyfe (1999b).

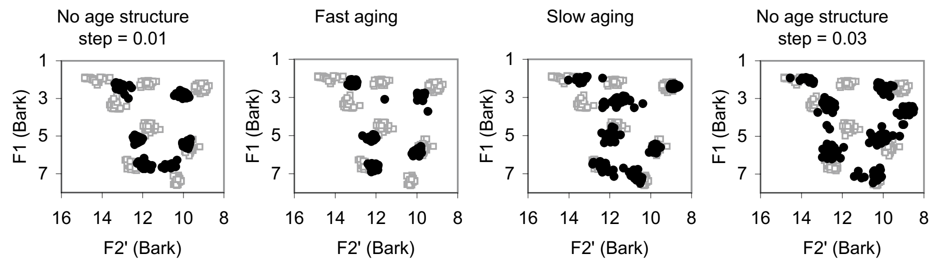

Figure 6: Preservation of vowel systems in changing populations. The original vowel system is indicated in gray, the vowel system after 15 000 imitation games is shown in black. The leftmost and rightmost frames do not age. The center two frames do age. Note better preservation of the original vowel system in the populations with aging.

2.6.2 An example: *vowel systems in changing populations*

The example that will be presented here has been taken from de Boer and Vogt (1999). In this paper it was investigated how vowel systems emerge in changing populations and how existing vowel systems can remain stable in changing populations. Here the focus will be on how vowel systems can remain stable. Understanding stability is also important if one wants to understand language change. Although it is true that languages can change rapidly over time, it is equally true that they can remain stable for relatively long periods of time. As there are many factors that can destabilize language in a population, such as the rate in which speakers are born and die, the influence of other languages and the fact that speakers often reduce their articulations in rapid speech, it is an interesting question how language remains stable in a changing population.

The model is entirely the same as the imitation game described in the example of a model of the evolution of speech. There are two important differences, however. The first is that whereas the population in the original imitation game is entirely static, in the modified imitation game, individuals can be removed and added to the simulation. This is called the population flux. Individuals that are added to the population are empty and have to learn the existing vowel system from scratch, while the linguistic knowledge of individuals that are removed is lost. Individuals are removed and added at random so that there can be fluctuations of population size. Removal (death) is independent of the age of agents. The second difference is that the ease with which agents can learn a language can now be varied over time. The 'ease' is modeled by the step size with which agents change their vowel representations in reaction to an imitation game. In the original imitation game this step size is fixed, while in the modified imitation game, this step size can vary, so that young agents use a larger step size than old agents.

The influence of the first change is that for large population flux, vowel systems tend to collapse into simple three-vowel systems. For small population flux, there is no discernable difference with a static population. For intermediate population flux, vowel systems tend to become simpler than ones that would emerge in static populations. These results are intuitive (the less stable the population, the simpler the vowel system) but one should be careful in making direct comparison with human populations. The way in which vowel systems are learned is quite different in the imitation games than in humans. Also, factors like the meaning of words, the phonetic context in which sounds occur and the influence of neighboring languages are important in human language change, whereas they are not modeled here. However, there are some indications that sound systems tend to be simple in languages that are used in very unstable populations (for example Creole languages that have recently emerged) while complex sound systems are typically found in relatively isolated regions (for example Khoisan languages that are spoken by people living in the Kalahari desert, or languages spoken in the Caucasus).

The second change causes the repertoires to remain more stable over time. Agents that change from a large step size to a small step size as they age preserve a vowel system better than populations of agents that either have a fixed large step size or a fixed small step size. This is illustrated in figure 6. All populations investigated in this experiment had the same population flux. The population flux was sufficiently low that there was almost no disruption of the vowel system. The observed phenomenon can be explained by the fact that conservative old agents provide a stable target towards which younger, empty agents can move rapidly. As the younger agents become older, they provide stable targets to newly added agents in turn. The behavior of the model appears to provide an explanation for why adults are so much worse at learning a language than infants. Perhaps this is not just an unfortunate side effect of aging, but an adaptation to keeping language stable.

The results of this model show how computer models can be applied to questions of language change and stability, but also that relatively small and simple changes to one model (the basic imitation game) can provide interesting insights into other aspects of language.

3 Caveats and communication

Computational modelling is a form of synthetic science, as opposed to analytic science which gathers and interprets real world data. Building and running computer models and interpreting their results is not always easy, and often mistakes are made which compromise the reputation which computational modelling has in other disciplines. In this section we wish to elaborate on some of the most common mistakes and issue a few warnings, or caveats, for the beginning modeller. Research does not exist if it is not communicated properly, and especially with cross-disciplinary research this requires a careful approach. We try to provide some kind advice and hope that this will be helpful in avoiding the mistakes that we have made.

Computer models are an abstraction of reality. In general, this is what science is about: scientists abstract reality and try to uncover general laws governing the universe, life and everything. Newton's laws of motion are an abstraction of reality, but still they serve extremely well to predict the behaviour of bodies, from crashing cars to comets swirling through our solar system. Nevertheless, Newton's laws are mere abstractions of the real world. But they are good abstractions, for they serve to make accurate descriptions and even predictions of how the world behaves.

No one has criticised Newton's laws for their abstractness (agreed, Einstein relativity theory does a better job at predicting how bodies travelling near light speed behave), but chances are that your computer model will receive quite some headwind for it being "too abstract". The reason for this is that some people mistake simplicity for inaccuracy. It is therefore imperative that you justify every design choice of your model, and admit to any simplifications and assumptions that you were forced to make before others will hold these against you.

Try also to run only the necessary experiments. When a computer model is finished it is always tempting to run it for hundreds of different parameter settings. This is good to "test drive" the model, but often this fiddling with your model will last for months without ever obtaining any useful results. Try to make out for yourself which hypothesis you will test and construct a number of well aimed experiments.

Sometimes however it is necessary to run the model with hundreds, or sometimes even thousands of different parameter settings; for example, three parameters each having 20 possible values, already leaves you with 3^{20} runs to execute. Make sure that the model is bug free before starting those runs, and try to optimize your code as much as possible so that simulations are, if not speedy, at least not mindbogglingly slow. Also, when doing simulations it is tempting to log as much as your hard drive can possibly store. This is not a good idea, as you will never have the courage or the time to analyse all that data. Think before you start experimenting on what data you will need, and build your logging according to those needs. Also, it is useful to spend some time thinking about measures. The measures will give you a quantitative impression of your results; for example, when studying lexicon change, you will need a measure for the amount of change in the lexicons of agents. And last, but not least, try to visualise your results.

A simple graph will give you a quick impression of the quality of your results, something which can not be gotten from looking at ASCII log files.

Another problem with computer modelling is that it often is, in the words of John Maynard Smith, fact-free science. People construct a computer model, let is run and observe results which resemble —more or less— some real-world phenomenon. There is however no reason to believe that the dynamics at work in the simulation are in any way similar to the dynamics at work in the real-world phenomenon. When simulating flocks of birds and schools of fish (Reynolds, 1987) you can mimic flocking behaviour by letting the agents abide to three simple rules: move towards the centre of your local mates, orient towards the average direction of the flock and avoid collisions. Although the simulations exhibit flocking behaviour which resembles actual flocking behaviour so closely that even Hollywood got interested, there is nothing which warrants the conclusion that actual flocking animals use those three rules. It is important to realise this and assume the necessary humility when drawing conclusions from your simulation results.

Eventually you will have built a good computer model and will have obtained good results: you are ready to write an academic paper. First comes the choice of the audience you will write for: will you submit your paper to conference or journal with likeminded researchers, or will you try to cross the disciplinary gorge and write for an audience outside your field of expertise? Always make sure that your style of writing and the terminology you use are in sync with the audience you are addressing. Whichever audience you write for, your paper will have to contain a clear definition of the research context, the problem at hand and the goal that you wish to reach or the question to which you seek an answer. Try to explain your model such that it is self-contained and reproducible: the hall mark of a good paper is that anyone can reimplement your model and fiddle around with it. Also make sure to have enough experimental data, carefully reported using the proper statistics. Try not to just include experimental data, but make sure that the experimental data is compared to some baseline. This gives the reader an idea of how good your results are. If, for example, you are studying the influence of language on categories acquisition, make sure to include a run where language has been switched off: this is your baseline against which your results should be compared. Make sure that you report a statistically significant number of runs, do *not* present only one lucky run. Also, make sure to do control experiments. Like the apocryphic story of the man drinking a glass of gin followed by a glass of water, and after that a glass of whiskey followed by water, and then rum and water. At the end he concludes that as water was the only constant in his experiment, it must the water which caused drunkenness. Of course, if he had done the control experiment —only drinking water— he would have realised he was wrong.

4 Wrap up

Up to date a lot has been achieved by using robots to study the origins and evolution of language. However, it is clear from the overview in this chapter, that we are still far from understanding the entire picture. Can we ever design a group of robots that can evolve languages similar to human languages? Personally, I[6] think that the human brain and body is so complex that we may never be able to unravel all its secrets —like we may never unravel the complete working of our universe, which is similarly complex. Nevertheless, I think that we can use robotics profitably to answer some of the questions that are posed in our effort to understand language evolution. Much future research will need to focus on ecological models, models of grammar, categorisation of more human-like concepts, and on models of the theory of mind. Furthermore, the models will have to be scaled up at many levels, such as population size, sensorimotor complexity, and complexity of the world in which the robots operate.

A recently started European project called New Ties aims at developing a simulation in which a large community of robots (over 1,000 agents) evolve a cultural society, including language. The society 'lives' in an environment where they have to cooperate in order to survive. The agents will evolve and learn techniques to deal with the constraints set by the environment in order to improve their viability. In addition, the agents will be designed to evolve language as the motor for evolving the cultural society (Vogt and Divina, 2005). The techniques these agents will use are heavily based on the language game techniques that have been developed so far. One of the major innovations - apart from its complexity and ecological setting - will be a design on the Theory of Mind, which at a later stage is intended to become subject of the evolution.

Given that many problems in modelling language evolution on robots relate to the difficulties in establishing joint or shared attention to the reference of the communication, studies into the nature of this ability is extremely important. Only few studies are known that investigate how joint attention can emerge (e.g. Kaplan and Hafner, 2004). One of the key aspects with respect to joint attention and related issues is that agents need to infer the intentions of other agents. This ability can loosely be characterised by the Theory of Mind (Premack and Woodruff, 1978). It may well be that the ToM is one of the major innovations of the human species with respect to language origins, and therefore its origins deserves more attention in models of language evolution.

With respect to the origins and evolution of grammar and compositionality, research needs to done to study how the learning mechanisms have evolved that allow individuals to construct grammatical and

[6] Paul Vogt.

compositional structures. Up to now, all studies on grammaticalisation have assumed that such learning mechanisms exist and therefore only investigate how grammar can emerge given these mechanisms. Another important direction that needs to be tackled is in relation to the grounding of more abstract and higher level symbols, such as - for instance - number systems, arithmetic, planning and 'feelings' (internal states). Up to now, all research has focused on the emergence of language about events or objects that are directly observable to the robots. We humans often use language to communicate about events that happened in the past or that may happen in the future. Some work on the development of time concepts is done (De Beule, 2004), but it would be good if a robot could communicate, for instance, the presence of an interesting object at a distant location, which the robot has visited before.

Most techniques used so far are based on simple language games where some aspect of a visible entity is communicated in one direction to investigate learning techniques. However, human language use is much more based on dialogues. Future robotic models should investigate how dialogues can aid in evolving language. This could be particularly interesting for applications where robots develop their own language in order to cooperate in environments we don't know, such as planets, or using sensors that we find difficult to read, such as infrared, sonar or other exotic sensors.

An excellent collection of literature with relation the language evolution is maintained by Jun Wang at `http://www.isrl.uiuc.edu/~amag/langev/index.html`.

References

Altmann, G. (1985). On the dynamic approach to language. In T. T. Ballmer, editor, *Linguistic Dynamics: Discourses, Procedures and Evolution*, pages 181–189. Walter de Gruyter, Berlin.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight, editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge University Press, Cambridge.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press.

Belpaeme, T. (2001a). Reaching coherent color categories through communication. In B. Kröse, editor, *Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC01), Amsterdam, The Netherlands*, pages 41–48.

Belpaeme, T. (2001b). Simulating the formation of color categories. In B. Nebel, editor, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'01), Seattle, WA*, pages 393–398, San Francisco, CA. Morgan Kaufmann.

Belpaeme, T. and Bleys, J. (2005a). Colourful language and colour categories. In K. Dautenhahn, C. Nehaniv, R. te Boekhorst, D. Cañamero, D. Polani, and J. Hewitt, editors, *Proceedings of the AISB 2005 convention, Hertfordshire, UK*. To appear.

Belpaeme, T. and Bleys, J. (2005b). Explaining universal colour categories through a constrained acquisition process. *Adaptive Behavior*. To appear.

Belpaeme, T., Steels, L., and van Looveren, J. (1998). The construction and acquisition of visual categories. In A. Birk and Y. Demiris, editors, *Proceedings of the 6th European Workshop on Learning Robots*, Lecture Notes on Artificial Intelligence 1545, pages 1–12, Berlin. Springer.

Berlin, B. and Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. University of California Press, Berkeley, CA.

Berrah, A.-R. and Laboissière, R. (1999). Species: An evolutionary model for the emergence of phonetic structures in an artificial society of speech agents,. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life, Lecture Notes in Artificial Intelligence*, volume 1674, pages 674–678. Springer, 1999, Berlin.

Berthouze, L. and Tijsseling, A. (2002a). Acquiring ontological categories through interaction. *The Journal of Three Dimensional Images*, **16**(4), 141–147.

Berthouze, L. and Tijsseling, A. (2002b). Acquiring ontological categories through interaction. *Journal of Three Dimensional Images*, **16**(4), 141–147.

Billard, A. and Dautenhahn, K. (2000). Experiments in social robotics: grounding and use of communication in autonomous agents. *Adaptive behavior*, **7**(3/4).

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

Booth, A. and Waxman, S. (2002). Object names and object functions serves as cues to categories for infants. *Developmental Psychology*, **38**, 948–957.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, **8**(1), 25–54.

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

Briscoe, T., editor (2002). *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press, Cambridge.

Buzing, P., Eiben, A., and Schut, M. (2005). Emerging communication and cooperation in evolving agent societies. *Journal of Artificial Societies and Social Simulation*, **8**(1).

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

Cangelosi, A. and Parisi, D. (1998). The emergence of ”language” in an evolving population of neural networks. *Connection Science*, **10**, 83–93.

Cangelosi, A. and Parisi, D., editors (2001). *Simulating the Evolution of Language*. Springer Verlag, London.

Chandler, D. (2001). *Semiotics: the Basics*. Routledge, London. Also available at http://www.aber.ac.uk/media/Documents/S4B/semiotic.html.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, **3**, 1–61.

Crockford, C., Herbinger, I., Vigilant, L., and Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology*, **110**, 221–243.

Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik, editors, *Universals of Human Language*, volume Volume 2 Phonology, pages 93–152. Stanford University Press, Stanford.

Crowley, T. (1994). *An introduction to historical linguistics*. Oxford University Press, Auckland.

Davidoff, J., Davies, I., and Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, **398**, 203–204.

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press, Oxford.

De Beule, J. (2004). Creating temporal categories for an ontology of time. In *Proceedings of the 16th Belgian-Dutch Conference on Artificial Intelligence (BNAIC04)*.

de Boer, B. (1999a). *Self-organisation in vowel systems*. Ph. D. Thesis AI-lab Vrije Universiteit Brussel, Brussels.

de Boer, B. (1999b). *Self Organisation in Vowel Systems*. Ph.D. thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel.

de Boer, B. (2000). Self-organisation in vowel systems. *Journal of Phonetics*, **28**(4), 441–465.

de Boer, B. and Vogt, P. (1999). Emergence of speech sounds in changing populations. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artifical Life, Lecture Notes in Artificial Intelligence 1674*, pages 664–673. Springer Verlag, Berlin.

de Jong, E. (1999a). Autonomous concept formation. In T. Dean, editor, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99)*, pages 344–349. Morgan Kaufmann, San Francisco, CA.

de Jong, E. D. (1999b). Analyzing the evolution of communication from a dynamical systems perspective. In *ECAL99*, pages 689–693. Springer-Verlag.

de Saussure, F. (1974). *Course in General Linguistics*. Fontana/Collins, New York. Manuscript from 1916, translated by Baskin, W.

Di Paolo, E. (1998). An investigation into the evolution of communication. *Adaptive Behavior*, **6**(2), 285–324.

Dunbar, R. (1996). *Grooming, gossip and the evolution of language*. Faber and Faber Ltd, London.

Fairchild, M. D. (1998). *Color Appearance Models*. Addison-Wesley, Reading, MA.

Gärdenfors, P. (2000). *Conceptual spaces*. The MIT Press, Bradford Books, Cambridge, MA.

Gardenfors, P. (2004). Cooperation and the evolution of symbolic communication. In K. Oller and U. Griebel, editors, *Evolution of Communication Systems: A Comparative Approach*, pages 237–256. MIT Press, Cambridge, MA.

Gong, T., Ke, J., Minett, J. W., and Wang, W. S.-Y. (2004). A computational framework to simulate the co-evolution of language and social structure. In *ALife 9*, Boston, MA, U.S.A.

Gumperz, J. J. and Levinson, S. C. (1996). *Rethinking Linguistic Relativity*. Studies in the Social and Cultural Foundations of Language 17. Cambridge University Press, Cambridge.

Hamilton, W. (1963). The evolution of altruistic behavior. *American Naturalist*, **97**, 354–356.

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

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

Hauser, M. D. and Konishi, M., editors (1999). *The design of animal communication*. The MIT Press, Cambridge, MA.

Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, **77**(2), 187–222.

Kaplan, F. and Hafner, V. (2004). The challenges of joint attention. In L. Berthouze, H. Kozima, C. Prince, G. Sandini, G. Stojanov, G. Metta, and C. Balkenius, editors, *Proceedings of the Fourth Conference on Epigenetic Robotics: Modelling Cognitive Development in Robotic Systems, Genova, Italy*, volume 117 of *Cognitive Studies*, pages 67–74. Lund University.

Kay, P. and Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, **100**(15), 9085–9089.

Ke, J., Ogura, M., and Wang, W. S.-Y. (2003). Optimization models of sound systems using genetic algorithms. *Computational Linguistics*, **29**(1), 1–18.

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

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

Kirby, S. (2002). Natural language from artificial life. *Artificial Life*, **8**(2), 185–215.

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

Labov, W. (1994). *Principles of Linguistic Change:internal factors*. Blackwell, Oxford.

Ladefoged, P. and Maddieson, I. (1996). *The Sounds of the World's Languages*. Blackwell, Oxford.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press, Chicago.

Laver, J. (1994). *Principles of Phonetics*. Cambridge University Press, Cambridge.

Levin, M. (1995). The evolution of understanding: A genetic algorithm model of the evolution of communication. *BioSystems*, **36**, 167–178.

Lieberman, P. (1998). *Eve spoke: Human language and human evolution*. W.W. Norton, New York.

Liljencrants, J. and Lindblom, B. (1972). Numerical simulatons of vowel quality systems. *Language*, **48**, 839–862.

Lindblom, B., MacNeilage, P., and Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of language universals. In M. Butterworth, B. Comrie, and O. Dahl, editors, *Explanations for language universals*, pages 181–203. Walter de Gruyter & Co., Berlin.

Livingstone, D. and Fyfe, C. (1999a). Modelling the evolution of linguistic diversity,. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life, Lecture Notes in Artificial Intelligence*, volume Volume 1674, pages 704–708. Springer, Berlin.

Livingstone, D. and Fyfe, C. (1999b). Modelling the evolution of linguistic diversity,. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life, Lecture Notes in Artificial Intelligence*, volume Volume 1674, pages 704–708. Springer, Berlin.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press, Cambridge.

Maddieson, I. and Precoda, K. (1990). Updating upsid. *UCLA Working Papers in Phonetics*, **74**, 104–111.

Marocco, D. and Nolfi, S. (in press). Emergence of communication in embodied agents: Co-adapting communicative and non-communicative behaviours. In A. C. et al., editor, *Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*. Singapore: World Scientific.

Marocco, D., Cangelosi, A., and Nolfi, S. (2003). The emergence of communication in evolutionary robots. *Philosophical transactions: Mathematical, Physical and Engineering Sciences*, **361**(1811), 2397–2421.

Maynard Smith, J. (1964). Group selection and kin selection. *Nature*, **201**, 1145–1147.

Maynard Smith, J. (1979). The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology*, **47**, 209–221.

Nettle, D. (1999). Using Social Impact Theory to simulate language change. *Lingua*, **108**, 95–117.

Noble, J. (2000a). Co-operation, competition and the evolution of prelinguistic communication. In J. R. H. Chris Knight and M. Studdert-Kennedy, editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge University Press, Cambridge.

Noble, J. (2000b). Talk is cheap: Evolved strategies for communication and action in asymmetrical animal contests. In J.-A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, and S. Wilson, editors, *Proceedings of the Conference on Simulation of Adaptive Behavior 2000*, pages 481–490, Honolulu, Hawaii. MIT Press.

Nolfi, S. (2005). Emergence of communication in embodied agents: co-adapting communicative and non-communicative behaviours. *Connection Science*. In press.

Ogden, C. and Richards, I. (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and the Science of Symbolism*. Routledge and Kegan Paul Ltd., London.

Oliphant, M. (1996). The dilemma of Saussurean communication. *BioSystems*, **37**(1-2), 31–38.

Oller, D. K. and Griebel, U., editors (2004). *Evolution of Communication Systems: A Comparative Approach*. MIT Press, Cambridge, MA.

Oudeyer, P.-y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, **233**(3), 435–449.

Peirce, C. S. (1931-1958). *Collected Writings*, volume 1 - 8. Harvard University Press, Cambridge, MA.

Pfeifer, R. and Scheier, C. (1999). *Understanding Intelligence*. The MIT Press, Cambdrige, MA.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, **1**(4), 515–526.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In J. Kelemen and P. Sosík, editors, *Proceedings of the European Conference on Artificial Life 2001*, Lectures Notes in Computer Science, pages 357–366, Berlin. Springer.

Redford, M. A., Chen, C. C., and Miikkulainen, R. (1998). Modeling the emergence of syllable systems. In *Proceedings of the 20th Annual Meeting of the Cognitive Science Society*, pages 882–886. Erlbaum, Hilsdale, NJ.

Reynolds, C. (1987). Flocks, herds, and schools: a distributed behavioral model. *Computer Graphics*, **21**(4), 25–34.

Roberson, D., Davies, I., and Davidoff, J. (2000). Color categories are not universal: replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General*, **129**(3), 369–398.

Rosenstein, M. and Cohen, P. (1998). Symbol grounding with delay coordinates. In *Working notes of the AAAI-98 Workshop on The Grounding of Word Meaning*, Menlo Park, CA. AAAI Press.

Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997a). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, **25**, 255–286.

Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997b). Major trends in vowel system inventories. *Journal of Phonetics*, **25**, 233–235.

Senghas, A., Kita, S., and Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, **305**(5691), 1779–1781.

Seyfarth, R., Cheney, D., and Marler, P. (1980). Vervet monkey alarm calls: Semantic communication in a free-ranging primate. *Animal Behavior*, **28**, 1070–1094.

Sierra-Santibáñez, J. (2001). Grounded models as a basis for intuitive reasoning. In B. Nebel, editor, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'01), Seattle, WA*, pages 401–406, San Francisco, CA. Morgan Kaufmann.

Smith, A. D. (2001). Establishing communication systems without explicit meaning transmission. In J. Kelemen and P. Sosik, editors, *Proceedings of the European Conference on Artificial Life (ECAL01), Prague*, pages 381–390, Berlin. Springer.

Smith, A. D. M. (2003). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, **9(2)**, 559–574.

Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology*, **228**(1), 127–142.

Smith, K., Brighton, H., and Kirby, S. (2003). Complex systems in language evolution: The cultural emergence of compositional structure. *Advances in Complex Systems*, **6**(4), 537–558.

Steels, L. (1996a). Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the International Conference on Multiagent Systems (ICMAS-96)*, pages 338–344, Menlo Park, CA. AAAI Press.

Steels, L. (1996b). Self-organizing vocabularies. In C. Langton and T. Shimohara, editors, *Proceedings of the Conference on Artificial Life V (Alife V) (Nara, Japan)*, Cambridge, MA. The MIT Press.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in cognitve science*, **7**(7), 308–312.

Steels, L. (2004). Constructivist development of grounded construction grammars. In D. Scott, W. Daelemans, and M. Walker, editors, *Proceedings Annual Meeting Association for Computational Linguistics Conference, Barcelona*, pages 9–16.

Steels, L. and Baillie, J.-C. (2003). Shared grounding of event descriptions by autonomous robots. *Robotics and Autonomous Systems*, **43**(2-3), 163–173.

Steels, L. and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language. A case study for colour. *Behavioral and Brain Sciences*, **24**(8). To appear as target article.

Steels, L. and Kaplan, F. (1998). Stochasticity as a source of innovation in language games. In G. Adami, R. Belew, H. Kitano, and C. Taylor, editors, *Proceedings of the Conference on Artificial Life VI (Alife VI) (Los Angeles, California)*, Cambridge, MA. The MIT Press.

Steels, L. and Kaplan, F. (1999). Bootstrapping grounded word semantics. In T. Briscoe, editor, *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press, Cambridge.

Steels, L. and Kaplan, F. (2002). AIBO's first words: The social learning of language and meaning. *Evolution of Communication*, **4**(1), 3–32.

Steels, L. and McIntyre, A. (1999). Spatially distributed naming games. *Advances in complex systems*, **1**(4), 301–324.

Steels, L. and Vogt, P. (1997). Grounding adaptive language games in robotic agents. In P. Husbands and I. Harvey, editors, *Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), Complex Adaptive Systems*, Cambridge, MA. The MIT Press.

Steels, L., Kaplan, F., McIntyre, A., and Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray, editor, *The Transition to Language*, pages 252–271. Oxford University Press, Oxford, UK.

Tinbergen, N. (1952). Derived activities: Their causation, biological significance, origin and emancipation during evolution. *Quarterly review of biology*, **27**, 1–32.

Tonkes, B. and Wiles, J. (2002). Methodological issues in simulating the emergence of language. In A. Wray, editor, *The Transition to Language*, chapter 11. Oxford University Press, Oxford.

Vennemann, T. (1988). *Preference Laws for Syllable Structure*. Mouton de Gruyter, Berlin.

Vogt, P. (2000). *Lexicon grounding on mobile robots*. Ph.D. thesis, Artificial Intelligence Lab, Vrije Universiteit Brussel, Belgium.

Vogt, P. (2001). Bootstrapping grounded symbols by minimal autonomous robot. *Evolution of Communication*, **4**(1), 89–118.

Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research*, **3**(3), 429–457.

Vogt, P. (2003a). Anchoring of semiotic symbols. *Robotics and Autonomous Systems*, **43(2)**, 109–120.

Vogt, P. (2003b). THSim v3.2: The Talking Heads simulation tool. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, and J. Ziegler, editors, *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL)*. Springer Verlag Berlin, Heidelberg.

Vogt, P. (2005a). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence*. In press.

Vogt, P. (2005b). Language evolution and robotics: issues on symbol grounding and language acquisition. In A. Loula, R. Gudwin, and J. Queiroz, editors, *Artificial Cognition Systems*. in press.

Vogt, P. (2005c). On the acquisition and evolution of compositional languages. *Adaptive Behavior*. In press.

von Frisch, K. (1969). *The Dance Language and Orientation of Bees*. Seale Hayne.

Wang, W. S.-Y. and Minett, J. W. (2005). The invasion of language: emergence, change and death. *trends in Ecology & Evolution*, **20**(5), 263–269.

Wilson, E. (1975). *Sociobiology: A new synthesis*. Belknap Press, Harvard, MA.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, **18**, 47–67.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, **85**, 223–250.

Yanco, H. and Stein, L. A. (1992). An adaptive communication protocol for cooperating mobile robots. In J.-A. Meyer, H. L. Roitblat, and S. Wilson, editors, *Proceedings of the Conference on Simulation of Adaptive Behavior 1992*, pages 478–485. MIT Press.

Ziemke, T. and Sharkey, N. E. (2001). A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica*, **134(1-4)**, 701–746.

Image sources

- brain: `http://www.phon.ox.ac.uk/~jcoleman/`
- dictionary: `www.henrylim.org/Dictionary.html`
- robots: `http://arti.vub.ac.be/pictures/Ecal_97_Beach/Page.html`
- yoda: `http://www.formfunctionemotion.net/mt-static/images/`
- granma: `http://www.scu.edu.au/news/images/`