RESEARCH REPORT

Language evolution as a Darwinian process: computational studies

Pierre-Yves Oudeyer · Frédéric Kaplan

Abstract This paper presents computational experiments that illustrate how one can precisely conceptualize language evolution as a Darwinian process. We show that there is potentially a wide diversity of replicating units and replication mechanisms involved in language evolution. Computational experiments allow us to study systemic properties coming out of populations of linguistic replicators: linguistic replicators can adapt to specific external environments; they evolve under the pressure of the cognitive constraints of their hosts, as well as under the functional pressure of communication for which they are used; one can observe neutral drift; coalitions of replicators may appear, forming higher level groups which can themselves become subject to competition and selection.

Keywords Language evolution · Language change · Darwinian evolution · Genes · Memes · Replicators · Learning bias · Functional pressure · Self-organization

Mapping biology to language

Biological evolution happens through the differential replication and selection of genes, with variation coming from mutations or cross-over. Cultural evolution, and in particular language evolution, happens through the social interactions between people, mediated by many different kinds of behaviours and contexts. While these two processes seem at first sight rather different, researchers have in fact considered strong parallels between the two since the invention of the theory of natural selection by Charles Darwin (1859). As a matter of fact, soon after Darwin published his book On the Origins of Species, describing how mechanisms of inheritance, variation, and selection could explain the evolution of biological organisms, August Schleicher published a book using very similar concepts to describe the birth and death of languages, and even the phenomena of language speciation (Schleicher 1863). Interestingly, Schleicher had developed the notion of language trees, representing their genealogy, which inspired the biologists' cladogram or phylogenetic trees (Mufwene 2005).

This mapping between biological evolution and language evolution was put forward again and made more detailed when the substrate and the structure of the genetic code was discovered (Watson and Crick 1953). There are two strands of correspondences that were developed. A first strand tried to map units and structures in the genetic space directly to units and structures in the linguistic space. It was based on the observation that genetic information was organized as a sequence of discrete nucleotides, forming DNA molecules. The particularity of this organization is that there is a molecular system for reading the sequence, in which nucleotides are read three by three (these groups are called codons) and command the formation of a corresponding sequence of amino-acids, which will then fold up and form a three-dimensional protein. The

P.-Y. Oudeyer (✉)
Sony Computer Science Laboratory Paris, Paris, France
e-mail: py@csl.sony.fr
URL: http://www.csl.sony.fr/~py

F. Kaplan
Ecole Polytechnique Federale de Lausanne, CRAFT,
Lausanne, Switzerland
URL: http://www.fkaplan.com

association between triplets of nucleotides and amino-acids seems to be largely arbitrary. The discovery of these structures led researchers to think of nucleotides as an alphabet, codons as words, and genes as sentences whose meaning would be the proteins associated to the genes (Berlinski 1972). This led to the proposal that the genetic code itself was a language (Searls 2002). Nevertheless, this mapping is controversial and a number of strong arguments against the relevance of these detailed correspondences at the levels of units were developed (Stegmann 2004; Tsonis et al. 1997).

A second kind of parallel was developed in which the focus was on the process of evolution rather that on the units themselves (Mufwene 2005; Croft 2000, 2002; Steels 2004). This was in fact a continuation of the nineteenth century view of language evolution as a (neo-)Darwinian process. The common process between genome and language evolution was here the following: (1) there exists a population of units capable of replication, (2) replication is not perfect: modifications can appear, (3) the units have different levels of efficiency in replication, which produces differential replication. This high-level formulation, sometimes conceptualized as a generalization of Darwin's theory of natural selection (Hull 1988), has the advantage of specifying neither the structure of units nor the mechanisms of replication and variation. And indeed, researchers found ways to instantiate it into biological or language evolution by filling in those missing slots with the corresponding specific structures and mechanisms (Croft 2000). As far as biology is concerned, the units are genes, the mechanisms of replication are those associated with meiosis/mitosis, and the mechanisms of variation are mutations and cross-over. As far as language is concerned, a wide variety of instantiations have been proposed. The units of replication were conceived as *ideas*, *mnemotypes*, *idene*, *culture-type*, *socio-genes*, *tuition* (van Driem 2003), ranging from simple abstract concepts like words or expressions to complex neural structures implementing associations between phonological forms and meaning. Perhaps the most well-known notion of cultural unit of replication is the *meme* introduced by (Dawkins 1976):

> Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. If a scientist hears,

or reads about, a good idea, he passed it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain. As my colleague N.K. Humphrey neatly summed up an earlier draft of this chapter: '... memes should be regarded as living structures, not just metaphorically but technically. When you plant a fertile meme in my mind you literally parasitize my brain, turning it into a vehicle for the meme's propagation in just the way that a virus may parasitize the genetic mechanism of a host cell. And this isn't just a way of talking—the meme for, say, 'belief in life after death' is actually realized physically, millions of times over, as a structure in the nervous systems of individual men the world over'. (Chapter 11 from Richard Dawkins, "The Selfish Gene")

Linguistic memes, sometimes called *linguemes* (Croft 2000), are themselves a population of very diverse kinds of units: phonological features, phonemes, syllables, rules of phoneme sequencing, lexicons, rules of syntax, semantic categories, systems of world categorization, constructions mapping combinations of words and complex meanings, prosodic structures, social conventions involving gestures and gaze to coordinate linguistic interactions, etc. Dawkins gives imitation as an example of mechanism of replication for units of language evolution. As a matter of fact, all kinds of linguistic activities, which can be much more complex than just imitation, like conversation or reading, provoke the replication of linguistic units. The consequence is that "leaping from brain to brain" is a very complex process that can happen through a variety of mechanisms. What provokes variation is therefore also very diverse: bad perception, erroneous interpretation, exaggeration, etc.

This shows that the conceptualization of language evolution as a Darwinian process may take quite different forms for different authors and is often presented only at a rather general level. Yet, in order to be useful, this conceptualization must be precise, detailed and operational. A first attempt to do so is to focus on particular linguistic examples, typically particular phonemes or words for which we have good data about their evolution in the past, and to find out a detailed and causal explanation in terms of Darwinian processes (Croft 2000; Blevins 2004, 2006). Nevertheless, if language evolution is a Darwinian process, then it means that many of its features are systemic: they are the outcome of the complex

interactions between replicators and environment. But this kind of system is extremely complicated, and this is why in biology the use of mathematical and computational modelling, in particular in population genetics, has been crucial in order to test and develop the neo-Darwinian theory of evolution (Crow and Kimura 1970; Maynard-Smith 1982; Ancel and Fontana 2000). These kinds of models could also be used for the study of language evolution as a Darwinian process. Computational models of the origins of language have been flourishing in the recent years (Cangelosi and Parisi 2002; Kirby 2002; Steels 2003). They allow us to develop intuitions about the complex phenomena characterizing language dynamics. Thus, they should be considered as tools for thought, and a number of them already had a significant impact on the debates and theories of the origins of language (Cangelosi et al. 2006). Some of them were designed to model the origins of particular features of language in a cultural Darwinian perspective. Nevertheless, there exists to our knowledge no general description focusing on the cultural Darwinian perspective and showing how a range of computational models presented in the light of Darwinian concepts can illustrate the diverse properties of the associated mechanisms. This is the purpose of this paper. A number of experimental studies will now be presented, showing operational implementations of linguistic replicators and their associated mechanisms of differential replication. These computational studies are tools for thought. We will try to show how they can help articulate the possible similarities and the differences between linguistic and genetic replicators. The common point of all experiments is that they consist of populations of agents initially devoid of linguistic convention, and that will progressively and culturally build each time a new (simple) linguistic system. A first series of examples will focus on the coherence constraints imposed on linguistic replicators due to their use as communication systems by their hosts. In particular we will see that only very specific mechanisms of replication allow for the efficient formation of shared linguistic conventions. Then, we will present an experiment showing how linguistic replicators can adapt and evolve under the specific constraints due to the external environment. We will then review an experiment studying the role of learning biases in the replication process and see how it can influence language evolution. Finally, we will describe an experiment in which we can observe the formation of coalition of replicators, forming groups which are themselves subject to competition and selection.

Viewing language evolution as a Darwinian process: experimental studies

Linguistic replicators have specific properties compared to biological replicators: they form a system that permits communication. For instance, in a vocabulary in which words are associated to concepts/meanings and are used to draw the attention of several speakers towards a particular referent in a given context, synonymy and homonymy tend to be reduced ensuring efficient communication. We will see that not every differential replication process permits the emergence of such communication systems. Typically, each linguistic interaction involves the semiotic triangle: there is a form (e.g. a word), an associated meaning, and an associated referent in a particular context. This entails that three kinds of entities can be replicated through communication: forms, meanings, and associations within certain forms and certain meanings. As a matter of fact, each of these macro-entities consists itself of a variety of entities which are also replicators. For example, words are composed of syllables which are themselves composed of vowels and consonants through phonotactic rules, and all these hierarchical entities can replicate differentially. Although all these replicators are constantly interacting, the experiments we will now present make a number of simplifications that allow us to develop a better understanding of the fundamental dynamics associated to various functional, internal and environmental constraints. For example, in the first experiment we will describe, we suppose that there is only one meaning in the world that agents inhabit, and that two possible words can be associated to this meaning. This experiment will show some basic properties of the replication mechanism so that a simple convention can be adopted by a population, i.e. so that linguistic coherence can be reached (speakers associate the same word to the same meaning). This experiment will then be made more complex in following sections, allowing to study progressively more complex phenomena like linguistic distinctiveness (speakers associate different words to different meanings).

Basic dynamics of linguistic coherence

Let us consider a simple problem: N agents have to choose between two conventional names c_1 and c_2. We will consider three simple models representative for many more complex ones studied in the field. The first model is an imitation-based model (model A) and the two others are frequency-based (model B and C). In model A (last heard policy), the speaker simply

Fig. 1 Model A (last heard policy). The speaker simply produces the conventional name he heard last as a listener. Competition between two conventional names c_1 and c_2 in a population of 100 agents. N1 and N2 represent, respectively, the number of agent using c_1 and c_2. Initially, 50 agents choose c_1 and 50 other agents choose c_2. Several oscillations are observed before convergence (reprinted from Kaplan 2005)
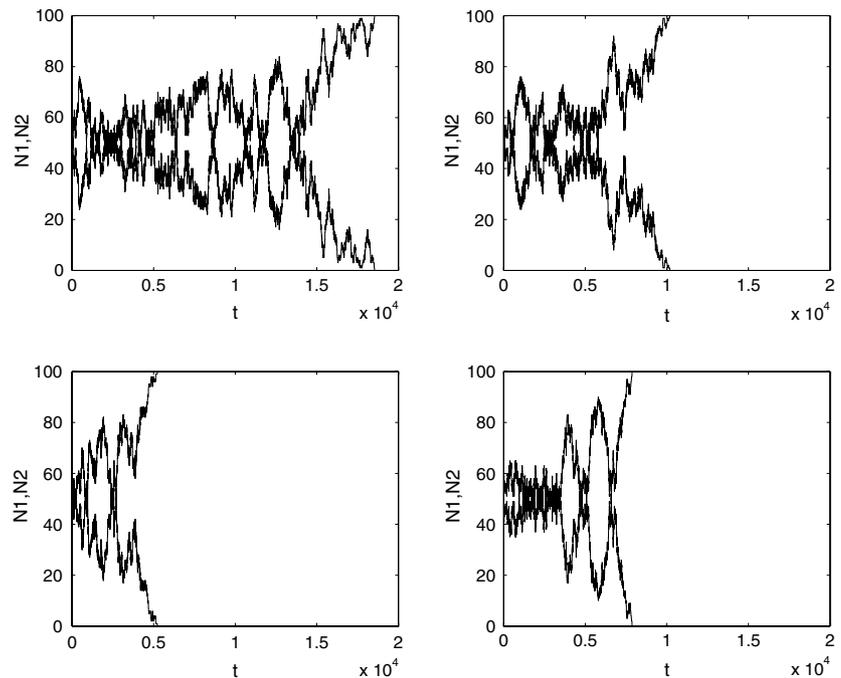
produces the conventional name he heard last as a listener. In model B (most frequently heard policy), the speaker produces the name that he has heard most frequently as a listener. In model C (proportional to the frequency heard policy), the speaker produces any name that he has heard with a frequency f as a listener with a probability proportional to f. These three types of replication processes could be seen a possible models of how cultural replication occurs.

Figures 1, 2, 3 present four representative sample evolutions respectively for models A, B and C. With imitation-based model A, the population eventually converges to a state of complete co-ordination. However, convergence happens only after a long series of oscillations. With model B, convergence towards a single conventional name also occurs. However, the oscillations observed are much smaller. As soon as a convention spreads more in the population than the other, its domination seems to amplify even more over time. For model C, on the contrary, dynamics tend to maintain the distribution of c_1 and c_2 over time, after an initial drift.

An in-depth study of these three models reveal that despite their apparent similarity the types of dynamics they create are extremely different. Among the three models studied, only model *B* creates self-reinforcing dynamics that permit a fast coordination of the entire population towards the use of a single conventional name. Model *A* is approximatively similar to a random walk, converging in quadratic time. On the contrary, the dynamics of model *C* tend

to maintain the distribution of the convention at a fixed level.[1]

What we must remember from these results is that not all cultural transmission systems create a differential replication process that ensures the domination of some linguemes over others. Simple models of cultural transmission solely based on imitation are not sufficient to permit linguistic co-ordination. In that sense the dynamics of linguistic replication are likely to be different from the ones characterizing epidemiological processes, which have often been presented as a possible metaphor for cultural transmission (Sperber 1984).

Implicit evaluation

The positive feedback loop introduced in model B creates a winner-take-all situation where one conventional name dominates. The replicator that eventually wins has no special properties and any new run of the simulation would lead to the selection of a different one. Let us now consider a set of replicators of unequal qualities, i.e. for example words which sounds are more or less prone to be deformed when transferred from one agent to the other due to noise. We model these

---

[1] Experimental results and qualitative interpretations suggest that self-reinforcing dynamics of model *B* converge in $N \cdot \log(N)$, where N is the population size. These experimental results can also be interpreted using various formalisms including Markov chains, stochastic games and Polya processes (see Kaplan 2005 for a review).

Fig. 2 Model B (most frequently heard policy). The speaker produces the name that he has heard most frequently as a listener. Competition between two conventional names c_1 and c_2 in a population of 100 agents. N1 and N2 represent respectively the number of agent using c_1 and c_2. Initially, 50 agents choose c_1 and 50 other agents choose c_2. Dominance of one name over the other tends to increase over time (reprinted from Kaplan 2005)
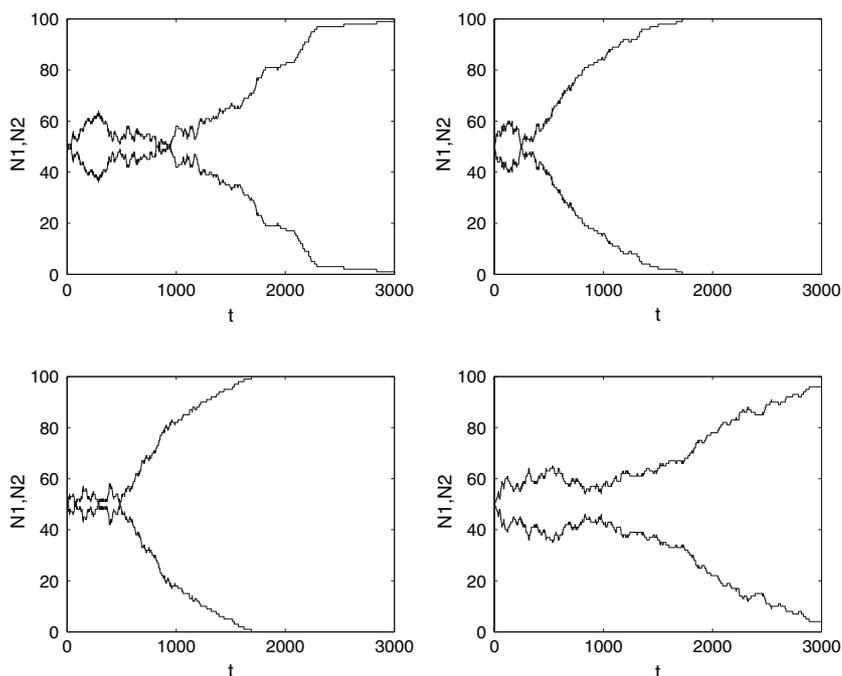
differences in a crude way by associating to each replicator a mutation probability $P_m(c_i)$ between 0 and 100%. During each interaction, a random test is done to check whether the replicator has been transmitted successfully or not. In case of failure, the replicator is transformed to another replicator randomly picked among all the possible ones.

More precisely, let us consider a new experiment in which there is still only one meaning possible, but this time there are 50 possible words/replicators that can be associated to it by individuals. Replicators have a mutation probability that grows linearly with their index.[2] Figure 4 shows, for 1,000 simulations, the distribution of the winning replicator for a population of 50 agents. Thirty-nine percentage of the runs converge towards the three best replicator (replicators with low mutation rates). The agents perform a *collective optimization*: they spontaneously converge towards the best names. The phenomena is based on an *implicit evaluation* of the solutions similar to the one described for foraging behaviour in ant colonies (Dorigo et al. 1997). It means that the agents are not evaluating individually the quality of each convention for choosing the more robust ones. Ill-adapted replicators simply

mutate more often and cannot propagate as easily as the others (Croft 2000).

Reorganization in the presence of a population flux

Figure 4 shows that although there is a clear tendency to focus towards the most robust replicators, suboptimal solutions are sometimes collectively chosen. This is due to *premature convergence*. If, for instance, a very good replicator appears in the population later during the experiment, it is probable that it will not be picked up because the positive feedback loop would have already caused the agents to converge towards a suboptimal one. The situation is different in an open system where agents are entering and leaving the population. Indeed, new agents entering the population have no special preference for the dominant replicator. They can discover a better replicator and maybe, if it is really more robust than the one currently dominating, such an "outsider" might eventually win.

However, introducing new agents in the population is not without risk. Intuitively, we can imagine that if the flux of new agents is too big, convergence towards a shared system of replicators may not be possible anymore. Let us define P_r as the probability of replacing an old agent before every new interactions during an experiment. Figure 5 summarizes the results of a large number of experiments investigating the relationship

---

[2] Thus for replicator c_i, the probability of mutation can be expressed by the following formula:

$$P_m(c_i) = \frac{i}{\text{Number of replicators}}. \tag{1}$$

Fig. 3 Model C (proportional
to the frequency heard
policy). The speaker produces
a name that he has heard as a
listener with a probability
proportional to the frequency.
Competition between two
conventional names c_1 and c_2
in a population of 100 agents.
N1 and N2 represent
respectively the number of
agent using c_1 and c_2. Initially,
50 agents have a bias toward
c_1 and 50 others a bias toward
c_2. After an initial drift
period, the distribution tends
to be maintained (reprinted
from Kaplan 2005)
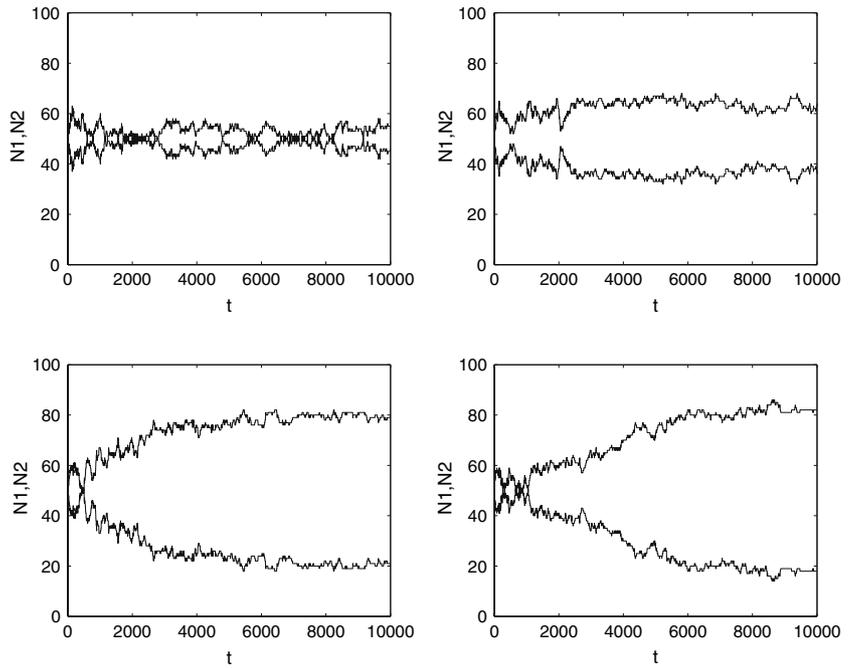
between the average convergence time T_c over 100
simulations for different population size and agent re-
newal rate P_r. The results suggests the existences of a
threshold over which convergence is either not possible
or extremely long. This threshold depends on the
population size.

It can be argued that this behaviour shows some
analogies with phase transitions occurring in certain
physical systems, for instance in experiments about
ferromagnetism. The renewal rate P_r plays the role of
the *temperature* in such systems. A too high tempera-
ture leads to a complete disorganization of the ferro-
magnetic system. In the same manner, if the renewal
rate is too important the cultural transmission process
is broken and convergence towards a shared set of
replicators is not obtained any more. In physics, this

threshold is called *critical temperature*. By analogy we
can use the term *critical flux* (Kaplan 2001).

High temperatures have both good and bad effects
on physical systems. They can lead to their disorga-
nization but also they can permit restructurations
leading to more optimal configurations (optimization
techniques such as simulated annealing are based on
this effect). Following our initial remark and in order
to see if an agent flux can lead to a better selection of
the replicator, we can measure for several values of P_r
the proportion of simulation runs that end up with
one of the three best replicators dominating in the
experiment similar to the one of Fig. 4. Figure 6 plots
the proportion of simulation runs converging toward
one of the three best solutions for different popula-
tion flux P_r before the critical flux is reached.

Fig. 4 Implicit evaluation of
linguistic replicators.
Distribution for 1,000
simulation runs of the
winning replicator for a
population of 50 agents.
Thirty-nine percent of the
runs converge towards the
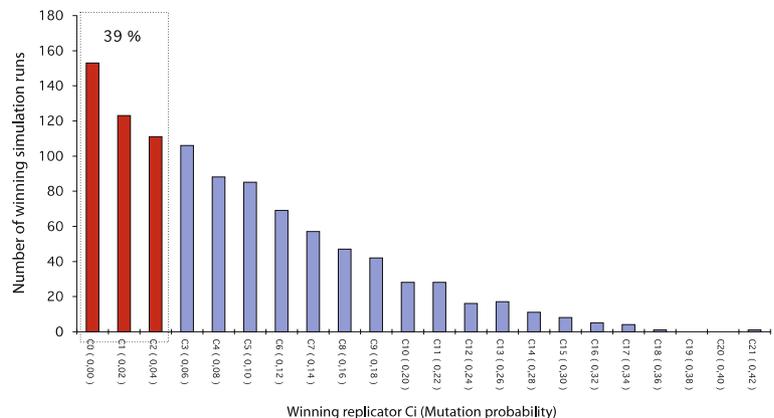three best replicators
(replicators with low
mutation rates)

Fig. 5 Comparison of mean convergence time T_c for different population flux P_r and for different population size N. The results suggest the existence of a threshold depending on the population size N above which convergence is either not possible or extremely long. This threshold characterizes the maximal renewal rate that the system can tolerate
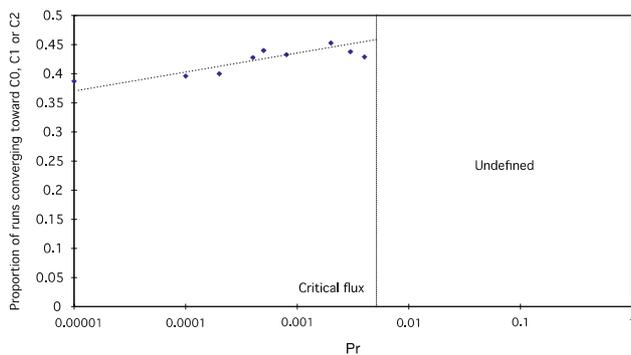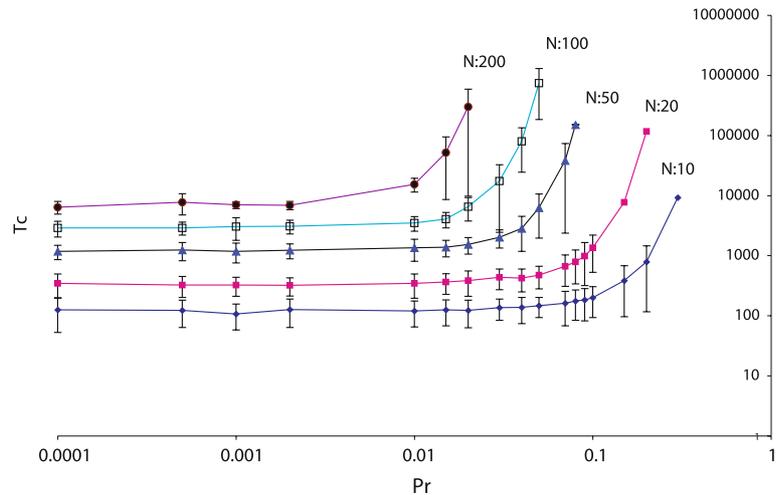
Fig. 6 Proportion of simulation runs (100 trials) converging toward one of the three best solutions for different population flux P_r for a population of $N = 100$ agents. Although the effect is small, the results suggest that reorganization is optimal near the edge of this threshold. In these cases, the flux permits to decrease suboptimal convergences

Although the effect is small, the results suggest that reorganization is optimal near the edge of this threshold. In these cases, the flux permits to decrease suboptimal convergences.

To summarize, if each agent uses the most widely spread replicator from its own point of view, then a positive feedback loop is created, leading to the domination of one replicator. This dynamic is "blind" and does not prefer any replicator per se. But if some replicators are more difficult to transmit, they will implicitly be left aside. Thus, the best replicator tends to be chosen by the population. Finally, the presence of a flux of agents in the system avoids premature convergence. Better replicators (more robust, easier to learn) tend to be selected. The arrival of new agents enables a continuous parallel search for solutions that can replace the ones currently dominating. If needed it can cause reorganization in the linguistic system.

Basic dynamics of linguistic distinctiveness

For efficient communication, it is better that different words are associated with different meanings and vice versa. This obvious remark actually constrains systems of linguistic replicators in many important ways. Indeed, the replication process must not only address the issue on linguistic coherence but also permit linguistic distinctiveness. Let us consider a model in which individuals have to establish conventionalized associations between several words and several meanings. Each agent is now equipped with an associative memory, which is a list of word-meaning pairs with a numeric score. It is used to find the best word associated to a given meaning and reversely to find the best meaning associated to a given word. As in the models of the previous section, agents choose the association with the highest score when several solutions are possible. The associative memories of the agents are initially empty. Associations are progressively created as the agent interacts with other agents.

Studies of such systems were initiated by Steels (1996) in the mid 1990s. Several other experiments rapidly showed how collective dynamics could permit that each name eventually becomes associated with a single context and each context with a single convention (Hutchins and Hazlehurst 1995; Oliphant 1997; Arita and Koyama 1998; Kaplan 2001; De jong and Steels 2003; Vogt 2005; Baronchelli et al. 2006).

Some of the most interesting dynamics of such self-organizing lexicons are obtained in the presence of noise. Let us consider that each word/replicator c_i is modelled with an integer value between 0 and 1,000. Each time a word/replicator is transmitted, a random number between $-B/2$ and $+B/2$ is added to its value. Thus, B is a measure of the global noise level. Each

agent is equipped with a *filter* permitting to select all the words/replicators in its associative memory of which the values are at a distance D less than $D = B$. The structure of an interaction is the following: agent 1 randomly chooses a meaning s_1 among the different meanings available and uses a word c_1 to express this meaning. If it does not have words associated with this meaning, the agent creates a new one (a random integer between 0 and 1,000). Then, c_1 is transmitted to agent 2 with an alternation between $-B/2$ and $+B/2$. Then, agent 2 selects all the possible associations with a word close to the integer received (at a distance less than B). If several associations are possible, agent 2 chooses the one with the highest score: (c_2, s_2). If $s_1 = s_2$ the interaction is a *success*, in the other cases the interaction is a *failure*. If no association is close enough in agent 2's memory, the agent creates a new association between the received integer and the meaning s_1. In case of success, agent 2 increases the score of the association (c_2, s_2) with $+ \delta$ and diminishes the score of competing associations $((c_2, *)$ and $(*, s_2))$ where $*$ is any meaning or word in the memory of the agent) with $-\delta$. In case of failure, agent 2 decreases the score of (c_2, s_2) with $-\delta$ (see Kaplan 2001; De jong and Steels 2003; Vogt 2005 for discussions of the importance of such forms of lateral inhibition). Associations are initially created with a 0 score.

We have previously shown that collective implicit evaluation led to choose the "best" word/replicator associated to a given meaning. What are the best words/replicators in the current model? A good word/replicator is a replicator that an agent will not confuse with another one that has a different usage. A "good" lexical system should have sets of words clearly distinct
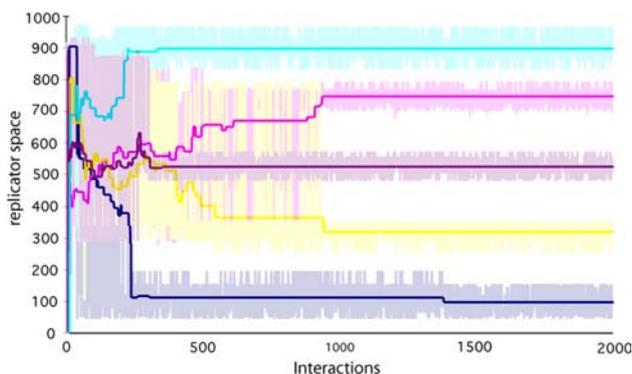

Fig. 7 Evolution in the word space (idealized acoustic space) of the words associated with five meanings in an experiment involving ten agents with a noise level $B = 100$. After a first period of ambiguity, five well separated bands appear associated with each meaning. Evolution of the 'average' values (in the population) of the words associated with each meaning is *highlighted* in the *middle* of each band

from one another depending on the meanings they associated to. Figure 7 shows the evolution in the word space (i.e. an idealized acoustic space) of the words associated with five meanings in an experiment involving ten agents with a noise level $B = 100$. After an initial ambiguity period, five well separated bands in the word space are clearly identifiable. Agents do not converge towards a unique value for each context. Each agent uses a different one. But these values tend to be very close. The "band" for one context is clearly distinct from bands associated with other contexts. No confusion is possible. Figure 7 plots also the 'average' value of each band. Thus, it is easier to see the collective optimization of distinctivity leading to a solution compatible with the level of noise present in the environment. We also see on this graph that with this level of noise we approach the limit of expressiveness possible in this medium. If the agents had to communicate about a larger number of distinct meanings, ambiguity will inevitably arise.

Compromise between distinctivity and robustness

Words are certainly not well modelled as integers. Some words are long and difficult to transmit, others are short but can be easily be confused with one another. To study this issue, let us now consider a model where each word/replicator is now a numeric chain of variable length. Each character of the chain is a number between 1 and 9. Noise is modelled by a probability of alteration P_m equal for each character. When a character mutates, it is simply replaced by a random character between 1 and 9. As in the previous model, agent 2 can look up the chains that are "close" to the transmitted word in its memory. We can define a distance D_c between chains, similar to the traditional Hamming distance.[3] In the interaction agent 2 selects the chains which are at a distance less than the threshold D.

---

[3] Let c_1 and c_2 be two chains the length of c_1 being either smaller or equal to the length of c_2. Let $k_1(i)$ and $k_2(i)$ be the character in position i in each of the chain. We define D_f as being the sum of the distance between the character of both chains to which is added ten times their length difference, $l_2 - l_1$.

$$D_c(c_1, c_2) = \sum_i \|k_1(i) - k_2(i)\| + 10.(l_2 - l_1) \qquad (2)$$

For instance the chains 1-4-5-2 and 1-4-5-7-3 are at a distance $5 + 10 = 15$. The arbitrary details of this particular distance are not important for the dynamics characterized. They are provided here only to permit a reproduction of the experimental results described.

With such a model, too long or too short chains are naturally less adapted. Indeed, the longer a chain is, the more risk it has to be altered during transmission. In a previous model, we have seen that such conventions generally lose the competition. But on the other side, if the conventional system is only composed of very short words, a single mutation might very often lead to confusion. Short words are robust but easily confused, long words are easy to distinguish but difficult to transmit correctly. A *compromise between robustness and distinctivity* must be found (for a longer discussion of this aspect refer to Kaplan 2001). We have analysed the distribution of all the words used by the agents after 5,000 games (at this point, we have observed experimentally that the replicator system reaches a stable state). The results of the distribution of the chain length are shown in Fig. 8. The distribution has a ''peak`` around chains of length 3. As expected, words/replicators that are too long or too short are less present in the final lexical systems.

These types of compromise is commonly observed in natural languages. Language typically contain a small set of short words used very frequently (e.g. auxiliaries). On the contrary, long words tend to have very specific meanings, used rarely.

Neutral drift

External factors like language contact between populations are often cited as a major cause of language evolution. But it is also known that language can change spontaneously based on internal dynamics (Labov 1994). We have seen with the previous model that in a noisy environment, agents can converge on a stable system in which distinct bands in the word/acoustic space are associated with distinct contexts. As we see in Fig. 7, this repartition in separated bands
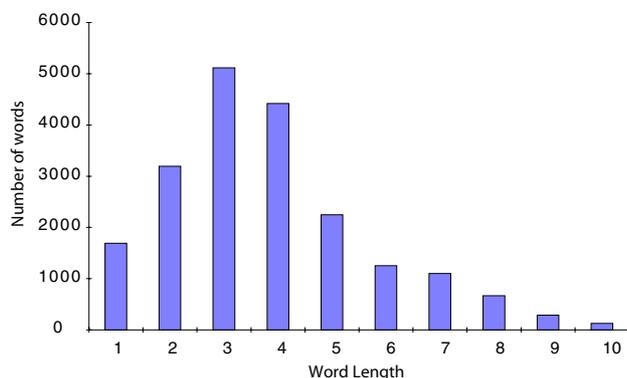

Fig. 8 Distribution of word length on 100 simulation runs for ten agents and 20 contexts with $D = 20$

does not evolve anymore once a stable solution has been found. Figure 9 shows the evolution of the average word in the presence of an agent flux defined by a probability of replacing an old agent by a new one $P_r = 0.01$, for a population of 20 agents and with two possible meanings. The centre of the bands are spontaneously evolving as new agents are entering the system. This is an example of a *neutral drift*.

This effect is easily understandable. A new agent tends to converge on words belonging to the existing bands for each meaning to express. But within this band, it does not converge towards the exact centre of the band. Thus the centre is moving as the flux of new agents enters the system. The higher the agent tolerance on noise, the higher the amplitude of this drift (see Steels and Kaplan 1998 for a first description of this phenomenon).

In this experiment, words/replicators are evolving spontaneously without any functional drive. However, external pressures can direct these dynamics in one direction or another. This neutral drift provides novelty and thus can lead to a more efficient reorganization if needed. In some way, this effect is similar to the role of neutral mutation in evolution (Kimura 1983).

Experiments on computational models of phonological systems have shown how similar collective dynamics in the presence of noise lead a population of agents to converge towards a set of vowels optimally distributed in the phonological space in order to favour distinctiveness between them (de Boer 1997; De Boer 1999; Oudeyer 2005b). Such emerging phonetic systems have high similarity with real ones as observed in natural languages.

To summarize, noise during word transmission favours sets of words that are clearly distinct from one another when they are associated with different meanings. When words can have different complexities, a compromise must be found between distinctivity and resilience to noise. Short words are easy to transmit but easily confused, long words are difficult to transmit correctly but are easily distinguishable from one another. We experimentally observe the convergence towards words of intermediary sizes. Finally, in the presence of noise and agent flux, we experimentally observe a spontaneous non functional evolution. This continuous exploration can lead to a more efficient reorganization of the replicator system if needed.

Linguistic replicators adapt to particular environments

Until now, we have only considered simple models of linguistic replication. Linguistic phenomena are
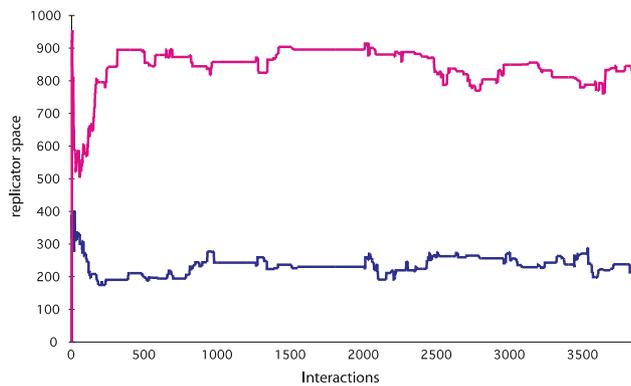
Fig. 9 Example of a neutral drift: spontaneous evolution of the 'average' forms in presence of an agent flux
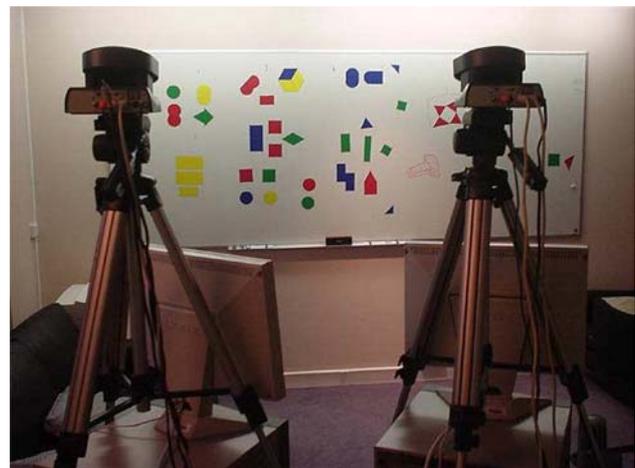

Fig. 10 The Talking Heads set up. Two robotic cameras are placed in front of white board. On the board, objects of various shaped and colours are placed. The robots have to construct categories and words to name the objects on the board and have the other agent guess the right object based on that word. Categories referring to colour, position, size or shape can be used

obviously more complex. The previous experiments have focused on the replication of words, but their associated meanings and semantic categories are also entities that can replicate from brain to brain through linguistic interactions. In many computational models, these categories are modelled as points in category space. But more complex systems of meanings, also referred as categories or concepts, were also investigated (Steels and Belpaeme 2005). The Talking Heads experiment conducted between 1999 and 2000 by Steels and co-workers have provided a large set of data on how systems of categories can adapt to particular environments (Steels and Kaplan 2002; Kaplan 2001). In this experiment robotic agents were capable of segmenting the image perceived through the camera into objects and of collecting various sensory data about each object, such as the colour, position, size, shape, etc. A couple of robots were placed in front of a white board on which various types of objects were placed. At each interaction, the speaker chose one object from this context, reused or constructed a category that would identify this object from the other object present in the background and uttered a word associated with that category. Based on this word, the other robot had to guess which object was named (Fig. 10).

In the first run of the experiment, a total of 8,000 words and 500 concepts were created, with a core vocabulary consisting of 10 basic words expressing concepts like up, down, left, right, green, red, large, small, etc. The dynamics that pushed the population towards coherence and distinctivity ensured the collective choice of a set of word-category associations adapted to the environment that the robots were perceiving. Interestingly, some features like shape were used very rarely to specify categories whereas position, colour and size categories were preferred. Interpreting such type of "preferences" is not always easy as they

can result both from the categorization mechanism used by the agents and the specific types of environments that they encounter. In the present case, to refer to specific objects in the types of scenes the robots perceived, position and colours were largely sufficient. Moreover, shapes were more difficult to distinguish than colour and position based on the features available to the agents, and given the kinds of objects present in their environment.

Such types of indirect competitions between perceptual categories were observed during the whole experiment. Some categories were general and other specific (e.g. one was used to describe a particular shade of green, and another one to describe green contexts in general). Usually, general categories were preferred because they were both easier to learn by the agent and adapted to a larger number of contexts (see also Smith 2005 for another series of experiments in this line). However, in several cases, a precise category adapted to reoccurring specific context survived as other categories were present to "back it up". Therefore, when analyzing these types of complex dynamics, considering competition between isolated categories is not always sufficient. The quality of a category needs to be evaluated regarding the category set to which it belongs and the adaptivity of the whole to particular environments.

Another interesting phenomenon was observed. Most of the words of the core vocabulary were coherently interpreted as having distinct meanings. However, in some cases, two competing meanings

co-occurred for a long time. For instance the word 'bozopite' was associated concurrently with two types of categories: large area (large) and large width (wide). This co-occurrence was due to the fact that in the types of environments that the robotic agents encountered most objects that were large in area were also large in width. This is an example of residual polysemy.

This brings us to a remark. As collective dynamics select sets of replicators that are well adapted to the environment in which the agents are communicating, we might be tempted to say that the "quality" of the replicators increases. But, like for species in natural evolution, *optimization stops once adaptation is reached*. We have seen in the previous section that in the presence of noise, well separated bands of replicators were emerging. However, once a stable solution was found, this optimization of distinctivity stops. The same effect occurs in more complex architectures where residual polysemy is observed. In all these situations, there is no absolute optimization, only the search for stable adapted solutions.

Linguistic replicators adapt to the cognitive constraints of their hosts

When linguistic replicators leap from brain to brain, they in fact do so through perpetual cycles of production, perception and learning. Whatever these replicators are, they need to go through a representation in the brain of speakers and hearers at some point. The process of updating one's brain to incorporate some new information defines learning. Learning theory, and in particular machine learning theory (Mitchell and Weinmann 1997), has shown that all learning systems are characterized by a number of biases which mean that every single system will be good at learning certain things and bad at learning other things. For example, learning algorithms such as recurrent neural networks are good to learn to predict complex time series but they are quite inefficient to learn fine categorical distinctions in high-dimensional static spaces, whereas support vector machines are good in high-dimensional static spaces but pretty bad when they have to learn time-dependent phenomena (Duda et al. 2001). Learning biases also apply to human brains. For example, when the human brain learns a new concept or a new sound, it will do so typically by using the representation of an already known concept or sound and modify it a little bit. The consequence is that learning a new concept or a new sound will only be effective if the corresponding brain already knows not too dissimilar concepts or sounds. This imposes strong constraints on the replication of linguistic memes, which are not only defined by the generic cognitive constraints of all human brains, but also by the particular cognitive structures that were built during the ontogeny of each of them. This means that for a given brain, some linguistic memes will be easily learnt and replicated, but some other linguistic memes will be strongly deformed often to the point that no replication at all takes place. And the linguistic memes which are easy to learn and replicate for this brain may prove to be difficult for another brain which had a different history.

What is then the consequence of all this on the dynamics of language evolution? We will now present the outline of a computational model of the origins of syllable systems which provides an answer (this model is described in detail in Oudeyer 2005a). This model involves a population of agents which can produce, hear, and learn syllables, based on an auditory and a motor apparatus that are linked by abstract neural structures. These abstract neural structures are implemented as a set of prototypes or templates, each of them being an association between a motor program that has been tried through babbling and the corresponding acoustic trajectory. Thus, agents store in their memory only acoustic trajectories that they have already managed to produce themselves. The set of these prototypes is initially empty for all agents, and grows progressively through babbling. The babblings of each agent can be heard by nearby agents, and this influences their own babbling. Indeed, when an agent hears an acoustic trajectory, this activates the closest prototype in its memory and triggers some specific motor exploration of small variation of the associated motor program. This means that if an agent hears a syllable S that it does not already know, two cases are possible: (1) he already knows a quite similar syllable and has a great chance to stumble upon the motor program for S when exploring small variations of the known syllable, (2) he does not already know a similar syllable and so there is little chance that he incorporates in its memory a prototype corresponding to S. This process means that if several babbling agents are put together, some islands of prototypes, i.e. networks of very similar syllables, will form in their memory and they will develop a shared skill corresponding to the perception and production of the syllables in these networks. Nevertheless, the space of possible syllables was large in these experiments, and so the first thing that was studied was whether agents in the same simulation could develop a large and shared repertoire of syllables. This was shown to be the case (Oudeyer 2005a). Interestingly, if one runs two simulations, the population of agents will always end up with their own particular repertoire of syllables.

Then, a second experiment was run: some fresh agents were tested for learning syllable systems that were formed by another population of interacting agents, and some other fresh agents were tested for learning a syllable system which was generated artificially as a list of random syllables. The results, illustrated in Fig. 11, were that the fresh agents were always good at learning the syllable systems developed by other similar agents, but on the contrary rather bad at learning the random syllable systems. In other terms, the syllable systems developed culturally by agents were adapted to their cognitive biases, and the random systems were not. Thus, the replicators constituted by syllables evolved and were selected in a cultural Darwinian process so as to fit to the ecological niche defined by the cognitive structures of agents, fitness being here learnability.

Several other computational systems have been developed to study the mechanisms that allow the cultural selection for learnability of linguistic replicators. Zuidema (2003) presented abstract simulations of the formation of syntactic structures and detailed the influence of cognitive constraints upon the generated syntax. Brighton et al. (2005) presented a thorough study of several simulations of the origins of syntax (Kirby 2001) which were re-described in the light of this paradigm of cultural selection for learnability.
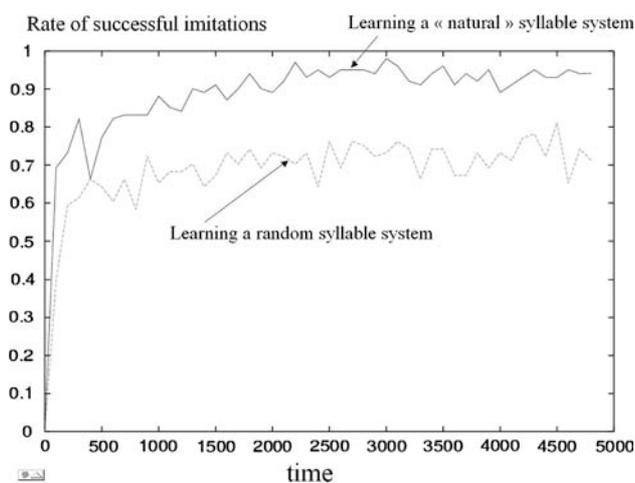

Fig. 11 Evolution of the rate of successful imitations for a child agent which learns a syllable system established by a population of agents (*top curve*), and for a child agent which learns a syllable system established randomly by the experimenter (*bottom curve*). The child agent can only perfectly learn the vocalization systems which evolved in a population of agents. Such vocalization systems were selected for learnability (reprinted from Oudeyer 2005a)
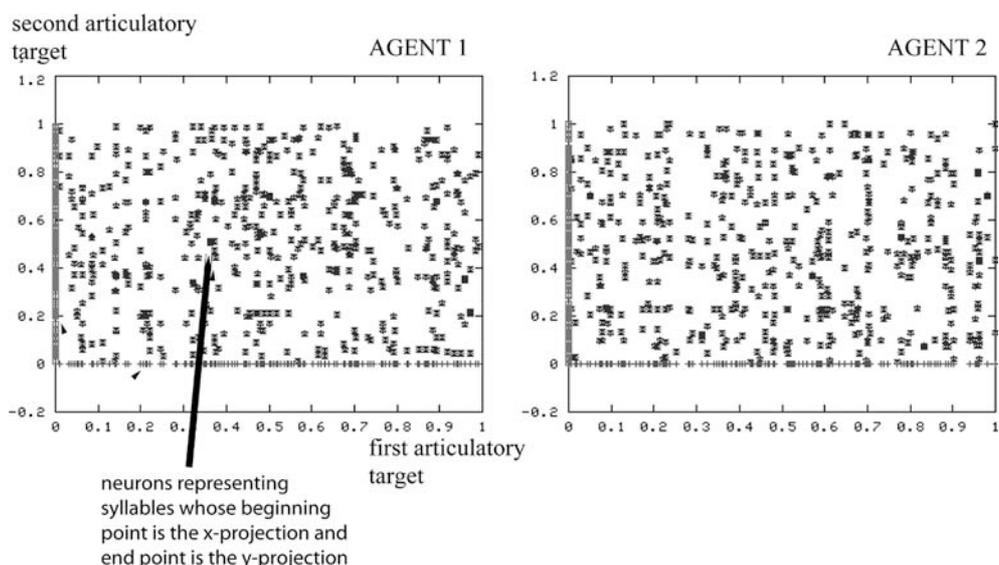
The self-organization of higher-level replicators: from individuals to groups

In the last section, we saw how it was possible that the ability of a linguistic replicator to replicate from one brain to another depended on the presence of other replicators in the receiving brain. As a consequence, this allows in principle for the formation of groups or coalition of mutually reinforcing replicators. In this section, we will illustrate with a related computer simulation how these groups can themselves become units subject to competition and selection.

This simulation is based on the same general principles as the simulation presented in the last section but the neural substrate was implemented in a lower-level and more realistic manner (Oudeyer 2006). In particular, prototypes were represented directly in terms of bundles of artificial neurons, and one of the crucial characteristics of these neurons was that their survival depended on their activation: neurons often activated were kept alive and neurons with low activation were pruned. Another difference was that initially there was a phase of generation of a large number of random neurons which then were progressively modified and recruited to represent particular acoustic or motor patterns. This implied that a given syllable could possibly recruit many redundant bundles of neurons, and another syllable could be encoded only by very few neurons. This introduces two interacting sources of competition for syllables: (1) competition for activation, (2) competition for neural resources. Indeed, the activity of agents was here merely babbling, which amounted to the random activation of neurons. The consequence was that the more a syllable got neural resources, the more it was produced, and so the more the neural resources were activated, both within the speaker's network and in the hearer's network. This activation was of course at the expense of other syllable's neural resources activation, and thus contributed to their progressive death. This introduced a positive feedback loop: the more a syllable was produced, the higher the probability to produce it again. Now, an interesting point was that the production of a syllable did not only activate the corresponding neural structure in agents, but also the neural structures of rather similar syllables, i.e. syllables which shared parts of the acoustic or motor trajectory. This made the formation possible of mutually re-inforcing groups of syllables.

Now, what happened precisely? Some systematic experiments were conducted in abstract and simple acoustic/motor spaces, so that they could be depicted as in Fig. 12. Here, we represent only the articulatory space, which is one-dimensional, and syllables are

Fig. 12 The neural maps of two agents at the beginning of the simulation. The neural map of one agent is represented on the *left*, and the neural map of the other agent is represented on the *right*

second articulatory target

neurons representing syllables whose beginning point is the x-projection and end point is the y-projection

defined by a beginning and an end point in this space. The beginning points are represented on the *x*-axis, and the ending points are represented on the *y*-axis. The little crosses in the middle represent neural bundles corresponding to particular syllables. Figure 12 shows the set of these neurons in two agents at the beginning of a simulation: they are randomly spread across the space, meaning that the syllables that agents produce cover uniformly the space of possible syllables in that space. Figure 13 shows what happens if one lets agents interact a few thousands times. We observe that lines and columns have appeared, which means that agents do not produce all possible syllables anymore but only a small subset, which is the same in both agents, and which is structured in a very particular manner. Indeed, now agents produce syllables composed of systematically re-used beginning or end points. In other terms, this corresponds to the transition from holistic to phonemically coded syllable

systems. The lines and columns show coalitions of syllables either sharing their beginning point or their end point, which have entered into competition with single syllables or other groups that died meanwhile. The state shown on Fig. 13 is in fact a convergent state in which the population of lines and columns has reached a threshold such that there is enough activation for all neurons (and thus their corresponding syllables) to survive.

A last point is that in these simulations, not all self-organized phonemes get re-used in the same manner: for example, on Fig. 13, if we call the self-organized phonemes

$$p_1, p_2, \ldots, p_8$$

then we can summarize the repertoire of allowed sequences by:

$$(p_6, *), (p_8, *), (*, p_7)$$

Fig. 13 The neural maps of the same two agents after 1,000 interactions. We observe that many neurons have died and the surviving ones are organised into *lines and columns*: this means that phonotactic rules have appeared, that the repertoire of vocalisation can be organised into patterns, and that some phonemes get re-used systematically for building vocalisations, i.e. vocalisations are now combinatorial
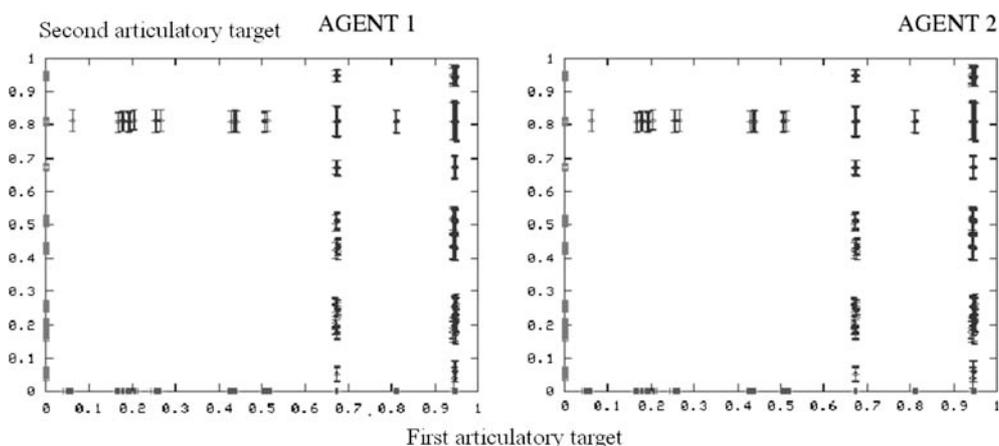
where * means "any phoneme in $p_1,...,p_8$". This corresponds to the formation of rules of sound syntax regulating the possible phonemic combinations, and is called the phonotactics of a syllable system in phonology: this is indeed a feature present in all human syllable systems. What is interesting here is that this is a self-organized side-effect of the process of cooperation/competition between groups of syllables: there was no functional pressure in the system that forced it to appear. Nevertheless, it has been widely described in the linguistics literature how phonotactics are used as constraints that help speakers in multiple ways: for example, it is useful for better speech recognition and for morpho-syntactic analysis. So one can envision easily an evolutionary scenario of language in which phonotactics appeared first as a side effect of a more fundamental process, i.e. the formation of a shared set of speech sounds, and was recruited later to constrain various linguistic processing functions: this corresponds to the well-known concept of exaptation in biology (Gould and Vrba 1982).

Conclusion

In this paper, we have shown how computational experiments could allow us to elaborate and refine the scientific understanding of language evolution as a Darwinian process. In particular, we have seen that there is potentially a wide variety of linguistic replicators, different both in nature and in replication mechanisms, and interacting at multiple levels. This shows that the most relevant analogy between language and biological evolution is probably neither at the level of units nor at the level of replication mechanisms, but rather at the general level of systems defined as unspecified sets of units subject to differential replication with inheritance and variation. Using computational experiments allows us to validate the coherence of this view of language evolution as a Darwinian process, which in turn participates in the refoundation of linguistics (Croft 2000). Indeed, in the last 50 years language has been mainly considered as a fixed and idealized system which could be studied independently of its use and of its users. In this traditional body of theories, individual variation, and more broadly language evolution were either ignored or left unsolved, and biological evolution was used to explain the particularly good adaptation of our brains to the learning of nowadays idiosyncratic languages. On the contrary, viewing language as a system of replicators constantly replicating from particular brains to particular brains, with variation as a central concept, allows one to understand how languages change over time, why there is so much linguistic diversity, and provides a different account of the ease with which children learn languages. Indeed, as we showed in "Linguistic replicators adapt to the cognitive constraints of their hosts", this framework allows to understand that languages themselves probably evolved in a cultural Darwinian manner so as to become easily learnable by their users. And the peculiarities of the pre-existing learning systems of these users can explain the apparent idiosyncratic properties of languages. The paradigm shift induced by viewing language evolution as a Darwinian process also sets up new problems to be solved. In particular, it highlights the fact that the sharing of complex and intricate linguistic conventions must be explained: how can a system of competing replicators interacting at the level of individuals converge to a coherent and distinctive system adopted by all the population? We have shown with several computational experiments how this problem could be solved for lexical systems, thanks to the use of specific replication mechanisms based on positive feedback loops and self-organization. Yet, future work will have to show how several interacting levels of conventions, ranging from phonology to grammar and pragmatics, can be formed through a cultural Darwinian process.

References

Ancel LW, Fontana W (2000) Plasticity, evolvability, and modularity in RNA. J Exp Zool 288:242–283

Arita T, Koyama Y (1998) Evolution of linguistic diversity in a simple communication system. In: Adami C, Belew R, Kitano H, Taylor C (eds) Proceedings of Artificial Life VI. The MIT Press, Cambridge, MA, pp 9–17

Baronchelli A, Felici M, Loreto V, Caglioti E, Steels L (2006) Sharp transition towards shared vocabularies in multi-agent systems. J Stat Mech P06014

Berlinski D (1972) Philosophical aspects of molecular biology. J Philos 69(12):319–335

Blevins J (2004) Evolutionary phonology: the emergence of sound patterns. Cambridge University Press, Cambridge

Blevins J (2006) New perspectives on english sound patterns: natural and unnatural in evolutionary phonology. J Engl Linguist 34:6–25

de Boer B (1997) Generating vowel systems in a population of agents In: Husbands P, Harvey I (eds) Proceedings of the Fourth European Conference on Artificial Life. The MIT Press, Cambridge, MA

Brighton H, Kirby S, Smith K (2005) Cultural selection for learnability: three hypotheses underlying the view that language adapts to be learnable. In: Tallerman M (ed)

Language origins: perspective on evolution. Oxford University Press, Oxford

Cangelosi A, Parisi D, (2002) Simulating the evolution of language. Springer, Heidelberg

Cangelosi A, Smith AM, Smith K (eds) (2006) The evolution of language. In: Proceedings of the 6th International Conference on the Evolution of Language. World Scientific, Singapore

Croft W (2000) Explaining language change. Linguistics. Longman, London

Croft W (2002) The darwinization of linguistics. Selection 3:75–91

Crow J, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York

Darwin C (1859) On the origin of species. John Murray, London

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

De Boer B (1999) Self-organizing phonological systems. Ph.D. thesis, VUB University, Brussels

De Jong E, Steels L (2003) A distributed learning algorithm for communication development. Complex Syst 14(4–5):315–334

Dorigo M, Maniezzo V, Colorni A (1997) The ant system: optimisation by a colony of cooperating agents. Trans Syst Man Cybern Part B 26(1):29–41

van Driem G (2003) The language organism: the leiden theory of language evolution. In: Mírovsky J, Kotesovcova A, Hajicova E (eds) Proceedings of the XVIIth International Congress of Linguists Prague: Matfyzpress vydavatelství Matematicko-fyzikální fakulty Univerzity Karlovy, Prague

Duda R, Hart P, Stork D (2001) Pattern classification. Wiley, New York

Gould SJ, Vrba E (1982) Exaptation—a missing term in the science of form. Paleobiology 8:4–15

Hull D (1988) Science as a process: an evolutionary account of the social and conceptual development of science. University of Chicago Press, Chicago

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

Kaplan F (2001) La naissance d'une langue chez les robots. Hermes Science, Paris

Kaplan F (2005) Simple models of distributed co-ordination. Connect Sci 17(3–4):249–270

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kirby S (2001) Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Trans Evol Comput 5(2):102–110

Kirby S (2002) Natural language and artificial life. Artif life 8:185–215

Labov W (1994) Principles of linguistic change, vol 1. Internal factors. Blackwell, Oxford

Maynard-Smith J (1982) Evolution and the theory of games. Cambridge University Press, Cambridge

Mitchell B, Weinmann L (1997) Creative design for the WWW. Lecture Notes, SIGGRAPH 1997

Mufwene SS (2005) Language evolution: the population genetics way, vol 29. Gene, Sprachen, und ihre Evolution, pp 30–52

Oliphant M (1997) Formal approaches to innate and learned communicaton: laying the foundation for language. Ph.D. thesis, University of California, San Diego

Oudeyer P-Y (2005a) How phonological structures can be culturally selected for learnability. Adapt Behav 13(4):269–280

Oudeyer P-Y (2005b) The self-organization of speech sounds. J Theor Biol 233(3):435–449

Oudeyer P-Y (2006) Self-organization in the evolution of speech. Oxford University Press, Oxford

Schleicher A (1863) Die darwinsche Theorie und die Sprachwissenschaft: Offenes Sendschreiben an Herrn Dr. Ernst Häckel. Böhlau, Weimar

Searls DB (2002) The language of genes. Nature 420:211–217

Smith ADM (2005) The inferential transmission of language. Adapt Behav 13(4):311–324

Sperber D (1984) Anthropology and psychology: towards and epidemiology of representations (the malinowsju memorial lecture 1984). Man 20:73–89

Steels L (1996) A self-organizing spatial vocabulary. Artif Life J 2(3):319–332

Steels L (2003) Evolving grounded communication for robots. Trends Cogn Sci 7(7):308–312

Steels L (2004) Analogies between genome and language evolution. In: Pollck J (ed) Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems. The MIT Press, Cambridge, MA, pp 200–207

Steels L, Belpaeme T (2005) Coordinating perceptually grounded categories through language: a case study for colour. Behav Brain Sci 28:469–529

Steels L, Kaplan F (2002) Bootstrapping grounded word semantics. In: Briscoe T (ed) Linguistic evolution through language acquisition: formal and computational models. Cambridge University Press, Cambridge, pp 53–74

Steels L, Kaplan F (1998) Spontaneous lexicon change. In: Proceedings of COLING-ACL 1998. Morgan Kaufmann, San Francisco, CA, pp 1243–1250

Stegmann UE (2004) The arbitrariness of the genetic code. Biol Philos 19(2):205–222

Tsonis AA, Elsner JB, Tsonis PA (1997) Is DNA a language? J Theor Biol 184:25–29

Vogt P (2005) The emergence of compositional structure in perceptually grounded language games. Artif Intell 167(1–2):206–242

Watson J, Crick F (1953) Molecular structure of nucleic acids. Nature 171:737–738

Zuidema W (2003) How the poverty of the stimulus solves the poverty of the stimulus. In: Becker S, Obermayer K (eds) Advances in Neural Information Processing 15. MIT Press, Cambridge, pp 51–68