Chapter 9
Self-Organization: Complex Dynamical Systems in the Evolution of Speech

Pierre-Yves Oudeyer

Abstract Human vocalization systems are characterized by complex structural properties. They are combinatorial, based on the systematic reuse of phonemes, and the set of repertoires in human languages is characterized by both strong statistical regularities—universals—and a great diversity. Besides, they are conventional codes culturally shared in each community of speakers. What are the origins of the forms of speech? What are the mechanisms that permitted their evolution in the course of phylogenesis and cultural evolution? How can a shared speech code be formed in a community of individuals? This chapter focuses on the way the concept of self-organization, and its interaction with natural selection, can throw light on these three questions. In particular, a computational model is presented which shows that a basic neural equipment for adaptive holistic vocal imitation, coupling directly motor and perceptual representations in the brain, can generate spontaneously shared combinatorial systems of vocalizations in a society of babbling individuals. Furthermore, we show how morphological and physiological innate constraints can interact with these self-organized mechanisms to account for both the formation of statistical regularities and diversity in vocalization systems.

Keywords Self-organization · Natural selection · Evolution of speech · Combinatoriality · Computational model

Human vocalization systems are characterized by complex structural properties. They are combinatorial, based on the systematic reuse of phonemes, and the set of repertoires in human languages is characterized by both strong statistical regularities—universals—and a great diversity. Besides, they are conventional codes culturally shared in each community of speakers. What are the origins of the forms

P.-Y. Oudeyer (✉)
INRIA Bordeaux Sud-Ouest, 13, C1 route du Brassan,
33640 Ayguemorte-les-Graves, France
http://www.pyoudeyer.com
e-mail: pierre-yves.oudeyer@inria.fr

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_9,

of speech? What are the mechanisms that permitted their evolution in the course of phylogenesis and cultural evolution? How can a shared speech code be formed in a community of individuals? This chapter focuses on the way the concept of self-organization, and its interaction with natural selection, can throw light on these three questions.

The tendency of many complex physical systems to generate spontaneously new and organized forms, such as ice crystals or galactic spirals, is indeed present as much in the inorganic world as in the living world. Thus, the explanation of the origins of forms and structures in the living can not only rely on the principle of natural selection, but should be complemented by the understanding of physical mechanisms of form generation in which self-organization plays a central role. This applies to the social and cultural forms of the living, in particular to the forms of speech and language. As a consequence, I will begin by articulating in a general manner the relationships between self-organization, natural selection and Neo-Darwinism in explanations of the genesis of forms in the living. Then, I will instantiate these relations in the context of the three questions formulated above. After that, I will explain why the use of computer models and simulations is fundamental for progress in our understanding of these issues. Finally, I will present the example of an experiment based on a computer model which shows that certain simple mechanisms coupling perception and production of sounds can generate combinatorial systems of vocalizations, characterized by the universal/diversity duality, and shared culturally by the members of a speech community. I will conclude with the presentation of evolutionary scenarios that this computer experiment complements or renews.

9.1 Self-Organization and the Evolution of Forms in the Living

9.1.1 Physics, the Caldron of Self-Organized Forms

Nature, especially inorganic nature, is full of fascinatingly organized forms and patterns. The silhouettes of mountains are the same, whether one views them at the scale of a rock, a summit, or a whole mountain range. Sand dunes often arrange themselves in long parallel stripes. Water crystallizes into symmetrical serrated flakes when the temperature is right. And when water flows in rivers and hurtles over cataracts, trumpet-shaped vortices appear and the bubbles collect together in structures which are sometimes polyhedral. Lightning flashes draw plant-like branches in the sky. Alternating freezing and thawing of the rocky ground of the tundra leaves polygonal impressions in the earth. The list of these forms rivals many human artefacts in complexity, as can be seen in Fig. 9.1. And yet they are not designed or conceived by anyone or anything, not even natural selection, Dawkins' 'blind watchmaker' (Dawkins 1986). What, then, are the mysterious factors that explain their existence?

In fact, all these organized structures have a feature in common: they are the macroscopic outcomes of local interactions between the many components of the system

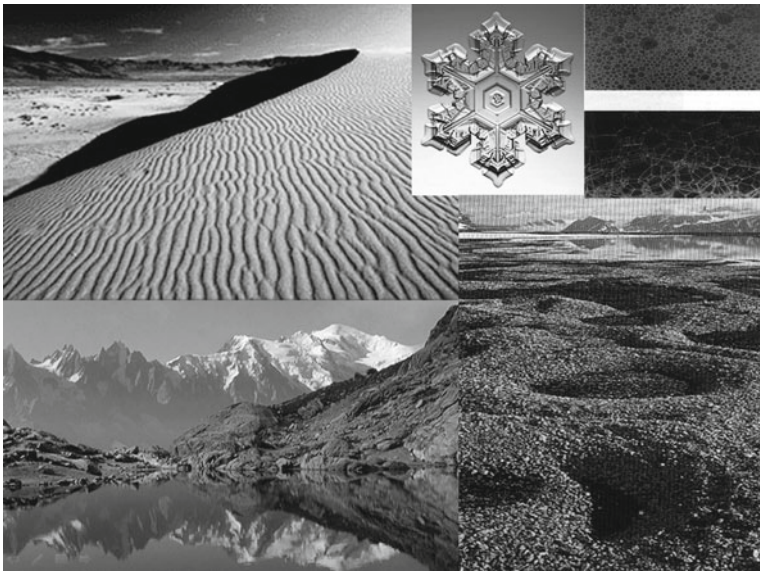

Fig. 9.1 Nature is full of organized forms and patterns without there being anywhere any plans which might have served to build them; they are said to be self-organized. Here, parallel stripes running through sand dunes, water bubbles on the surface of liquid which has been stirred up and the polyhedral structures which are left when they dry out, an ice crystal, mountains whose shapes are the same whether one views them on the scale of a rock or a whole peak (Photos: Nick Lancaster, Desert Research Institute, Nevada)

from which they emerge. Their global organizational properties are not to be found at the local level. Indeed the properties of the shape of a water molecule, as well as its individual physico-chemical properties, are qualitatively different from the properties of ice crystals, whirlpools, or polyhedral bubbles. The polygonal impressions in the tundra do not correspond to the shape of the stones composing them, and have a spatial organization quite different from the temporal organization of freezing and thawing. This is the hallmark of a newly discovered phenomenon—self-organization.

In nature self-organization characterizes very diverse physical systems, but several typical properties can be identified: non-linearity, symmetry breaking, presence of dynamical systems, "attractors", and historicity. For example, when one heats from below a thin layer of oil spread out on a flat surface, convection currents with peculiar geometric shapes (lines or polygons) self-organize and these shapes change dramatically when the temperature goes over given thresholds (see Fig. 9.2). On the contrary, between these thresholds the shapes remain globally stable even if they are perturbed, constituting attractors. Another property of many self-organized dynamical systems is historicity, often associated with the sensitivity to initial conditions in chaotic systems: the attractor in which the systems falls, i.e. the shapes/forms that are produced by the complex system, can be very different depending on slight variations in the initial conditions. For example, this is the case of ferromagnetization: each of the

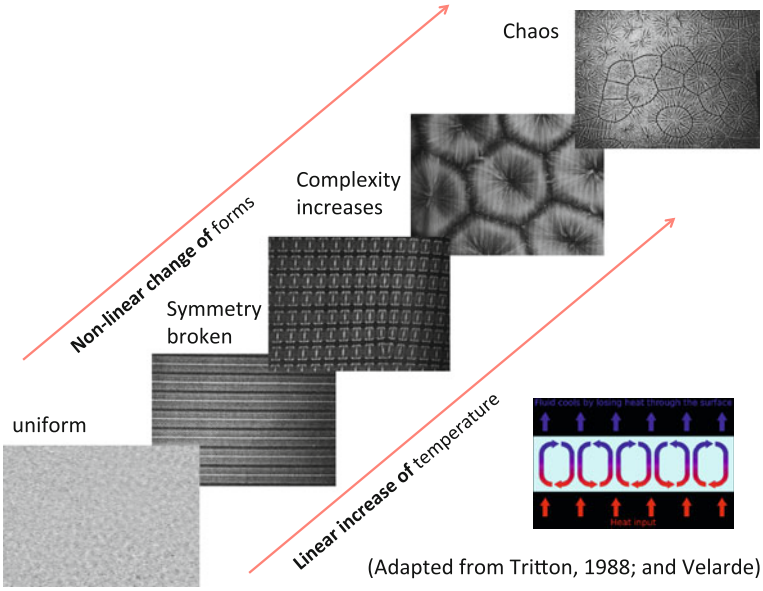

Fig. 9.2 Rayleigh-Bénard cells: when one heats from below a thin layer of oil spread out on a plane surface, convection currents with peculiar geometric shapes (lines or polygons) self-organize and these shapes change brutally when the temperature goes over given thresholds. This kind of non-linearity characterized many self-organized systems both in the inorganic and in the living world (Photos adapted from Tritton (1988), and Manuel Velarde, Universidad Complutense, Madrid)

atoms of an iron plate can be viewed as a sort of magnet that can have several possible orientations, and whose orientation is random if the temperature is high enough. Yet, if the temperature goes below a certain threshold, a self-organized phenomenon happens: all atoms spontaneously adopt the same magnetic orientation. This shared orientation is quasi-impossible to predict beforehand and tiny random variations of initial orientations of atoms can lead the plate to be magnetized in a quite different orientation. These variations in initial conditions are typically linked to contingent events that interacted with the iron plate: this is why the final state of the plate depends both on its history and on its intrinsic physical mechanisms, whence the term "historicity".

This fundamental concept is the touchstone of the paradigm shift undergone by the sciences of complexity in the twentieth century (Ashby 1956; Nicolis and Prigogine 1977; Kauffman 1996; Ball 2001). Ever since Newton, good science was supposed to be reductionist, and consisted in decomposing natural systems into simpler subsystems. For example, to understand the functioning of the human body, it was appropriate to study the respective parts, such as the heart, the nervous system, or the limbic system. Moreover, things did not stop there, and the study of the nervous system, for example, was subdivided into study of the cortex, of the thalamus, or of the peripheral motor innervations, and each of these sub-parts was studied by hyper-specialists in

separated dedicated university departments. This method has obviously enabled us to accumulate an impressive bank of knowledge. But the prophets of complexity have broken up this paradigm. Their credo is "the sum of the parts is greater than the parts taken independently".

9.1.2 The Impact of Self-Organization on the Origins of Forms in the Living

Complex systems, i.e. systems composed of many interacting sub-systems, abound in nature and have a strong tendency to self-organize. The examples of the previous section were chosen deliberately from inorganic systems to show that the property of self-organization can be found in systems subject to laws which have nothing to do with natural selection. However, self-organization applies similarly to living systems. It is a concept widely used in several branches of biology. It is particularly central to theories which explain the capacity of insect societies to build nests or hives, to hunt in groups or to explore in a decentralized and effective way the food resources of their environment (Camazine et al. 2002). In developmental biology, it is used, for example, to explain the formation of coloured patterns on the skins of animals like butterflies, zebras, jaguars or ladybirds (Ball 2001).

Thus, it seems possible that there are shape- and pattern-forming mechanisms in biological systems which are orthogonal to natural selection, through their property of self-organization. Now natural selection is at the heart of almost all the arguments of biologists when it comes to explaining the presence of a shape, a pattern or a structure in an organism. What, then, is the relationship between the theory of natural selection and self-organization?

Some researchers have suggested that self-organization casts doubt on the centrality of natural selection in explaining the evolution of living organisms. Waldrop explains:

> Complex dynamical systems can sometimes go spontaneously from randomness to order; is this a driving force in evolution? Have we missed something about evolution—some key principle that has shaped the development of life in ways quite different from natural selection, genetic drift, and all the other mechanisms biologists have evoked over the years? . . . Yes! And the missing element. . . is spontaneous self-organization: the tendency of complex dynamical systems to fall into an ordered state without any selection pressure whatsoever (Waldrop 1990).

However, this is not the position I take in this article. Rather than seeing self-organization as a concept which minimizes the role of natural selection by suggesting competing form-creating mechanisms, it is more accurate to see instead as belonging to a somewhat different level of explanation and more importantly as describing mechanisms which actually increase the power of natural selection by

orders of magnitude. Systems with the self-organizing property are completely compatible with the mechanism of natural selection in explaining the evolution of forms and structures in biology.

9.1.3 Classic Neo-Darwinism

To see the matter clearly, it is first necessary to recall what the mechanism of natural selection, or Neo-Darwinism, comprises. It is a mechanism characterizing a system composed of individuals each having particular traits, shapes or structures. In addition, the individuals in this system are capable of replication. This replication must occasionally produce individuals which are not exact copies of their ancestors, but are slight variants. These variations are the source of diversity among individuals. Finally, each individual has a greater or lesser capacity for replication, according to its surrounding environment. This generates differential replication of individuals and gives rise to "selection" of those who are most capable of replicating themselves. The combination of the processes of replication with variation and selection means that, over the generations, the structures or traits of individuals which help them to reproduce themselves are preserved and improved upon.

Now there is one crucial point on which the theory of natural selection is neutral: it is the way in which variation is generated, and more generally the ways in which the individuals with their shapes, traits and structures are produced. A number of Neo-Darwinist arguments consider the mechanisms of variation of forms as secondary in comparison with the reproductive advantages of these forms when it comes to explain their evolution. This states implicitly that the relation between the level of genes, considered as the main space in which variations operate through mutations and cross-overs, and the level of phenotypes, considered as an isomorphic image of the space of genes, is simple and linear. According to this vision, exploration of the space of phenotypes (which determines, along with the environment, the relative effectiveness of the genes at replicating) can simply be carried out by studying the way things change in the space of genotypes. Now the mechanisms of mutation which actually bring about these changes are of small amplitude (most mutations only affect a minimal fraction of the genome when replication succeeds), and thus random variations in genes lead to uniform exploration of the space of genotypes. What this means is that under the hypothesis that phenotypic and genotypic space have the same structure and can be mapped approximately linearly, the space of possible biological forms can be searched quasi-continuously, by successive little modifications of pre-existing forms. Fortunately for the appearance of complex life-forms, this is not the case. In fact, although this mechanism of small successive variations in form is notably effective in the delicate regulation of the structures of organisms, it would make the search for forms as complex as those of human organisms equivalent to the search for a needle in a haystack because genomes are much too high-dimensional (Keefe and Szostak 2001).

9.1.4 Self-Organization Constrains the Space of Forms to be Explored: Not All Forms can Emerge Equally Easily

It is here that the concept of self-organization comes to the rescue of this naive search mechanism in the space of phenotypic forms in the Neo-Darwinian theoretical framework. In fact the relation between genes and the forms of organisms is characterized by its complexity and its non-linearity that are expressed through the ontogenetic and epigenetic development of each organism. Organisms are constructed starting from a stem cell containing a whole genome. This stem cell can be seen as a dynamic system parameterized by its genome and under the influence of perturbations imposed by the environment. This dynamic system is above all a self-organizing system with the same sorts of properties as the self-organizing systems described in the previous section. The genome is a set of parameters analogous to temperature and the viscosity of liquids in Bénard systems, and the environment is analogous to noise (but evidently highly structured noise!). Thus the development of an organism from a stem cell shares many properties with physical systems: shapes, structures and patterns appear at the global level, and are qualitatively different from those implementing functioning at the local level, that is, different from the patterns characterizing the structure of the stem cell and its genome. The hexagonal pattern which can appear as a result of a simple difference in temperature in a homogeneous liquid gives an idea of the way in which a simple sequence of nucleotides enclosed in a system of molecules which transforms them automatically into proteins can generate a bipedal organism endowed with two eyes and ears and an immensely complex brain.

As with Bénard systems or ferromagnetic plates, dynamic systems defined by the cells and their genomes are characterized by a landscape of attractors: there are large regions in the parameter space within which the dynamic system systematically adopts behaviour which is more or less the same. For Bénard systems, there is a range of temperatures giving rise to parallel stripes which is wide enough to locate easily. For ferromagnetic plates the range of temperature in which the system settles to global magnetic coherence is also very wide. Thus for living organisms it is not only possible to generate self-organizing structures with complex global properties, but in addition these structures are generated by genomes belonging to broad subspaces of genome space called basins of attraction. The structuring of genome space into basins of attraction by this kind of dynamic system facilitates the evolutionary search of the space of forms so that it does not resemble a search for a needle in a haystack.

As in ferromagnetic systems, structured noise imposed by the environment on the development of the dynamic system can lead it to take different developmental pathways. For pieces of iron at low temperatures, this corresponds to magnetization in one direction or another. For a living organism, this corresponds to its possible shapes; this is how it happens that even monozygotic twins can show quite important morphological differences. This is also the reason why the relationship between

genes and the forms of organisms is not only complex and nonlinear, but also non-deterministic. Moreover, and as in Bénard systems where search of the parameter space of temperature can sometimes lead to fast and qualitative changes in the behaviour of the system (for example the change from parallel stripes to square cells), which have been called phase-transitions, the search within genome space can also lead to fast qualitative changes. This possibly corresponds to many observations of rapid form-changes in evolution, as witnessed by the fossils studied by anthropologists, and which are the basis of the theory of punctuated equilibrium proposed by Elredge and Gould (1972).

To summarize, the self-organizing properties of the dynamic system composed by the cells and their DNA brings essential structuring to the phenotypic space by constraining it, making the discovery of complex robust forms by natural selection much easier. On the one hand, these properties enable a genome to generate complex, highly organized forms without the need for precise specification of each detail in the genome (in the same way as Bénard's polygonal shapes are not specified precisely, or encoded in a plan, in the properties of the liquid's molecules). On the other hand, the self-organizing properties structure the landscape of these possible forms into basins of attraction within which they resemble each other greatly (here is where gradual evolution happens, involving fine tuning of existing structures), and between which there can be substantial differences among forms (transitions from one basin to another are what provide abrupt and powerful innovations in evolution). To give a simple picture, self-organization provides a catalogue of complex forms distributed over a landscape of valleys in which and between which natural selection moves and makes its choices: self-organization proposes, and natural selection disposes. Obviously this is only an image to facilitate understanding, because with its movements natural selection actually enables new mechanisms, themselves self-organized, to appear, and these in turn structure the space of forms within which it moves; thus natural selection participates in the formation of these mechanisms which help it to move effectively in the space of forms; vice versa, the mechanism of natural selection certainly appeared in the history of life due to the self-organized behaviour of systems which were as yet completely unconnected to natural selection; natural selection and self-organizing mechanisms thus help each other reciprocally in a sort of spiral which enables complexity to increase during the course of evolution.

The consequence of this interplay between natural selection and self-organization is that any explanation of the origins and evolution of forms and structures in living organisms requires at least two kinds of argumentation. The first one, classic, is the neo-Darwinian functionalist argumentation: it consists in identifying the ecological context in which a given trait may have appeared and in articulating the balance between the associated costs and reproductive advantages. The second kind of argumentation is more rarely used but is equally essential: it consists in identifying the developmental/epigenetic mechanisms, and the associated constraints, which may have permitted, or made difficult, the genesis of these novel traits. And the concept of self-organization is central to the way developmental mechanisms impact the genesis of forms.

9.2 Self-Organization and the Evolution of Forms and Structures of Language and Languages

The question of how speech and language arose in humans, and the question of how new languages form and evolve, are among the most difficult that science has to address. After being put aside from scientific enquiries during most of the twentieth century, partly because of the ban pronounced by the Société Linguistique de Paris, they are now again the focus of attention of a whole scientific community. There is an emerging consensus among researchers who are today getting down to questions of the origin of the human language faculty and the evolution of languages: this research must be interdisciplinary. It in fact poses a puzzle with immense ramifications which go beyond the competence of each individual discipline on its own. Firstly, it is because the two big questions, that of the origins of language and that of the origins of languages, must be decomposed into subquestions which are themselves already quite complex: What, in fact, is the language faculty? What is a language? How are sounds, words, sentences and representations of meaning related to each other? How does the brain represent and process these sounds and sentences and the concepts which they convey? How do we learn to speak? What are the respective roles of nature and nurture? What is language for? What is its role in a community? How does a language form and change in the course of successive generations of speakers (Croft, this volume; Kirby, this volume)? What do we know about the history of each particular language? Why are languages and the language faculty the way they are? Why do we see universal tendencies and at the same time great diversity in languages? How does language influence the way we perceive and understand the world? What do we know about the history of the human capacity for speech? Is it mainly the result of genetic evolution, like the evolution of the eyes, or a cultural invention, like writing? Is language an adaptation to a changing environment? An internal change in an individual which increased its chances of reproduction? Is it an exaptation, a side effect of changes which were not at first tied to communicative behaviour? What are the evolutionary prerequisites which paved the way for the capacity of speech? And how did these prerequisites appear? Independently? Genetically? Culturally? (Dediu, this volume; Zuidema, this volume).

Placed against the diversity of these questions is an even greater diversity of research disciplines and methods. Linguists, even though they continue to provide critical data on the history of languages, are no longer the unique actors. Developmental and cognitive psychologists and neuropsychologists carry out behavioural studies of language acquisition and language pathology, and these often reveal cognitive mechanisms involved in language processing (Wonnacott, this volume). Neuroscience (Müller, this volume), especially with equipment for brain imaging allowing us to see which brain regions are active for given tasks, attempts to find neural correlates of verbal behaviour, to discover its organization in the brain. Some researchers also study the physiology of the vocal tract, to try to understand how we produce speech sounds. The physiology of the ear, the essential receptor in the speech-decoding chain (or vision, in the case of signed languages), is also a focus of research. Archeologists

examine fossils and artefacts left by the first hominids, and try on the one hand to deduce our anatomical evolution (especially of the larynx) and on the other hand to get an idea of what activities they were engaged in (What tools did they make? How did they use them? What can these tools tell us about the degree of cognitive development?). Anthropologists do fieldwork on isolated peoples, and report on cultural differences, especially those related to languages and the meanings they convey. Primatologists try to report on the communicative capacities of animals that may have some ancestors in common with human and to compare them with our own communicative capacities. Geneticists on the one hand sequence the human genome and that of potential ancestral species when it is possible to specify their phylogenetic relatedness, and on the other hand use genetic information from different people across the planet to help in reconstructing the history of languages, which is often correlated with the genetic history of their speakers (Dediu, this volume).

Thus language involves a multitude of components interacting in complex ways in parallel on several timescales: the ontogenetic timescale, characterizing the growth of an individual person, the glossogenetic or cultural timescale which characterizes the evolution of cultures, and the phylogenetic timescale, which characterizes the evolution of species (see Fig. 9.3). In particular, language is characterized by complex physical and functional interactions among multiple cerebral circuits, several organs, the individuals who are equipped with them, and the environment in which they live. Now, as we have seen in previous paragraphs, not only is it essential to study each

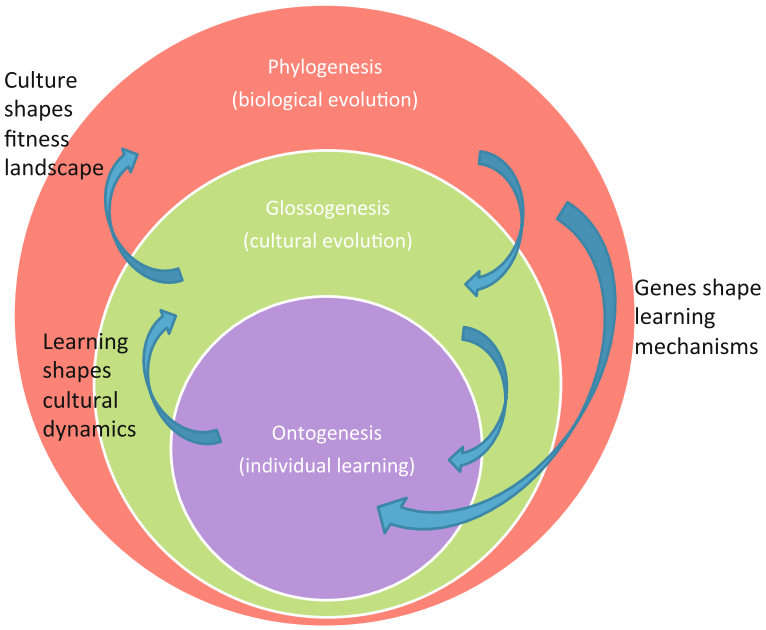

Fig. 9.3 Multiple interaction scales involved in the origins of language

of these components independently to reduce the complexity of the problem; it is also necessary to study their interactions. Thus, a growing number of researchers have proposed the idea that many properties of language and languages may not be encoded in any particular component involved, i.e. in certain specific cerebral structures or in properties of the auditory and vocal apparatus, or even in an individual considered independently of others, but rather may be the self-organized outcomes of the complex dynamic interactions among the components and individuals. Yet, these self-organizational phenomena are often complicated to understand or to foresee intuitively, and to formulate in words, whence the necessary use of mathematical and computer modelling that I will now present.

9.3 Computer Models and Simulations of the Evolution of Language

9.3.1 Experimenting with Complex Dynamical Systems

Nowadays, one of the most efficient ways to improve our understanding of the dynamics of self-organized systems is the use of computers or robots. Indeed, they allow us to implement operational models of which we know all the assumptions, to run them, and to observe the resulting behaviour as a function of the values of parameters set in the context of these models. This is why, in addition to linguists, psychologists, anthropologists, neuroscientists, geneticists, and physiologists, mathematicians and computer scientists/roboticists have now a critical role in this research.

An operational model is one which defines the set of its assumptions explicitly and above all shows how to calculate their consequences, that is to prove that it leads to a certain set of conclusions. There are two main types of operational model. The first, used by mathematicians and some theoretical biologists, consists in abstracting from the phenomenon of language a certain number of variables along with the rules of their evolution in the form of mathematical equations. Most often this resembles systems of coupled differential equations, and benefits from the framework of dynamic systems theory. The second type, which allows for modelling of more complex phenomena than the first, is that used by researchers in artificial intelligence: it consists in the construction of artificial systems implemented in computers or in robots. These artificial systems are made of programs which most often take the form of artificial software or robotic agents, endowed with artificial brains and bodies. These are then allowed to interact with an artificial environment (or a real environment in the case of robots), and their dynamics can be studied. This is what one calls the "method of the artificial" (Steels 2001) or the "synthetic methodology" (Pfeifer and Scheier 1999). The use of computational machines to simulate and study natural phenomena is not new: Early computers were used by Lorenz to study the behaviour of climatological models, by Fermi to simulate non-linear interactions among magnetized particles,

by Turing to imagine how morphogenetic processes could self-organize, and by Von Neumann to study self-replication.

More recently, this method has allowed ethology to progress significantly in the understanding of behaviour performances of social insects (Bonabeau et al. 1997). Computer simulations of social insects were built based on the concept of software or robotic agents modelling each insect individually, hence the term agent-based modelling. This has permitted to establish sufficient characteristics of behaviour and insect capabilities that lead to the formation of collective structures, such as the construction of termite nests, or the formation of organized groups for hunting or foraging in ants, or the formation of fish shoals, thermoregulation in beehives or the formation of social structures in wasps. In general, these computer simulations have shown that it was often not necessary that insect be equipped with complex cognitive structures in order to produce the collective formation of complex structures.

Physicists have also a tradition of using computers to make simulations of complex systems that allow them to elaborate their intuitions. For example, through experimentation with cellular automata—sorts of grids which cells can be in an "on" or "off" state and evolution depends on the state of their neighbours according to simple rules—they discovered how, starting either from initially random structures or completely uniform structures, complex patterns with non-trivial symmetries could be formed, resembling those observed in ice crystals, in the distribution of avalanches in sand piles or in mountains, dunes in the desert, the shape of fluvial deltas, galaxies or polyhedral bubbles in water cascades. For physicists, cellular automata are not what could be called physical models of ice crystals or avalanches, but they have played the role of metaphors and analogies which triggered a renewal in the way their community perceived and understood these phenomena (Vichniac et al. 1989; Weisbuch 1991; Bak 1996; Ball 2001).

9.3.2 Computer Science and the Origins of Language and Languages

It is also possible to use computers and agent-based simulations not only to help us understand the phenomena that characterize self-organization of matter, simple biological structures, or insect societies, but also to help us understand phenomena that characterize humans and their societies. The time has come to use computers and robots as scientific tools in human sciences. Thus, building artificial systems in the context of research into language origins and the evolution of languages is enjoying a growing popularity in the scientific community, exactly because it is a useful tool for studying the phenomena of language in relation to the complex interactions of its components (Steels 1997; Oudeyer and Kaplan 2007; Kaplan et al. 2008). These systems are put to two main types of use: (1) they serve to evaluate the internal coherence of verbally expressed theories already proposed, by clarifying all their hypotheses and verifying that they do indeed lead to the proposed conclusions (and

quite often one discovers errors in the assumptions as well as in the conclusions, which need to be revised); (2) they serve to explore and generate new theories, which themselves often appear when one simply tries to build an artificial system reproducing the verbal behaviour of humans.

A number of notable results have already been obtained and have opened the way for resolution of previously unanswered questions: the decentralized generation of lexical and semantic conventions in populations of agents (e.g. Kaplan 2001), the formation of shared inventories of vowels or syllables in groups of agents (e.g. Berrah et al. 1996; de Boer 2001; Oudeyer 2001, 2005a,b, 2006), with features of structural regularities greatly resembling those of human languages (e.g. Pierrehumbert 2001; Wedel 2006), the formation of conventionalized syntactic (e.g. Batali 1998) and grammatical structures (e.g. Steels 2005), the conditions under which combinatoriality, the property of systematic re-use, can be selected (Kirby 2001).

It is important to note that in the context of research on the origins of language, this methodology of the artificial is a *methodology for exploration*. It fits within an abductive scientific logic, i.e. a logic in which one searches for the premises that can lead to a given conclusion (instead of a deductive logic in which one searches for the conclusions that can be reached from a given set of premises).

The word ***model*** has here a different meaning than in its traditional use. Indeed, traditionally, modelling consists in observing a natural phenomenon and then abstracting fundamental mechanisms and variables on which a formalism is constructed that permits to predict reality as precisely as possible. Here, we are rather interested in qualitative investigations of broad types of mechanisms that may have been implemented in nature to solve given problems. Language is such a complex phenomenon that observations alone cannot allow researchers to *deduce* explanatory mechanisms. On the contrary, it is necessary to have beforehand a structured and rich conceptualisation of the space of hypotheses and mechanisms that might explain the complex properties of language evolution. This is where artificial computational systems, for which we use the term ***model***, come into play: they are used to refine our intuitions on the dynamics of language and languages formation, as well as to sculpt the space of hypotheses.

As a consequence, the aim of these computational models is not to establish the list of mechanisms responsible for the origins of given aspects of language. Rather, the objective is more modestly to try to build a list of potential candidates, and then to constrain the space of hypotheses, in particular by showing examples of mechanisms that are sufficient and examples of mechanisms that are not necessary.

9.4 The Speech Code

I will now illustrate this computer modelling work about the evolution of language and languages with the description of an experiment which focuses on the problem of the origins of speech, i.e. sound systems as physical vehicles of language (as gestures

can also be in signed languages). The goal of this experimentation is to foster the reconceptualisation of this scientific issue, through the evaluation of both existing and novel scientific hypotheses.

Discreteness and combinatoriality. Humans have a complex system of vocalizations. They are discrete and combinatorial, i.e. they are built from elementary units, "sculpted" in the auditory and phonatory continuum, which are systematically recombined and reused. These units exist at several levels (motor primitives to obstruct the air flow in the vocal tract, called *gestures*; gestures coordinations, called phonemes and which define vowels and consonants; syllables; etc. . .). Whereas the articulatory space is continuous and potentially permits an infinity of gestures and phonemes, each language discretizes this space in its own way, carving a repertoire of gestures and phonemes both small and finite (Studdert-Kennedy and Goldstein 2003). This is why it is sometimes referred as *phonemic coding*.

Universals and diversity. In spite of the great diversity of these elementary units in world languages, one can also find strong statistical regularities. For example, certain vowel systems are much more frequent than some others, such as the five-vowel system composed of vowels [e], [i], [o], [a] and [u]. The same fact can be stated for consonants. The way units are combined is also peculiar: one the one hand, not all sequences of phonemes are allowed to form syllables in each particular language, and on the other hand the associated sets of possible phoneme combinations can be grouped into generic types. This organization into generic types means that for example, one can summarize the possible phoneme combination to form syllables in Japanese ("moras") with types "CV/CVC/VC", where "VC" denotes syllables composed of two slots, with any Japanese vowel in the first slot and any Japanese consonant in the second slot.

Cultural sharing. Speech is a conventional code. Whereas statistical regularities can be observed across human languages, each linguistic community possesses its own way to perceive and categorize sounds as well as its own repertoire of rules for combinations. For example, native speakers of Japanese do not hear the difference between the *r* in *read* and the *l* in *lead*. How can a linguistic community come to form a code shared by all individuals and without a central coordinated control of the code?

Since the work of de Boer (2001) and Kaplan (2001), we have convincing hypotheses about how a new sound or a new word can propagate and be accepted in a given population. But these negotiation mechanisms, also called "consensus dynamics", assume the pre-existence of linguistic interaction conventions (Croft, this volume). Thus, the associated models concern mainly the formation and evolution of languages, but do not address directly the question of the origins of language. Indeed, when there was not already a conventional linguistic communication system, how could the first conventions have bootstrapped?

The model I will present focuses on this later question. It is obviously linked to the question of the origins of languages, because it is about understanding how a

speech code may have formed to be used as a basis for the first languages. The main difference among the two questions lies in the properties that shall characterize the mechanisms we are searching for. If one is interested in the origins of speech, one must search for an explanatory mechanism which assumes neither the existence of linguistic conventions, nor the existence of cognitive structures that are specific to language. Indeed, this would imply that we would have models of individuals that can already speak, and thus for which language would already have appeared.

9.5 Self-Organization and the Evolution of Speech

How did the first human speech codes form in a society with no language? As argued above, two kinds of answers must be given. First, a functional answer: they establish the function of vocalization systems, and shows that human systems, with the properties we described, are efficient to achieve this function. Liljencrantz and Lindblom (1972) proposed such an answer, and showed that the statistical regularities of human repertoires of phonemes were the most efficient in terms of the perceptual distinctiveness/articulatory cost compromise. This kind of answer is necessary, but not sufficient: it does not allow us to explain how evolution (genetic or cultural) could have found those quasi-optimal structures, and does not allow us either to explain how a given linguistic community can "choose" one solution among several quasi-optimal ones. In particular, it is possible that "naïve" Darwinian search with random mutations is not efficient enough to find complex structures such as those of the speech code: the search space is too large.

This is why a second kind of answer is necessary: we need to investigate how biological evolution might have generated and selected these structures. One possibility is to study how self-organization may have constrained the search space to help natural selection. This would consist in showing that a system much simpler than the structure we want to explain spontaneously *self-organizes into this structure*.

I will now present such a system and show how relatively simple premises—from an evolutionary point of view—can lead to the self-organized formation of speech codes.

9.5.1 A Computer Investigation of the Formation of Fundamental Structures of Speech

This computer model is agent-based: it consists in setting up virtual robots equipped with models of the auditory and phonatory apparatus coupled with a network of artificial neurons that connect perceptual and motor modalities. These artificial neurons determine the robots' behaviour, mainly consisting in vocal babbling. The babbling activity, coupled with the properties of plasticity characterizing neural networks,

allow the robots to learn the correspondences between the space of auditory perceptions and the space of vocal tract gestures. Finally, these robots are placed together in a shared environment where they can hear the vocalizations of their neighbours, which influence their own vocalizations, and wander around. I will show that a number of properties characterizing the vocalizations produced by robots in a given population emerge spontaneously.

More technically,[1] agents possess an artificial ear (whose properties can be modified to study their specific role: see below) capable of transforming an acoustic signal into neural impulses which stimulate neurons in an artificial perceptual neural map. They also possess a motor neural map whose neurons activation produce movements of a vocal tract model, which itself produces an acoustic wave (and which degree of realism can equally be modified). Both perceptual and motor maps are totally interconnected (see Fig. 9.4). Initially, internal parameters of all neurons, as well as the parameters of their connections, are random. To produce a vocalization, a robot randomly activates several motor neurons, whose internal parameters encode articulatory configurations which shall be reached in sequence, which in turn produces an articulatory trajectory and, through the vocal tract model, an acoustic signal that can

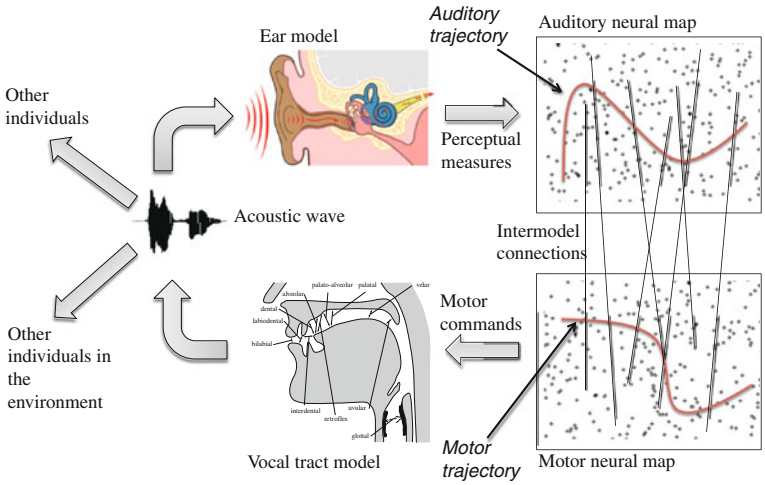

Fig. 9.4 The robot architecture in the artificial system. Each robot is equipped with an artificial ear, an artificial vocal tract, and two neural maps (perceptual and motor) that couple the ear and the vocal tract. These neural maps are initially random but characterized by two forms of plasticity: (1) intermodal connections evolve in such a way that the robot can learn the correspondences between auditory trajectories and associated motor trajectories when the robot is babbling; (2) neurons in each map evolve in such a way that they track the distribution of sounds heard by the robots. Thus, if one exposes a robot with a continuous flow of speech of a given language, its babbling will tune/align itself to the distribution of sounds in this language.

[1] We only give here a general description of the system: a detailed mathematical description is available in (Oudeyer 2006).

be perceived by the ear model. This is the basis of babbling, and explains why initially, robots produce vocalizations randomly spread in the vocal continuum. These neural networks are characterized by two forms of plasticity: (1) intermodal connections evolve in such a way that the robot learns the correspondences between auditory and motor trajectories perceived and produced when it is babbling[2]; (2) neurons in each map evolve in such a way that they tend to model the distribution of sounds heard by the robot[3]; (3) The connections between both neural maps are such that the distribution of sounds encoded in the motor map follows roughly the distribution of sounds encoded in the perceptual map. This implies that the neural architecture is such that robots have the tendency to produce the same distribution of sounds as the distribution of sounds that they hear around them. Thus, if one exposes a robot with a continuous flow of speech of a given language, its babbling will tune/align itself on the distribution of sounds in this language. For example, is this language contains the vowels [a,e,i] but not [o], the robot's babbling vocalizations will quickly contain [a,e,i] much more often than [o]. This behaviour corresponds to what is observed in young infants, and referred as "phonological attunement" (Vihman 1996).

9.5.2 A Unified Mechanism for the Self-Organization of Combinatoriality, of the Universals/Diversity Duality, and of Cultural Sharing

This type of architecture has frequently been used in the literature to model speech acquisition in children (Kohonen 1988; Sanguineti et al. 1998), in experiments in which the system learned to pronounce sounds/syllables of a language it was exposed to. Yet, the experiment I present here is different: one does not assume that a constituted speech code exists initially. On the contrary, one places a population of babbling robots together in a shared environment, such that they can both perceive their own babblings and those of their neighbours (see Fig. 9.5). Given that the properties of plasticity of their brains make them align their babbling vocalizations on those they hear around them, and as initially they all produce random vocalizations uniformly spread in the vocal space, the initial state is an equilibrium.

Yet, if one runs the simulation, one observes that this equilibrium is not stable. Indeed, noise—stochasticity—makes that from time to time, certain types of vocal-

[2] Connections between the two maps evolve according to Hebb's law: those that link neurons that are often activated in a correlated manner are reinforced, whereas those that link neurons with uncorrelated activation become weaker. These connections are initially random, and through babbling and Hebb's law, they self-organize and finally allow the robot to find motor commands that correspond to a given sound that he perceives.

[3] Neurons adapt to stimuli through sensitization: their dynamics is such that if a stimulus S is perceived, then they are modified such that if the same stimulus S would be presented just afterwards they would be more activated than the first time, and the amount of modification depends exponentially on their activation (strongly activated neurons are modified most).

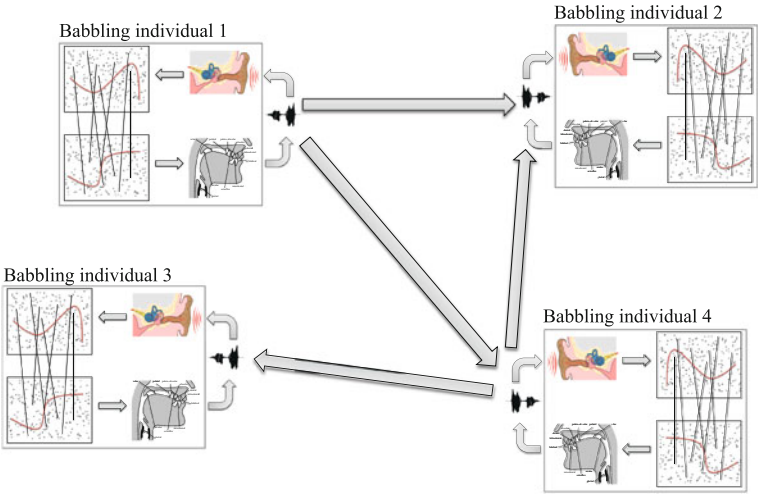

Fig. 9.5 In the experiment, babbling robots are placed in a shared environment and can hear each other's vocalizations in addition to their own. Their neural networks are initially random, thus their vocalizations are initially random and unorganized. Given that the properties of plasticity of their brains push them to align their vocalizations with the one they hear, and as they statistically all produce vocalizations randomly uniformly spread in the articulatory space, the initial state is an equilibrium. Yet, this equilibrium is unstable and the noise inherent to their interactions and to the neural dynamics provokes a symmetry breaking: vocalizations crystallize and become discrete, combinatorial and shared by all the members of the population.

izations are pronounced a little more often than others. Now, the mechanism that couples perception and production creates a positive feedback loop: these deviations from the mean get amplified when they are big enough, and the system's symmetry breaks. Neural maps self-organize into clusters of neurons encoding particular acoustic and articulatory configurations in the space of vocalizations (see Fig. 9.6). In brief, the continuous space of vocalization has been discretized. The vocalizations produced by agents are not holistic anymore, but discrete and combinatorial: they are systematically built through the sequencing of key configurations, that we can call phonemes. One sees the formation of phonemic coding, i.e. discrete combinatorial speech codes as described earlier. Besides, the system of phonemes that self-organize is shared by all robots of a given simulation, and is different in different simulations. Thus, one observes the formation of a "cultural convention" which can be diverse across groups.

As a matter of fact, several variants of this experiment can be set up and permit to refine the conclusions. First, it is possible to experiment what happens when there is only one single robot which is listening to itself babbling. In this case, one also observes a crystallization of vocalizations: it quickly ends up producing vocal trajectories that systematically reuse few key articulatory configurations. One can deduce that the formation of phonemic coding, i.e. discreteness and combinatoriality, is not

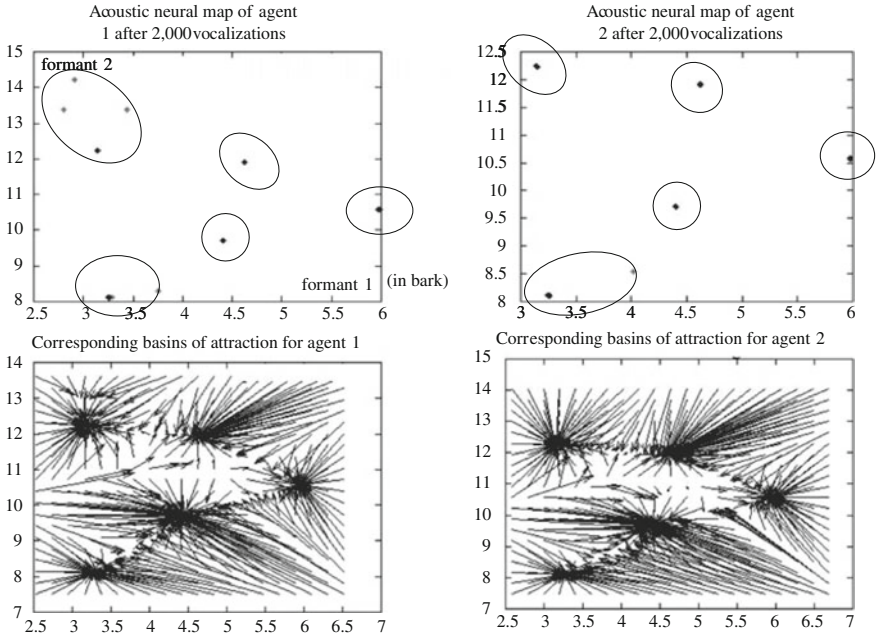

Fig. 9.6 Very quickly, initial symmetry in the system is broken, and neurons which initially encoded random vocal configurations now encode a small number of configurations which are systematically reused by agents when they babble: the vocal space has been discretized. Besides, these self-organized elementary configurations are the same in all robots of the same population, but different across populations. One can see it on this figure which represents the perceptual neural maps of two agents after 2000 vocalizations, where the configurations encoded by neurons are clustered (*top*), as well as their corresponding distributions (*bottom*). The auditory space is here projected on the first and the second effective formant, expressed in barks, which allows us to visualize the self-organized vowel systems

the result of social interactions but rather of the internal coupling between speech perception and production. Yet, whereas the vocalizations of isolated babbling agents will crystallize on different vocalization systems, these systems will spontaneously synchronize when they share the same environment and are capable of hearing each other: in this case, the self-organized systems are approximately the same in all agents on the same population.

A second significant variant of this experiment consists in varying the morpho-physiological properties of the auditory and phonatory systems so as to determine the impact of these properties on the systems that form (or do not form). In particular, an essential property of the speech organs is the non-linearity of the function that maps articulatory configurations to acoustic waves to auditory perceptions. The human vocal tract is indeed such that for certain articulatory configurations, small variations produce small variations in the perceived sound, while for other articulatory configurations, small variations produce large variations in the perceived sound.

Now, this property is central in several theories that propose to explain why speech is phonemically coded, such as in Stevens' quantal theory (Stevens 1989) or in the DRM model (Mrayati et al. 1988). It is possible to use models of the ear and the vocal tract that are realistic and include this type of non-linearity, but it is also possible to construct on purpose a non-realistic model to evaluate the specific impact of the non-linearities. These experiments were run, and with a linear model one observes that in a population of babbling robots the crystallisation we presented above still happens: vocalizations self-organize into a combinatorial system in which particular articulatory configurations are systematically reused in vocal trajectories. Thus, we can draw a first conclusion: these simulations show that phonemic coding can appear spontaneously without non-linearities in the auditory-phonatory system. This does not imply that non-linearities do not accelerate the formation of phonemic coding, but that they are not necessary, as proposed in the quantal theory or in the DRM model.

Yet, if one looks at the distribution of self-organized key vocal configurations when one uses the linear auditory-phonatory system (which one can see as kinds of phonemes), one observes that these configurations are globally positioned randomly uniformly in the space of possible vocal configurations. But when one uses a realistic ear and vocal tract model, reproducing in particular the properties of vowel production and perception[4], one observes an additional phenomenon. Besides the crystallization which is the qualitatively same as with the linear model, the vocalization systems that form are characterized by statistical regularities that share many similarities with human vowel systems. For example, one can collect statistics on the vowel systems that appear as key reused configurations in the self-organized systems when one runs many simulations. The results, illustrated in Fig. 9.7, show that on the one hand a diversity of systems appear, and on the other hand that certain vowel systems appear much more often than others. As a conclusion, one observes the same duality of universals and diversity that characterized human languages, and the simulation suggests a unified explanation:

(1) The dynamical system composed by the set of babbling robots and the internal sensorimotor couplings possesses a number of attractors which are culturally shared combinatorial vocalization systems;
(2) Under the influence of noise and small variations of initial conditions, the dynamical system falls in a particular attractor, which allows us to explain the "decentralized collective choice" made by the population to adopt a system rather than another;
(3) Non-linearities in the auditory and phonatory systems introduce asymmetries between attractors: some of them have a larger basin of attraction, in particular those for which phonemes are in zones where small articulatory variations provoke small perceptual variations, which increases the probability that the system falls in such attractors.

[4] See (Oudeyer 2006) for a precise description of the model based on the work of (de Boer 2001).

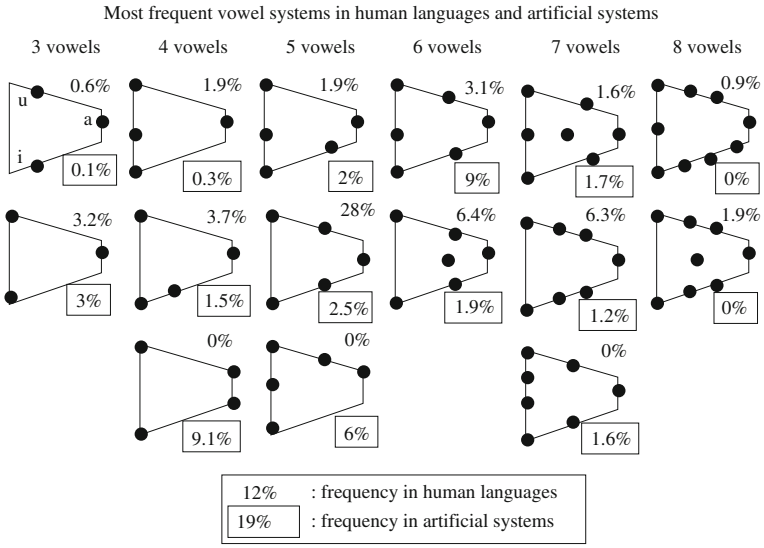

Fig. 9.7 Comparison between the distribution of vowel systems formed by populations of robots and in human languages of the UPSID database (Maddieson 1984). Vowel systems are represented on the vocalic triangle, which horizontal dimension corresponds to the first formant, and vertical dimension corresponds to the second effective formant. One observes that the most frequent systems in the artificial system are the same as in human languages (in particular the symmetric 5 vowel system /a,e,i,o,u/ with 25 % in artificial systems and 28 % in human languages)

Furthermore, there is not only a structural correspondence between simulations and reality, but the vowel systems that appear most frequently in robot populations are approximately the same, and in the same proportion, as those that appear most frequently in human languages. Thus, there is a quantitative relationship. One can conclude that the non-linearities of the auditory and phonatory systems are decisive to explain why certain systems of phonemes are statistically more frequent than others. Yet, the very existence of these phonemes, i.e. the existence of a vocalization system in which invariant articulatory and auditory configurations are systematically reused, is not necessitated by these non-linearities.

It is important to note that in these simulations, neural architectures are characterized by several parameters, and that all values of these parameters do not lead to the above mentioned results. Nevertheless, only one parameter has a critical influence on the results (Oudeyer 2006): neurons are characterized by a selectivity to stimuli (σ) that can be focused or wide. If this selectivity is too focused, no crystallization happens, but the robots remain capable to learn the relationships between the auditory space and the phonatory/motor space. If the selectivity is too wide, the system crystallizes in a degenerate state in which all vocalization are exactly the same and are stationary: there is only one phoneme. Yet, there is a large range of intermediary values between

these two extremes that allows us to observe a crystallization in which a combinatorial system with multiple phonemes self-organizes in a population of babbling robots.

9.5.3 Towards a Novel Vision of Evolutionary Scenarios of the Origins of Coded Speech

The preceding discussion has important consequences if one uses the model to imagine evolutionary scenarios that may have led to the formation of the first vocalization systems featuring the fundamental properties of contemporary human speech. Indeed, they imply that many parameter variations in the neural architecture permit to generate combinatorial speech systems shared by all members of a community. Also, they imply that with such a neural architecture, a combinatorial phonemically coded vocalization system can appear without assuming special properties of the auditory and phonatory system: the only assumption is that a certain variety of sounds can be produced, but non-linearities are not required. Finally, the architecture is in itself relatively primitive: it includes neural units which intrinsic properties (dynamics and plasticity) are classical and are functionally similar to most of neural units in mammal brains (Oudeyer 2006). The specificity of this architecture relies in the systematic and direct plastic connections between the auditory and motor maps. Yet, this specificity characterizes the basic building blocks of the capacity to learn to imitate holistic sounds, which we call here ***holistic adaptive vocal imitation*** and is sometimes also called adaptive vocal mimicry or vocal learning in the animal literature, which is a capacity functionally more restricted than the capacity to learn and share combinatorial phonemically coded speech sounds. This leads us to the following evolutionary scenario to conceptualize the origins of combinatorial vocalization systems culturally shared by all individuals of a community:

(1) Adaptive vocal imitation is present in many animals (Snowdown and Hausberger 1997; Hauser 1997) which possess shared and learnt vocalization systems, but which do not possess language. Ethologists have identified many potential reproductive advantages characterizing the capacity of adaptive vocal imitation in a community of individuals (i.e. this allows individuals to mark their group membership). Thus, it is reasonable to think that before being capable of speech and language, humans may have evolved the capacity to imitate vocally;
(2) Being capable of adaptive vocal imitation, as well as the related reproductive advantages identified in non-human animals, does not imply and does not necessitate a combinatorial phonemically coded vocalization system. As an example, the range of parameter values for which the selectivity σ is focused allows the robots to learn efficiently the vocal perceptuo-motor correspondences without generating a phonemic system;
(3) Now, if one imagines an ecological context in which the presence of a combinatorial vocalization system would provide a reproductive advantage to those that possess it, then the experiments that we described permit to state that a simple

change in the value of the σ parameter in motor and perceptual neural map would lead to the spontaneous formation of vocalization systems which feature several fundamental properties of contemporary human speech systems. This allows us to understand that what may have been a great leap forward for language, i.e. the formation of shared combinatorial vocalizations, may be the consequence of a small biological change together with the self-organizing properties of neural matter and multi-modal sensorimotor coupling.

Moreover, this scenario in which phonemically coded vocalization systems would have been selected thanks to the reproductive advantage that they may have provided is not the only one that the computational model can support and refine. Indeed, I explained above that in the range of σ values that allows combinatorial systems to be formed, the capacity of adaptive vocal imitation is intact and equally efficient. Besides, performance being equal, the transition of σ among this range and the range of more focused values for selectivity does not a priori imply a metabolic cost. This implies that in an ecological context in which those neural structures appeared under a selective pressure for adaptive vocal imitation, neutral mutation/variation and neutral drift may have happened and generated spontaneously shared phonemically coded vocalization systems without a selective pressure for language. One observation makes this scenario particularly stimulating: among animal species capable of adaptive vocal imitation in which culturally shared sound systems exist, but which do not possess language, several of them produce vocalizations or songs structured around the systematic reuse of basic units. For example, this can be observed in zebra finches (Brenowitz and Beecher 2005) or in humpback whales (Tyack 1981). The function of this quasi-phonemic structuring has been only little understood so far in ethology. Besides, because the model I presented is neutral with respect to many properties of the auditory and phonatory systems, and because the neural architecture that it assumes corresponds to the minimal equipment for adaptive vocal imitation, it can be applied to the formation of songs in these animals. In this case, it provides a hypothesis reinforced by the current scientific uncertainty about the function of combinatorial coding in these songs: combinatorial and systematically reused units may have been generated spontaneously as a collateral effect of the biological equipment for adaptive vocal imitation. Thus, it is also reasonable to imagine that this may have been the case in humans: combinatorial speech systems may have been recruited only later on to achieve their current linguistic function. This implies that several fundamental properties of contemporary human speech systems may be exaptations.[5]

[5] This term was introduced in (Gould and Vrba 1982). It refers to the use of a biological feature/structure for a function A which is different than the function B for which it was initially evolutionary selected.

9.6 Conclusion

Through the implementation and use of a computer model, I have shown how a relatively simple neural architecture coupling auditory and phonatory modalities permitted, via a self-organizing dynamics, the spontaneous formation of combinatorial phonemically coded vocalization systems shared by all members of a community, and characterized by the duality of universals and diversity. The first contribution of this work is that for the first time it provides a unified explanation of these three phenomena.

Besides, this multimodal coupling architecture corresponds to the minimal neural kit required by adaptive vocal imitation, and does not include biological elements that are specific to human speech. Given that crystallization happens in a large range of the parameter space, this shows that the transition from inarticulated vocalization systems to human-like speech codes may have been largely due to a modest biological innovation. Indeed, the model indicates that neuronal structures that encode a priori and specifically phonemic organization, as well as typical regularities of speech, do not need to be innately generated to allow the formation of such speech code. This is the second contribution of this work: it allows us to understand how the self-organizing properties of simple neural structures may have constrained the space of biological vocalization structures and how speech codes may have been generated and selected during phylogenesis.

These new hypotheses may not have been identified without the use of computer models and simulations, because the underlying dynamics are complex and difficult to anticipate just through verbal reasoning. This illustrates the potential importance of these new methodological tools in human and biological sciences. Yet, such computer models abstract many biological and behavioural mechanisms, and provide primarily a theoretical investigation of the space of hypotheses: once this space is reconceptualised, and the internal coherence of hypotheses evaluated through computer simulations, validation work and grounding of these hypotheses in biological field observations remains to be done. Thus, the third contribution of this work is, more than just the elaboration of specific hypotheses, also the construction of a framework and of tools that allow us to develop new intuitions and new concepts for our understanding of the origins and evolution of speech.

Acknowledgments *This work was in major part achieved in the Sony Computer Science Laboratory, Paris, and benefited from the support of Luc Steels.*

References

Ashby, W.R.: An Introduction to Cybernetics. Chapman & Hall, London (1956)

Bak, P.: How Nature Works: The Science of Self-Organized Criticality. Copernicus, New York (1996)

Ball, P.: The Self-Made Tapestry : Pattern Formation in Nature. Oxford University Press, Oxford (2001)

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

Berrah, A., Glotin, H., Laboissière, R., Bessière, P., Boë, L.J.: From form to formation of phonetic structures: an evolutionary computing perspective. In: 13th International Conference on Evolutionary Computing and Machine Learning (ICML), pp 23–29. MIT Press, Bari (1996)

Bonabeau, E., Theraulaz, G., Deneubourg, J.L., Aron, S., Camazine, S.: Self-organization in social insects. Trends in Ecology and Evolution **12**, 188–193 (1997)

Brenowitz, E.A., Beecher, M.D.: Song learning in birds : diversity and plasticity, opportunities and challenges. Trends in Neuroscience **28**(3), 127–132 (2005)

Camazine, S., Deneubourg, J.-L., Franks, N.R., Sneyd, J., Theraulaz, G., Bonabeau, E.: Self-Organization in Biological Systems. Princeton University Press, Princeton (2002)

Dawkins, R.: The Blind Watchmaker. W. W. Norton and Company, New York (1986)

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

Elredge, N., Gould, S.J.: Punctuated equilibria : an alternative to phylogenetic gradualism. In: Schopf, T.J.M. (ed.) Models in Paleobiology, pp. 82–115. Freeman, San Francisco (1972)

Gould, S.J., Vrba, E.S.: Exaptation: a missing term in the science of form. Paleobiology **8**(1), 4–15 (1982)

Hauser, M.D.: The Evolution of Communication. MIT Press, BradfordBooks, Cambridge (1997)

Kauffman, S.: At Home in the Universe : The Search for Laws of Self-Organization and Complexity. Oxford University Press, Oxford (1996)

Keefe, A., Szostak, J.: Functional proteins from a random sequence library. Nature **410**, 715–718 (2001)

Kaplan, F.: La naissance d' une langue chez les robots. Hermès (2001)

Kaplan, F., Oudeyer, P.-Y., Bergen, B.: Computational models in the debate over language learnability. Infant Child Dev. **17**(1), 55–80 (2008)

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

Kohonen, T.: The neural phonetic typewriter. Computer **21**(3), 11–22 (1988)

Liljencrantz, J., Lindblom, B.: Numerical simulation of vowel quality systems: the role of perceptual contrast. Language **48**, 839–862 (1972)

Maddieson, I.: Patterns of Sounds. Cambridge University Press, Cambridge (1984)

Mrayati, M., Carre, R., Guerin, B.: Distinctive regions and modes: a new theory of speech production. Speech Commun. **7**, 257–286 (1988)

Nicolis, G., Prigogine, I.: Self-Organization in Non-equilibrium Systems: From Dissipative Structures to Order Through Fluctuations. Wiley, New York (1977)

Oudeyer, P.-Y.: Origins and learnability of syllable systems, a cultural evolutionary model. In: Collet, P., Fonlupt, C., Hao, J., Lutton, E., Schoenauer, M. (eds.) Artificial Evolution; LNCS 2310, pp. 143–155 (2001)

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

Oudeyer, P.-Y.: The self-organization of combinatoriality and phonotactics in vocalization systems. Connection Sci. **17**(3–4), 325–341 (2005b)

Oudeyer, P-Y.: Self-Organization in the Evolution of Speech. Oxford University Press, Oxford (2006)

Oudeyer, P.-Y., Kaplan, F.: Language evolution as a darwinian process: computational studies. Cogn. Process. **8**(1), 21–35 (2007)

Pierrehumbert, J.: Exemplar dynamics: word frequency, lenition, and contrast. In: Bybee, J., Hopper, P. (eds.) Frequency Effects and the Emergence of Linguistic Structure. John Benjamins, Amsterdam (2001)

Pfeifer, R., Scheier, C.: Understanding Intelligence. MIT Press, Cambridge (1999)

Sanguineti, V., Laboissière, R., Ostry, D.J.: A dynamic biomechanical model for neural control of speech production. J. Acoust. Soc. Am. **103**(3), 1615–1627 (1998)

Snowdown, C.T., Hausberger, M.: Social Influences on Vocal Development. Cambridge University Press, Cambridge (1997)
Steels, L.: The synthetic modeling of language origins. Evolution of communication, **1**(1), 1-34 (1997).
Steels, L.: The methodology of the artificial. Behav. Brain Sci. **24**(6), 1077–1078 (2001)
Steels, L.: The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. Connection Sci. **17**(3–4), 213–230 (2005)
Stevens, K.N.: On the quantal nature of speech. J. Phonetics **17**, 3–45 (1989)
Studdert-Kennedy, M., Goldstein, L.: Launching language: the gestural origin of discrete infinity. In: Christiansen, M., Kirby, S. (eds.) Language Evolution: The State of the Art, pp. 235–254. Oxford University Press, Oxford (2003)
Tritton, D.J.: Physical Fluid Dynamics. Oxford University Press, Oxford (1988)
Tyack, P.: Interactions between singing hawaian humpback whales and conspecifics nearby. Behav. Ecol. sociobiol. **8**(2), 105–116 (1981)
Vichniac, G., Manneville, P., Boccara, N., Bidaux, R. (eds.): Cellular automata and modeling of complex systems. In: Workshop Les Houches. Springer, Heidelberg (1989)
Vihman, M.: Phonological Development: The Origins of Language in the Child. Blackwell, Cambridge (1996)
Waldrop, M.: Spontaneous order, evolution, and life. Science **247**, 1543–1545 (1990)
Wedel, A.: Exemplar models, evolution and language change. Linguist. Rev. **23**, 247–274 (2006)
Weisbuch, G.: Complex systems dynamics. Addison-Wesley, Redwood City, CA (1991)