# Origins and Learnability of Syllable Systems: a Cultural Evolutionary Model

Pierre-yves Oudeyer
py@csl.sony.fr

Sony Computer Science Lab, Paris

**Abstract.** This paper presents a model of the origins of syllable systems that brings plausibility to the theory which claims that language learning, and in particular phonological acquisition, needs not innate linguistically specific information, as believed by many researchers of the Chomskyan school, but is rather made possible by the interaction between general motor, perceptual, cognitive and social constraints through a self-organizing process. The strategy is to replace the question of acquisition in a larger and evolutionary (cultural) framework: the model addresses the question of the origins of syllable systems (syllables are the major phonological units in speech). It is based on the artificial life methodology of building a society of agents, endowed with motor, perceptual and cognitive apparati that are generic and realistic. We show that agents effectively build sound systems and how these sound systems relate to existing human sound systems. Results concerning the learnability of the produced sound systems by fresh/baby agents are detailed: the critical period effect and the artificial language effect can effectively be predicted by our model. The ability of children to learn sound systems is explained by the evolutionary history of these sound systems, which were precisely shaped so as to fit the ecological niche formed by the brains and bodies of these children, and not the other way around (as advocated by Chomskyan approaches to language).

## 1 Introduction

Children learn language, and in particular sound systems, incredibly easily and fast, in spite of its apparent idiosyncratic complexity and noisy learning conditions. Many researchers, especially those in the Chomskyan school, believe this can not be possible without a substantial genetically linguistically specific endowment. In fact the role of learning in language development is thought to be very minor ([20]) and reduced to the setting of a few parameters like in the Principles and Parameters theory ([5]) or in Optimality Theory ([1]). Yet, a growing number of researchers (but still the minority among language researchers) have challenged this view, and think that no linguistically specific innate neural device is necessary to account for the oddities of language learning (and structure): rather, they propose that these result from the complex interactions between a number of general motor, perceptual, cognitive, social and functional constraints, and this in a mainly cultural manner ([24]). The word "constraint" is used in its most general meaning: it can be "obstacle" or "opportunity". According to this view, language emerged and evolved so as to fit the ecological niche of initially

non-speaking human brains and bodies. In brief, the languages that humans speak were selected so as to be learnable (and not the other way around as suggested in ([19]).

As a consequence, if we take that view as we do here, it seems natural to put oneself in a cultural evolutionary framework: if one wants to understand the principles of language learning, one has to understand the principles of language emergence and evolution, i.e. language epigenesis. This paper follows this idea and illustrates the theory with the case-study of the origins and learnability of syllable systems, which are thought to be a fundamental unit of the complex phonological systems of human languages ([18]). We present a computational model in the spirit of past work on the origins of language ([24], [12]). Among related existing models of the origins of sound systems, there exists two models of the origins of vowel systems: Lindblom [14] showed that the optimization of a number of analytically defined perceptual constraints could predict the most frequent vowel systems, whereas de Boer ([4]) developed an operational multi-agent based model of how vowel systems could have been built culturally. Also, Redford ([21]) made a model similar to Lindblom's concerning syllable systems. Yet, this work is focused uniquely on the properties of sound systems, but does not give cues of how it could actually have been built and how it relates to the cognitive abilities of speakers. The model presented here is inspired from the work of de Boer, in particular for the evolutionary architecture (the imitation game). The difference is that first we are dealing with syllables here, and secondly we tried to model constraints in a more embodied and situated manner. Indeed, previous models have shown how important constraints are to the shape of sound systems: when dealing with too abstract constraints, there is a danger to find wrong explanations. Furthermore, Redford showed that certain phenomena can be understood only by considering the interactions between constraints, so models should try to incorporate most of them. The present model builds on a first very simple model detailed in ([27]). It is much more realistic and less arbitrarily biased at both morphological and cognitive levels, and while only studies of efficiency were performed with the previous model, structural properties and learnability of the produced sound systems are here presented. Due to space limitations, this paper focuses on the learnability aspects of the behavior of the model and its implications on theories of human sound systems. The fine details of the architecture will be described in a longer paper, and the structural properties are detailed in a companion paper ([25]).

The next section presents an overview of the model with its different modules. Then we summarize the efficiency of the system as well as the structural properties of the produced syllable systems. Finally, we explore in details their learnability and the implications on theories of language.

## 2 The model

### 2.1 The imitation game

Central to the model is the way agents interact. We use here the concept of game, operationally used in a number of computational models of the origins of language ([24], [27]). A game is a sort of protocol that describes the outline of a conversation, allowing agents to coordinate by knowing who should try to say

what kind of things at a particular moment. Here we use the "imitation game" developed by de Boer for his experiments on the emergence of vowel systems.

A round of a game involves two agents, one being called the speaker, and the other the hearer. Here we just retain from their internal architecture that they possess a repertoire of items/syllables, with a score associated to each of them (this is the categorical memory described below). The speaker initiates the conversation by picking up one item in its repertoire and utters it. Then the hearer tries to imitate this sound by producing the item in its repertoire that matches best with what he heard. The speaker then evaluates whether the imitation was good or not by checking whether the best match to this imitation in his repertoire corresponds to the item he uttered initially. He then gives a feedback signal to the hearer in a non-linguistic manner. Finally, each agent updates its repertoire. If the imitation succeeded, the scores of involved items increase. Otherwise, the score of the association used by the speaker decreases and there are 2 possibilities for the hearer: either the score of the association he used was below a certain threshold, and this item is modified by the agent who tries to find a better one; or the score was above this threshold, which means that it may not be a good idea to change this item, and a new item is created, as close to the utterance of the speaker as the agent can do given its constraints and knowledge at this time of its life. Regularly the repertoire is cleant by removing the items that have a score too low. Initially, the repertoires of agents are empty. New items are added either by invention, which takes place regularly in response to the need of growing the repertoire, or by learning from others.

### 2.2 The production module

**Vocal tract** A physical model of the vocal tract is used, based on an implementation of Cook's model ([6]). It consists in modeling the vocal tract together with the nasal tract as a set of tubes that act as filters, into which are sent acoustic waves produced by a model of the glottis and a noise source. There are 8 control parameters for the shape of the vocal tract, used for the production of syllables. Finally, articulators have a certain stiffness and inertia.

**Control system** The control system is responsible for driving the vocal tract shape parameters given an articulatory program, which is the articulatory specification of the syllable. Here we consider the syllable from the point of view of the frame-content theory ([18]) which defines it as an oscillation of the jaw (the frame) modulated by intermediary specific articulatory configurations, which represent a segmental content (the content) corresponding to what one may call phonemes. A very important aspect of syllables is that they are not a mere sequencing of segments by juxtaposition: co-articulation takes place, which means that each segment is influenced by its neighbors. This is crucial because it determines which syllables are difficult to pronounce and imitate. We model here co-articulation in a way very similar to what is described in ([17]), where segments are targets in a number of articulatory dimensions. The difference is that we provide a biologically plausible implementation inspired from a number of neuroscientific findings ([3]) and that uses techniques developed in the field of behavior-based robotics ([2]). This will be detailed in a forthcoming longer paper.

The constraint of jaw oscillation is modeled by a force that pulls in the direction of the position the articulators would have if the syllable was a pure frame, which means an oscillation without intermediary targets. This can be viewed as an elastic whose rest position at each time step is the pure frame configuration at this time step. Finally, and crucially, we introduce a notion of articulatory cost, which consists in measuring on the one hand the effort necessary to achieve an articulatory program and on the other hand the difficulty of this articulatory program (how well targets are reached given all the constraints). This cost is used to model the principle of least effort explained in ([14]): easy articulatory programs/syllables tend to be remembered more easily than others. Agents are given initially a set of pre-defined targets that can be thought to come from an imitation game on simple sounds (which means they do not involve movements of the articulators) as described in ([4]). Although the degrees of freedom that we can control here do not correspond exactly to the degrees that are used to define human phonemes, we chose values that allow them to be good metaphors of vowels (V), liquids (C1) and plosives (C2), which mean sonorant, less sonorant, and even less sonorant phonemes (sonority is directly related to the degree of obstruction of the air flow, which mean the more articulators are opened, the more they contribute to a high sonority of the phoneme).

### 2.3 The perception module

The ear of agents consists of a model of the cochlea, and in particular the basilar membrane, as described in ([16]). It provides the successive excitation of this membrane over time. Each excitation trajectory is discretized both over time and frequency: 20 frequency bins are used and a sample is extracted every 10 ms. Next the trajectory is time normalized so as to be of length 25. As a measure of similarity between two perceptual trajectories, we used a technique well-known in the field of speech recognition, dynamic time warping ([22]). Agents use this measure to compute which item in their memory is closest. No segmentation into "phonemes" is done in the recognition process: the recognition is done over the complete unsegmented sound. Agents discover what phonemes compose the syllable only after recognition of the syllable and by looking at the articulatory program associated to the matched perceptual trajectory in the exemplar. This follows a view defended by a number of researchers ([23]) who showed with psychological experiments that the syllable was the primary unit of recognition, and that phoneme recognition came only after.

### 2.4 The brain module

The knowledge management module of our agents consists of 2 memories of exemplars and a mechanism to shape and use them. A first memory (the "inverse mapping" memory) consists of a set, limited in size, of exemplars that serve in the imitation process: they represent the skills of agents for this task. Exemplars consists in associations between articulatory programs and corresponding perceptual trajectories. The second memory (the categorical memory) is in fact a subset of the inverse-mapping memory, to which is added to each exemplar a score. Categorical memory is used to represent the particular sounds that count

as categories in the sound system being collectively built by agents (corresponding exemplars are prototypes for categories). It corresponds to the memory of prototypes classically used in the imitation game ([4]).

Initially, the inverse mapping memory is built through babbling. Agents generate random articulatory programs, execute them with the control module and perceive the produced sound. They store each trial with a probability inverse to the articulatory cost involved (prob=1-cost). The number of exemplars that can be stored in this memory is typically quite limited (in the experiments presented below, there are 100 exemplars whereas the total number of possible syllables is slightly above 12000). So initially the inverse mapping memory is composed of exemplars which tends to be more numerous in zones where the cost is low than in zones where the cost is higher. As far as the categorical memory is concerned, it is initially empty, and will grow through learning and invention.

When an agent hears a sound and wants to imitate it, he first looks up in its categorical memory (if it is not empty) and find the item whose perceptual trajectory is most similar to the one he just heard. Then he executes the associated articulatory program. Now, after the interaction is finished, in any case (either it succeeded or failed), it will try to improve its imitation. To do that, it finds in its inverse mapping memory the item (it) whose perceptual trajectory matches best (it may not be the same as the categorical item). Then it tries through babbling a small number of articulatory variations of this item that do not belong to the memory: each articulatory trial item is a mutated version of it, i.e. one target has been changed or added or deleted. This can be thought of the agent hearing at a point "ble", and having in its memory the closest item being "fle". Then it may try "vle", "fli", or even "ble" if the chance decides so (indeed, not all possible mutations are tried, which models a time constraints: here they typically try 10 mutations). The important point is that these mutation trials are not forgotten for the future (some of them may be useless now, but very useful in the future): each of them is remembered with a probability inverse to its articulatory cost. Of course, as we have memory limitation, when new items are added to the inverse mapping memory, some others have to be pruned. The strategy chosen here is the least biased: for each new item, a randomly chosen item is also deleted (only the items that belong to categorical memory can not be deleted).

The evolution of inverse mapping memory implied by this mechanism is as follows. Whereas at the beginning items are spread uniformly across "iso-cost" regions, which means skills are both general and imprecise (they have some capacity of imitation of many kind of sounds, but not very precise), at the end items are clustered in certain zones corresponding to the particular sound system of the society of agents, which means skills are both specialized and precise. This is due to the fact that exemplars closest to sound produced by other agents are differentiated and lead to an increase of exemplars in their local region at the cost of a decrease elsewhere.

It is interesting to remark that what goes on in the head of each agent is very similar to what happens in genetic evolution. One can view the set of exemplars that an agent possess as a population of individuals/genomes, each defined by the sequence of articulatory goals. The fitness function of each individual/syllable is defines by how often it leads to successful imitation when it is used, in both speaker and hearer roles. This population of individuals evolve through a generate and select process, generation being performed through a combination of

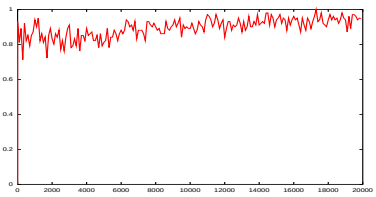


**Fig. 1.** Example of the evolution of success in interactions for a society of agents who build a sound system from scratch

completely random inventions and mutations of syllables (= one changes one articulatory goal), and selection using the scores of each syllable. The original thing here as compared to many simulations modeling either genetic or cultural evolution, is that the fitness function is not fixed but evolves with time: indeed the fitness of one syllable depends on the population of syllables in the heads of other agents whose fitness itself depends on this syllable. So we have a case of coupled dynamic fitness landscapes. As we will see, what happens is that those fitness landscapes synchronize at some point, they become very similar and stable. Also, the fitness of one syllable depends of the other syllables/exemplars in the memory of the agent: indeed, if a syllable is alone in its part of the space, for example, then few syllables of this area will be produced and other agents will have less opportunity to be practice imitation of this kind of syllable, and so there is a high probability that the syllable will be pruned. The consequence of this is that groups selection also happens.

## 3 Efficiency

The first thing one wants to know is simply whether populations of agents manage to develop a sound system of reasonable size and that allows them to communicate (imitations are successful). Figure 1 and 2 show an example of experiment involving 15 agents, with a memory limit on inverse-mapping memory of 100 exemplars, with vocalizations comprising between 2 and 4 targets included among 10 possible ones (which means that at a given moment, one agent never knows more than about 0.8 percent of the syllable space). In figure 1, each point represents the average success in the last 100 games, and on figure 2, each point represents the average size of categorical memory in the population (i.e. the mean number of syllables in agents' repertoires). We see that of course the success is very high right from the start: this is normal since at the beginning agents have basically one or two syllables in their repertoire, which implies that even if an imitation is quite bad in the absolute, it will still get well matched. The challenge is actually to remain at a high success rate while increasing the size of the repertoires. The 2 graphs shows that it is the case. To make these results convincing, the experiments was repeated 20 times (doing it more is rather infeasible since each experiment basically lasts about 2 days), and the average number of syllables and success was measured in the last 1000 games (over a total of 20000 games): 96.9 percent is the mean success and 79.1 is the mean number of categories/syllables.

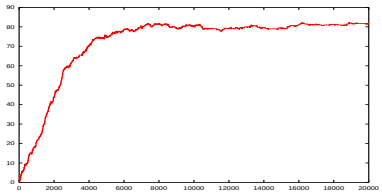


**Fig. 2.** Corresponding evolution of mean number of items/categories in repertoires of agents along with time

The fact that the success remains high as the size of repertoire increases can be explained. At the beginning, agents have very few items in their repertoires, so even if their imitations are bad in the absolute, they will be successfully recognized since recognition is done by nearest-neighbours (for example, when 2 agents have only 1 item, no confusion is possible since there is only 1 category!). As time goes on, while their repertoires become larger, their imitation skills are also increasing: indeed, agents explore the articulatory/acoustic mapping locally in areas where they hear other utter sounds, and the new sounds they create are hence also in these areas. The consequence is a positive feed-back loop which makes that agents who knew very different parts of the mapping initially tend to synchronize their knowledge and become expert in the same (small) area (whereas at the beginning they have skills to imitate very different kinds of sounds, but are poor when it becomes to make subtle distinctions in small areas).

## 4 Structural Properties

The properties summarized here are detailed in ([25]). The produced syllable systems have structures very similar to what we observe in human languages. On the one hand, a number of universal tendencies were found, like the ranking of syllable types along their frequency (CV $\geq$ CVC $\geq$ CCV $\geq$ CVVC/CCVC/CVCC); Also the model predicts the preference for syllables respecting the sonority hierarchy principle, which states that within a syllable, the sonority (or degree of obstruction of the air flow in the vocal apparatus), first increases until a peak (the nucleus) and then decreases. On the other hand, the diversity observed in human languages could also be observed: some syllable systems did not follow the trend in syllable type preference, and categorical differences exist (some syllable systems have certain syllable types not possessed by others). This constitutes a viable alternative to the mainstream view on phonological systems, optimality theory ([1]), which require the presence of innate linguistically specific constraints in the genome to account for universal tendencies (an example of constraint is the *COMPLEX constraint which states that syllables can have at most one consonant at an edge), and explains diversity by different orderings in the strengths of these constraints (which is basically the only thing that is learnt).

## 5 Learnability properties

The learnability of the produced systems by fresh agents confronted directly with the complete sound is an important question. Indeed, more generally, learnability of language has been the subject of many experiments, theories and debates. Experiments have shown for example that language acquisition is most successful when it is began early in life ([15]), which refers to the well-known concept of critical period ([13]). Also, learners of a second language typically have much more difficulties than learners of a first language ([9]). Until relatively recently, these facts were interpreted in favor of the idea that humans have an innate language acquisition device ([19]; [20]) which partly consists in pre-giving a number of linguistically specific constraint: for example, ([15]), argues that it is strong evidence for "maturationnaly scheduled **language specific** learning abilities". This view is also supported by a number of theoretical studies, like Gold's theorem ([10]), which basically states that in the absence of enough explicit negative evidence, one can not learn languages belonging to the superfinite class, which includes context free and context sensitive languages (but the applicability to human languages has been challenged, see ([8]).

Here we propose an alternative view, to which our model brings plausibility. It consists in explaining the fact that the learning skills of adults are lower than those of children by the fact that the brain resources needed to do so have already been recruited for other tasks or for a different language/sound system (see Rohde and Plaut, 1999 for a comparable view). Said another way, children are better to learn a completely new sound system than adults because their cognitive capabilities are less committed, whereas adults are already specialized. This is indeed what we observe in our model. To see that, a number of experiments were conducted in which on the one hand, some children agents had to learn a particular sound system, and on the other hand, adult agents had to learn a "second language" sound system. More precisely, in each experiment, first a society of agents was ran to produce a syllable system: after 15000 games, an agents was randomly chosen and called the teacher. This teacher was then used in the same game than described above, and with a second agent, the learner, except that here the teacher did not update its memory (he is supposed to know that he knows well the language as compared to the learner). The learner was each time in a first run a fresh agent (this models the child) and in a second run an agent taken from another society after 15000 games (which models an adult who knows already another sound system). This experiment was repeated 20 times. One example of success curve is on figure 1: the upper curve is the one for children learning success, and the lower curve for adults learning. Each point in the curve represents the mean success in last 100 games at a particular time t. The mean success after 5000 games of the 20 runs was of 97.3 percent for children against 80.8 percent for adults. This conform well to the idea of a critical period: adults never manage to learn perfectly another sound system. There is an explanation for that: whereas children start with a high plasticity in their inverse mapping memory (because they have no categories yet and so can freely delete and create many new items) and have no strong bias (in fact they are biased, as we will state in next paragraph, but not as much as adults) towards a particular zone of the syllable space, adults, on the contrary, are already committed to another sound system, and have more difficulties to create new items in the appropriate zone of the syllable space because their skills resources

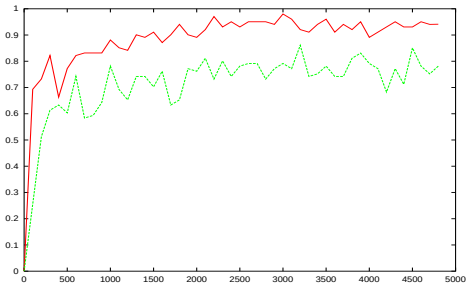


**Fig. 3.** Evolution of success in interactions during the learning of an established sound system: top curve is when agent is a child (fresh agent) and bottom curve when it is an adult (it already knows another sound system)

(which are items in inverse mapping memory that are not prototypes of one of their previous language categories) are much lower. Of course, some of these category prototypes may be pruned, and thus freeing some resources, because they are unsuccessful for the new sound system. But in practice it seems that enough of them allow successful imitations of items in the new sound system, though imprecise, so that still not enough resources can be freed to resolve the remaining confusions. To conclude this paragraph, we see that our model fits very well with the idea that critical periods/second language learning effect need not a genetically programmed language specific mechanism to find an explanation, and that the more parsimonious idea of (un-)commitment of the cognitive system can account for it.

Now, we saw that children could actually learn nearly perfectly a sound system. This result is not obvious since they are faced directly to the complete sound system, in the contrary of the agents who co-built it: the building was incremental and the sound system complexified progressively, which does not mean that their job was easier since negotiation had also to take place, but it was different. An experiment was performed that shows on the one hand how non-obvious the task is and on the other hand has implications over a number of existing theories. Children/fresh agents were put in a situation of trying to learn a random syllable system: the adult/teacher was artificially built by putting in its categorical repertoire items whose articulatory programs where completely random (chosen among the complete set of combinatorially possible less-than-5-phonemes articulatory programs). This experiment was repeated again 20 times. Figure 2 shows the curves of 2 experiments: the top one is for child learning success when the target language was generated by a population of agents and the bottom one for child learning success when the target language was random. The mean success over the 20 experiments after 5000 games is 97.3 percent for "natural" sound systems and 78.2 percent for random sound systems. We see that children never learn reasonably well the random sound systems. This result is experimentally and functionally very similar to an experiment about syntax described in (Christiansen, 2000), in which human subjects were asked to learn small languages whose syntax was either the one of an existing natural language or a random/artificial one. They found that indeed subjects were much better at

learning the language where the syntax was "natural" than the language where the syntax was "artificial". Deacon (1997) also made a point about this: "if language were a random set of associations, children would likely be significantly handicapped by their highly biased guessing".

This state of affair is in fact compatible with most of theories of language, which all basically suggest that human languages have many particular structures (that make them non-random) and that we are innately endowed with constraints that biases up us towards an easier learning of these languages, because they lead to the particular structure of languages. Now, where considerable disagreement comes in is again about the nature of these constraints and how they got there. On the one hand, the Chomskyan approach suggests that they are coded in a Universal Grammar genetically coded and linguistically specific, and consider language as a system mainly independent of its users (humans) who may have undergone biological evolution so as to be able to acquire and use it in an efficient way (this is suggested by [19]). This is not only true for syntax but also down to phonetics: this approach posits that we have an innate knowledge of what features (for example the labiality of a phoneme) and combination of features can be used in language ([5]). One of the problems with this approach is that the apparent "idiosyncrasies of language structure are hard to explain". On the other hand, a more recent approach considers that language itself evolved and its features were selected so as to fit to generic already existing learning and processing capabilities of humans (see for example ([7]), and that the coherent structures may have emerged through a process of self-organization at multiple levels ([24]). The fact that language evolved to fit to the primitive human brains ecological niche, and in particular to the brains of children, explains, as Deacon ([8]) puts it, why "children have an uncanny ability to make lucky guesses" though they do not possess innate linguistic knowledge. Again the present model tends to bring more plausibility to the second approach. Indeed, it is clear here that on the one hand innate generic motor, perceptual and cognitive constraints bias the way one explores and acquire parts of the syllable space, and on the other hand that the mechanism by which agents culturally negotiate which will be their particular sound system makes them select preferentially systems which allow easy imitation, hence easier learning. For instance, syllables that are very sensitive to noise will tend to be avoided/pruned since they lead to confusions. Also, syllable systems will tend to be coherent both with the process of exploration by differentiation and the tendency to remember better easy items than difficult ones: given a part of a syllable system, the rest may be found quite easily by focusing the exploration on small variants of items of this part, and exploration is also made maximally efficient by focusing on easy parts.

## 6 Conclusion

We have presented an operational model of the origins of syllable systems whose particularity is the stress on embodiment and situatedness constraints or opportunities, which imply the avoidance of many shortcuts usually taken in the literature. It illustrates in details (and brings more plausibility) the theory which states that language originated in a cultural self-organized manner, taking as a starting point a set of generic non-linguistically specific learning, motor and perceptual capabilities. In addition to the demonstration of how an efficient

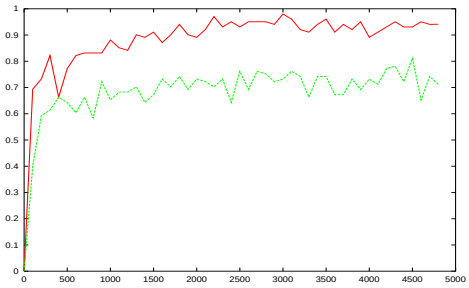


**Fig. 4.** Evolution of success in interaction during the learning of an established sound system by a child agent: top curve is when the sound system was generated with a population of agents with all constraints, bottom curve is when the sound system is completely random

communication system could be build with this parsimonious starting point and through cultural evolution, and to the fact that the produced sound systems had many structural similarities with human sound systems, we showed that the ability of children to learn sound systems so easily can be explained (contrarily to speculations of many Chomskyan researchers) by the evolutionary history of these sound systems, which were precisely shaped so as to fit the ecological niche formed by the brains and bodies of children, and not the other way around (as advocated by Chomskyan approaches to language). Yet, one has to note that we do not exclude that biological evolution driven by the need to adapt to a linguistic environment took a role; in fact it is very probable that genes (in particular those implicated in the development of the neural system) co-evolved with language, but, as Deacon puts it "languages have done most of the adapting".

## 7 References

## References

1. Archangeli D., Langendoen T. Optimality theory, an overview, Blackwell Pulishers (1997).
2. Arkin, R. Behavior-based Robotics, MIT Press (1999).
3. Bizzi E., Mussa-Ivaldi F., Giszter S.Computations underlying the execution of movement: a biological perspective, Science, vol. 253, pp. 287-291 (1991).
4. de Boer, B. Investigating the Emergence of Speech Sounds. In: Dean, T. (ed.) Proceedings of IJCAI 99. Morgan Kauffman, San Francisco. pp. 364-369 (1999).
5. Chomsky, N. and M. Halle (1968) The Sound Pattern of English. Harper Row, New york.
6. P. R. Cook, "Synthesis of the Singing Voice Using a Physically Parameterized Model of the Human Vocal Tract," Proc. of the International Computer Music Conference, pp. 69-72, Columbus, OH, 1989.
7. Christiansen, M., Using artificial language learning to study language evolution: Exploring the emergence of word order universals, in Language Evolution, Dessalles, Wray, Knight (eds.), Transitions to language, Oxford, Oxford University Press (2000).

8. Deacon T., The symbolic species, The Penguin Press (1997).
9. Flege J., Speech learning in a second language, In Ferguson, Menn, Stoel-Gammon (eds.) Phonological Development: Models, Research, Implications, York Press, Timonnium, MD, pp. 565-604 (1992).
10. Gold, E. Language identification in the limit. Information and Control 10, 447-474 (1967).
11. Hurford, J., Studdert-Kennedy M., Knight C., Approaches to the evolution of language, Cambridge, Cambridge University Press (1998).
12. Kirby, S., Syntax without natural selection: how compositionnality emerges from vocabulary in a population of learners, in Hurford, J., Studdert-Kennedy M., Knight C. (eds.), Approaches to the evolution of language, Cambridge, Cambridge University Press (1998).
13. Lenneberg, E. Biological foundations of language, New-york: Wiley (1967).
14. Lindblom, B., Phonological Units as Adaptive Emergents of Lexical Development, in Ferguson, Menn, Stoel-Gammon (eds.) Phonological Development: Models, Research, Implications, York Press, Timonnium, MD, pp. 565-604, (1992).
15. Long M. Maturational Constraints on Language Development, Studies in Second Language Acquisition 12, 251-285 (1990).
16. Lyon, R., All pole models of auditory filtering, in Lewis et al. (eds.) Diversity in auditory mechanics, World Scientific Publishing, Singapore (1997).
17. Massaro, D., Perceiving talking faces, MIT Press (1998).
18. MacNeilage, P.F., The Frame/Content theory of evolution of speech production. *Behavioral and Brain Sciences*, 21, 499-548 (1998).
19. Pinker, S., Bloom P., Natural Language and Natural Selection, The Brain and Behavioral Sciences, 13, pp. 707-784 (1990).
20. Piattelli-Palmarini, M., Evolution, selection and cognition: from "learning" to parameter setting in biology and in the study of language, Cognition, 31, 1-44 (1989).
21. Redford, M.A., C. Chen, and R. Miikkulainen Modeling the Emergence of Syllable Systems. In: Proceedings of the Twentieth Annual Conference of the Cognitive Science Society. Erlabum Ass. Hillsdale (1998).
22. Sakoe H., Dynamic programming optimization for spoken word recognition, IEEE Transaction Acoustic., Speech, Signal Processing, vol. 26, pp. 263-266 (1982).
23. Segui, J., Dupoux E., Mehler J. (1995) The role of the syllable in speech segmentation, phoneme identification, and lexical access, in Altman, (ed.), Cognitive Models of Speech Processing, Psycholinguistics and Computational Perspectives, MIT Press.
24. Steels, L., Synthesizing the origins of language and meaning using co-evolution, self-organization and level formation, in Hurford, Studdert-Kennedy, Knight (eds.), Cambridge University Press, pp. 384-404 (1998).
25. Oudeyer P-Y., The origins of syllable systems: an operational model, to appear in the Proceedings of The International Conference on Cognitive Science, COGSCI'2001, Edinburgh, Scotland, (2001).
26. Oudeyer P-Y, Coupled Neural Maps for the Origins of Vowel Systems, to appear in the proceedings of the International Conference on Artificial Neural Networks, ICANN'2001, Vienna, Austria, Springer Verlag (2001).
27. Steels L., Oudeyer P-y., The cultural evolution of syntactic constraints in phonology, in Bedau, McCaskill, Packard and Rasmussen (eds.), Proceedings of the 7th International Conference on Artificial Life, pp. 382-391, MIT Press (2000).