Language Learning Dynamics: Coexistence and Selection of Grammars

Valery Tereshko

School of Computing, University of Paisley, Paisley PA1 2BE, Scotland
valery.tereshko@paisley.ac.uk
http://cis.paisley.ac.uk/tere-ci0/

Abstract. Language learning dynamics is modelled by an ensemble of individuals consisting of the grammar carriers and the learners. Increasing the system population size results into the transition from the individual to the collective mode of learning. At low communication level, different grammars coexist in their own survival niches. Enhancement of the communication level in purely collective mode, when all individuals are the part of general communication network, leads to the selection of the fittest grammar. Adding the individual mode of learning results into the formation of the quasigrammar, with the dominant grammar prevailing over the set of coexisting grammars.

1 Introduction

A community of language users collectively developing a shared communication system can be viewed as the complex adaptive system subjected to a Darwinian evolution [1, 2, 3, 4, 5, 6, 7]. The language users are considered as the interacting agents. The central question is the phenomenon of spontaneous emergence of order in the ensembles of these agents. This order can be static or dynamic, i.e. represents stationary patterns or synchronized motions respectively, and most importantly, it appears without imposing any centralized control. Applying to a language, children are known to develop the grammatical competence by interactions with people, without any formal training [8]. This can be viewed as one of examples of the order that emerges entirely through individual's interactions.

In this paper, we consider language learning dynamics as a pattern formation phenomenon in a space of all available grammars, the grammar space. All individuals are divided into the grammar carriers, who are already learned and carry a particular grammar, and the learners who are not carry any grammar yet but potentially can learn one and become the grammar carriers themselves. Since the learners can interact with the carriers of different grammars, they can choose between the latter. On the other hand, the grammar carriers are competing for the pool of learners in the attempts to persuade the latter to their own grammars. The similar approach was used to model the molecular evolution [9] and the honey bee colony foraging dynamics [10, 11, 12].

We introduce the individual and the collective (through communication with other individuals) modes of learning. Increasing the population size results into

F. Almeida e Costa et al. (Eds.): ECAL 2007, LNAI 4648, pp. 415–424, 2007.

the transition from the individual to the collective learning characterizing by the sharp increase in the grammar carriers.

When the grammar carriers can access the learners only locally, different grammars coexist. Unlike, in purely collective mode, when the grammar carriers have access to the whole pool of learners, the only fittest grammar survives. When taking the individual mode of learning into account, the quasigrammar is developed. The latter is characterized by the formation of the dominant grammar prevailing over the set of coexisting grammars.

Unlike other models using the population dynamics approach to model an evolution of language [5, 6, 7], our model introduces the real space, allowing to consider not only the temporal but the spatio-temporal dynamics, and the individual mode of learning. The latter give extra modelling opportunities to produce more realistic outcomes. Indeed, in [7] the coexistence of grammars, or so called m-grammar solution, can be reached only for the case of equal fitness, which is unrealistic situation. Our model allows the coexistence of the different fitness grammars.

2 Basic Model

Let us describe the language dynamics mechanism in the terms of chemical reactions. The grammar carriers and the individuals who are not carry any particular grammar yet, i.e. the learners, are denoted by X and Y respectively; the grammar fitness by f:

$$Y + X \to 2X \tag{1}$$

Reaction (1) illustrates the autocatalytic nature of the communication process. If a carrier communicates its grammar to an individual involving the latter to it, this individual (as the new carrier) will in turn reinforce the process and communicate the chosen grammar to other individuals, and so forth.

A carrier can abandon an unrewarding grammar at the rate inversely proportional to that grammar's quality:

$$X \xrightarrow{\frac{1}{f}} Y \tag{2}$$

The kinetic equations corresponding to reactions (1-2) take the form

$$\begin{aligned}
\frac{\partial x(\mathbf{r},t)}{\partial t} &= \Big(y(\mathbf{r},t) - \frac{1}{f(\mathbf{r},t)}\Big)x(\mathbf{r},t) + D_x \frac{\partial^2 x(\mathbf{r},t)}{\partial \mathbf{r}^2} \\
\frac{\partial y(\mathbf{r},t)}{\partial t} &= -\Big(y(\mathbf{r},t) - \frac{1}{f(\mathbf{r},t)}\Big)x(\mathbf{r},t) + D_y \frac{\partial^2 y(\mathbf{r},t)}{\partial \mathbf{r}^2},
\end{aligned} \tag{3}$$

where x and y are the concentrations, and D_x and D_y are the diffusion constants of the grammar carriers and learners respectively, and $f(\mathbf{r},t)$ is the grammar landscape. Taking the Wright's idea of fitness landscape which assigns a fitness to each point in a genetic space, our grammar landscape gives a grammar quality value to each point in the grammar space.

3 Dynamics at Different Communication Levels

The grammar carrier diffusion is assumed to be small that means high accuracy of the grammar acquisition. The diffusion of learners determines their access to different grammars. Let us to compare two opposite cases: (i) the diffusion of Y is slow allowing the access of learners to local information only, and (ii) the diffusion of Y is high enough to ensure the global mixing of learners throughout the entire grammar space.

(i) small D_y

In the limit of vanishing diffusions, system (3) possesses integral of motion

$$\frac{\partial x(\mathbf{r},t)}{\partial t} + \frac{\partial y(\mathbf{r},t)}{\partial t} = 0, \tag{4}$$

which yields condition of constant *local* concentrations

$$x(\mathbf{r},t) + y(\mathbf{r},t) = x_0(\mathbf{r}) + y_0(\mathbf{r}) = C(\mathbf{r}). \tag{5}$$

This allows to eliminate variable y from system (3) and to reduce the latter to spatially extended logistic equation

$$\frac{\partial x(\mathbf{r},t)}{\partial t} = \big(\alpha(\mathbf{r}) - x(\mathbf{r},t)\big)x(\mathbf{r},t), \tag{6}$$

where

$$\alpha(\mathbf{r}) = C(\mathbf{r}) - \frac{1}{f(\mathbf{r})}. \tag{7}$$

The grammar acquisition threshold, $1/C(\mathbf{r})$, indicates the grammars that are attractive for the individuals. The latter begin to learn only those grammars whose fitness value exceeds threshold

$$f(\mathbf{r}) > 1/C(\mathbf{r}). \tag{8}$$

The problem nature allows us to take into account the set of n spatial modes corresponding to the local maxima of the grammar landscape. Considering only these modes, infinite-dimensional system (6) reduces to system of uncoupled equations describing the logistic growth of the grammar carriers at i-th spatial point

$$\dot{x}_i(t) = (\alpha_i - x_i(t))x_i(t), \tag{9}$$

where

$$\alpha_i = C_i - \frac{1}{f_i} \tag{10}$$

is the reproductive rate of i-th mode.

Every mode associated with a fitness exceeding the grammar acquisition threshold converges to attractor

$$x_i^s = C_i - \frac{1}{f_i},\ i = 1,..,n. \tag{11}$$

When the learners have access to only local information, different grammars have ability to *coexist* in the society.

(ii) large D_y
Let us consider now the opposite case when the learners have access to all available grammars. In the limit of Y's full mixing, system (3) reduces to system of integro-differential equations

$$\begin{aligned}\frac{\partial x(\mathbf{r},t)}{\partial t} &= \Big(y(t) - \frac{1}{f(\mathbf{r})}\Big)x(\mathbf{r},t) + D_x\frac{\partial^2 x(\mathbf{r},t)}{\partial \mathbf{r}^2}\\ \frac{\partial y(t)}{\partial t} &= -y(t)\frac{1}{S}\int_Q x(\mathbf{r},t)\,\mathrm{d}\mathbf{r} + \frac{1}{S}\int_Q \frac{1}{f(\mathbf{r})}x(\mathbf{r},t)\,\mathrm{d}\mathbf{r},\end{aligned} \tag{12}$$

where $y(t) = (1/S)\int_Q y(\mathbf{r},t)\,\mathrm{d}\mathbf{r}$ is the spatially-averaged concentration of carriers over domain Q with area $S = \int_Q \mathrm{d}\mathbf{r}$.

Integral of motion[1]

$$\frac{1}{S}\int_Q \frac{\partial x(\mathbf{r},t)}{\partial t}\,\mathrm{d}\mathbf{r} + \frac{\partial y(t)}{\partial t} = 0 \tag{13}$$

yields condition of constant *total* concentration

$$\frac{1}{S}\int_Q x(\mathbf{r},t)\,\mathrm{d}\mathbf{r} + y(t) = x_0 + y_0 = C, \tag{14}$$

which allows to eliminate variable y from system (12) and to reduces the latter to spatially extended Lotka-Volterra equation

$$\frac{\partial x(\mathbf{r},t)}{\partial t} = [\alpha(\mathbf{r}) - \frac{1}{S}\int_Q x(\mathbf{r},t)\,\mathrm{d}\mathbf{r}]x(\mathbf{r},t) + D_x\frac{\partial^2 x(\mathbf{r},t)}{\partial \mathbf{r}^2}. \tag{15}$$

In the limit of vanishing D_x, if only modes corresponding to the local maxima of the grammar landscape are taken into account, infinite-dimensional system (15) reduces to system of coupled equations for the spatial mode amplitudes

$$\dot{x}_i(t) = (\alpha_i - \sum_{i=1}^{n} x_i(t))x_i(t). \tag{16}$$

Dividing i-th and j-th equations on x_i and x_j respectively and subtracting one equation from another, one obtains

$$\frac{\dot{x}_i(t)}{x_i(t)} - \frac{\dot{x}_j(t)}{x_j(t)} = \frac{1}{f_j} - \frac{1}{f_i} \tag{17}$$

The integration of equation (17) results into

$$\frac{x_i(t)}{x_j(t)} = \frac{x_i(0)}{x_j(0)}\exp\left(\Big(\frac{1}{f_j} - \frac{1}{f_i}\Big)t\right). \tag{18}$$

---

[1] The integration over the space eliminates the diffusion term in the first equation of system (12) due to no-flux boundary conditions.

Expression (18) provides the analytical proof of the selection in the system. If m-th mode is the fittest and the unique, then $f_m > f_j$ for $\forall j \neq m, j = 1, .., n$. Hence, it immediately follows that when $t \to \infty, x_m/x_j \to \infty$ for $\forall j \neq m, j = 1, .., n$. However, condition of constant total concentration (14) and the positive definiteness of variables prevent the unlimited growth of the modes. This means that the amplitudes of all modes, excluding the fittest one, must tend to zero with time. If more than one mode are the fittest, then they all survive.

If the fitness of at least one mode exceeds the grammar acquisition threshold, then trivial equilibrium

$$x_i^s = 0, \; i = 1, .., n \tag{19}$$

loses stability and system (16) converges to non-trivial attractor

$$x_m^s = C_0 - \frac{1}{f_m}, \; x_i^s = 0, \; i = 1, .., n; \; i \neq m \tag{20}$$

where $f_m > f_i$, which corresponds to the *selection* of the fittest grammar.

The society of globally informed individuals is, thus, capable of the collective choice of the fittest grammar.

Let us perform the numerical simulations. We use the explicit method of the numerical integration of PDEs when space and time are divided into discrete uniform sub-intervals, and derivatives are replaced by their finite-difference approximations. The numerical integrations are performed on the 2D lattice with the space and the time steps are chosen to guarantee the stability and the convergence of explicit scheme. Throughout, the initial concentrations of grammar carriers and their diffusion constant are taken to be $x_0 = 0.01$ and $D_x = 0.01$ respectively.

Consider the grammar fitness landscape with three spatially separated regions (Fig. 1a). Two regions consist of two fitness peaks each, and one region consists of only one peak. Among the peaks, we have the highest one, two smallest ones, and two intermediate ones of the same height.

The diffusion length of learners in the grammar space, l_d, can be evaluated [13]:

$$l_d = \sqrt{D_y \tau_d}, \tag{21}$$

where characteristic diffusion time τ_d is approximated as

$$\tau_d \sim \frac{1}{\alpha(\mathbf{r})} = \left(C_0 - \frac{1}{f(\mathbf{r})}\right)^{-1}. \tag{22}$$

For the areas where $f(\mathbf{r}) \to 0$, τ_d and, hence, $l_d \to 0$. The latter means that for the weak diffusion of learners and the distantly separated grammars, the grammar carriers get their own exclusive resource for the development of their grammars. In others words, the grammar niches are developed, which leads to the coexistence of the different grammars in the system. Figure 1b illustrates the above arguments. Note that the weakest grammars corresponding to the smallest peaks in every niche get suppressed. This happens because the diffusion length becomes compared to the niche size. The grammars inside the niches compete

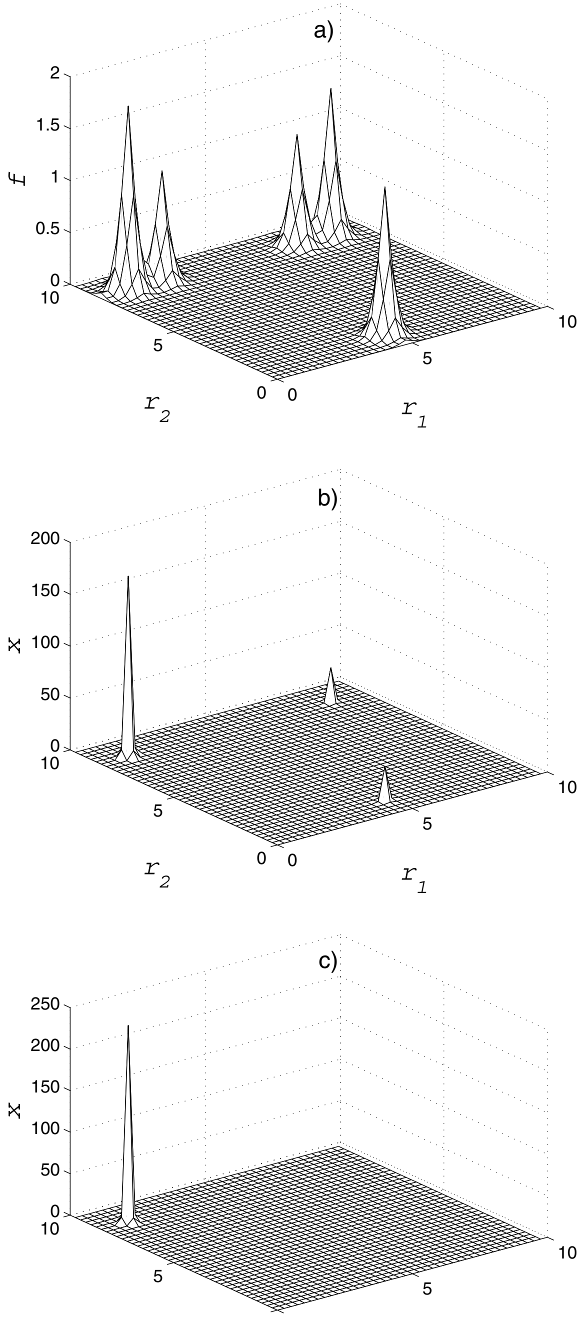

Fig. 1. a) Fitness landscape, b)-c) Concentration of grammar carries for system (3) at $D_y = 1$, $t = 3000$ and $D_y = 10$, $t = 75$ respectively.

for their common resource, y. As result, the niche's strongest grammars survive. Note that the above processes become completed at about $t = 300$, and the landscape similar to one in Fig. 1b is formed. Performing simulations up to $t = 3000$ gives the same picture, which ensures us that the grammar coexistence is stable. Increasing D_y creates a common communication niche throughout the grammar space, which leads to the selection of the fittest grammar in the system (see Fig. 1c).

4 Transition from Individual to Collective Learning and Formation of Quasigrammar

Model (1)-(2)accounts for the learners acquiring their grammars through the communication process with the grammar's carriers. Let us now to account for these who learns individually.

The following reaction represents the individual learning:

$$Y \xrightarrow{\epsilon} X_i. \tag{23}$$

The amount of individual learners are assumed to be much smaller that these learning through the communication, meaning ϵ is set to be small.

The kinetic equations of the updated model are

$$\begin{aligned}
\frac{\partial x(\mathbf{r},t)}{\partial t} &= \Big(y(\mathbf{r},t) - \frac{1}{f(\mathbf{r},t)}\Big)x(\mathbf{r},t) + \epsilon y(\mathbf{r},t) + D_x \frac{\partial^2 x(\mathbf{r},t)}{\partial \mathbf{r}^2} \\
\frac{\partial y(\mathbf{r},t)}{\partial t} &= -\Big(y(\mathbf{r},t) - \frac{1}{f(\mathbf{r},t)}\Big)x(\mathbf{r},t) - \epsilon y(\mathbf{r},t) + D_y \frac{\partial^2 y(\mathbf{r},t)}{\partial \mathbf{r}^2},
\end{aligned} \tag{24}$$

In the limit of vanishing D_x and high D_y, model 24 reads

$$\dot{x}_i = \Big(C - \frac{1}{f_i}\Big)x_i - \epsilon \sum_{k=1}^{n} x_k - x_i \sum_{k=1}^{n} x_k + \epsilon C. \tag{25}$$

Let us consider the simplest case of a single grammar and analyze the system dynamics depending on its size. In this case, the learners dynamics is described by simple logistic equation with small constant-growth term

$$\dot{x} = \Big(C - \frac{1}{f} - \epsilon\Big)x - x^2 + \epsilon C. \tag{26}$$

The only physical attractor the system converges to is

$$x^* = \frac{C - \frac{1}{f} - \epsilon}{2} + \sqrt{\frac{(C - \frac{1}{f} - \epsilon)^2}{4} + \epsilon C} \tag{27}$$

It is easy to see that x^* remains low up to the point $C = \frac{1}{f} + \epsilon$. Above this point, x^* sharply increases and eventually tends to C (at higher values of the latter). Fig. 2 illustrates this.

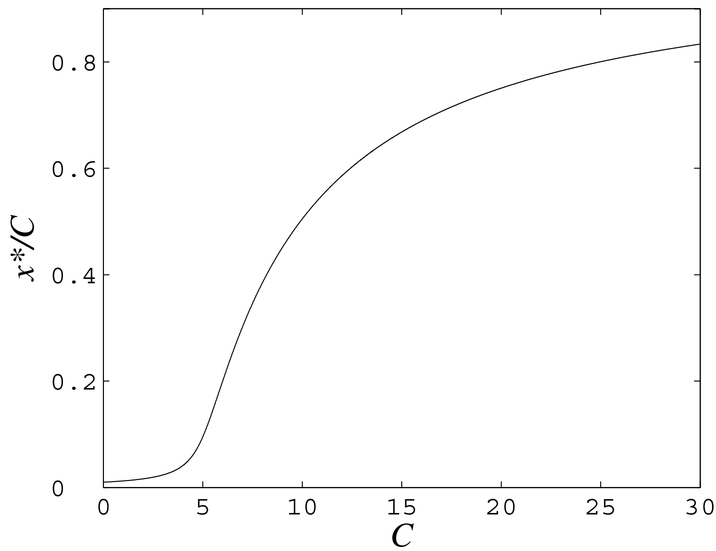

Fig. 2. Proportion of the stationary concentration of grammar carriers to the total population versus the population size.

The above figure elucidates the transition from the *individual* to the *collective* learning. Indeed, when $C < \frac{1}{f} + \epsilon$, the grammar carriers concentration growths entirely due to the free term of equation (26), i.e. due to the individual learning. Unlike, at $C > \frac{1}{f} + \epsilon$, the communication is the mechanism governing the learning dynamics. Note that we still have an increase of the grammar carriers caused by the individual learning. However, the dominant term contributing to the above increase, $(C - \frac{1}{f}\epsilon)x$, is determined by the communication.

In general case of n grammars, define the (total) grammar carriers flow as $\sum x_i$. Then, the flow dynamics is governed by equation

$$\sum_{i=1}^{n} \dot{x}_i = C \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \frac{1}{f_i} x_i - \epsilon \sum_{i=1}^{n} x_i - \Big(\sum_{i=1}^{n} x_i \Big)^2 + \epsilon C. \tag{28}$$

Taking into account that

$$\sum_{i=1}^{n} \frac{1}{f_i} x_i = \frac{\sum_{i=1}^{n} \frac{1}{f_i} x_i}{\sum_{i=1}^{n} x_i} \sum_{i=1}^{n} x_i = \Big\langle \frac{1}{f} \Big\rangle \sum_{i=1}^{n} x_i, \tag{29}$$

where $\langle \frac{1}{f} \rangle$ is the averaged (over the set of grammars) grammar quality, obtain:

$$\sum_{i=1}^{n} \dot{x}_i = \Big(C - \Big\langle \frac{1}{f} \Big\rangle - \epsilon \Big) \sum_{i=1}^{n} x_i - \Big(\sum_{i=1}^{n} x_i \Big)^2 + \epsilon C. \tag{30}$$

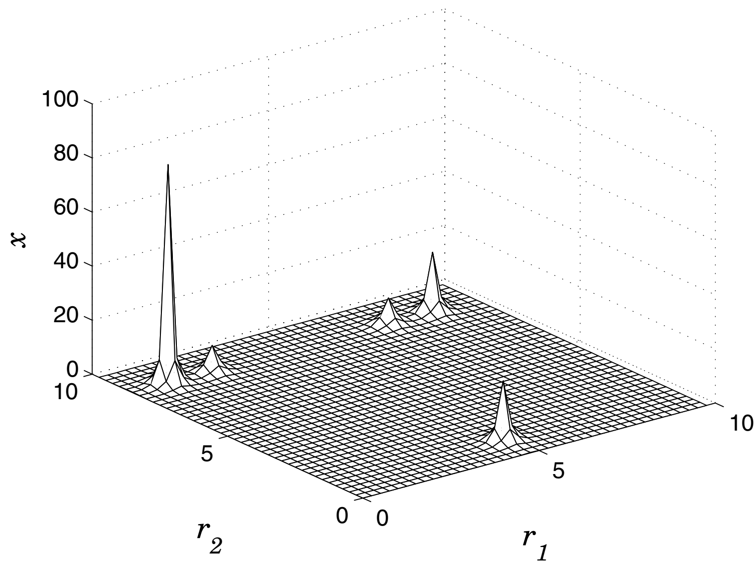

Fig. 3. Concentration of grammar carries for system (24) at $D_y = 10$, $t = 300$.

Obviously, all previous results obtained for the single grammar case are held here, and at $C = \left\langle \frac{1}{f} \right\rangle + \epsilon$ the population undergoes the individual-collective learning transition.

Let us look at the grammar fitness landscape in Fig. 1a. As we have already seen, at high communication level, the common communication niche throughout the whole system is created. In this case, the basic model dynamics leads to the survival of the fittest grammar in the system (Fig. 1c). Unlike, system (24) doesn't produce the pure selection: one observes the formation of the dominant grammar that prevails over all other coexisting grammars (Fig. 3). Taking the idea of quasispecies [14], we call the above the *quasigrammar.*

5 Conclusion

Our approach elucidates the natural selection that created the human's system of communication. The grammars are like species in competition. The fitness of the species is given by the grammar quality. A grammar can survive by continuing to circulate within the society, and is able to reproduce itself by recruiting new carriers who has learned and share it with others.

Increasing the population size results into the transition from the individual to the collective learning. The sharp increase in the grammar carriers demonstrates greater efficiency of the collective mode.

The level of spatial interactions determines the system behaviour. At low level, the learners are "locked" on their local knowledge sites. Different grammars can coexist in their own survival niches. The enhancement of the interaction level creates the global niche excluding the above scenario. In this case we have survival of the fittest grammar. In cultural sense it means simplification, and we can wonder whether such a globalization will be beneficial for the system.

At high level of communication in the system, the pure selection happens only in the case of purely collective learning. One can say that this is the case when all individuals are the part of general communication network. Taking into account the individual learning results into the formation of the quasigrammar.

References

1. Steels, L.: Language as a complex adaptive system. In: Deb, K., Rudolph, G., Lutton, E., Merelo, J.J., Schoenauer, M., Schwefel, H.-P., Yao, X. (eds.) Parallel Problem Solving from Nature-PPSN VI. LNCS, vol. 1917, pp. 17–26. Springer, Heidelberg (2000)
2. Steels, L.: The evolution of communication systems by adaptive agents. In: Alonso, E., Kudenko, D., Kazakov, D. (eds.) Adaptive Agents and Multi-Agent Systems. LNCS (LNAI), vol. 2636, pp. 125–140. Springer, Heidelberg (2003)
3. Kirby, S.: The evolution of language. In: Dunbar, R., Barrett, L. (eds.) Oxford Handbook of Evolutionary Psychology, pp. 669–681. Oxford University Press, Oxford (2007)
4. Oudeyer, P-Y., Kaplan, F.: Language evolution as a darwinian process: Computational studies. Cognitive Processing 8, 21–35 (2007)
5. Komarova, N.L., Niyogi, P., Nowak, M.A.: The evolutionary dynamics of grammar acquisition. J. Theor. Biol. 209, 43–59 (2001)
6. Nowak, M.A., Komarova, N.L., Niyogi, P.: Evolution of universal grammar. Science 291, 114–118 (2001)
7. Komarova, N.L.: Replicator-mutator equation, universality property and population dynamics of learning. J. Theor. Biol. 230, 227–239 (2004)
8. Jackendoff, R.S.: The architecture of the language faculty. MIT Press, Camridge (1977)
9. Tereshko, V.: Selection and coexistence by reaction-diffusion dynamics in fitness landscapes. Phys. Lett. A 260, 522–527 (1999)
10. Tereshko, V.: Reaction-diffusion model of a honeybee colony's foraging behaviour. In: Deb, K., Rudolph, G., Lutton, E., Merelo, J.J., Schoenauer, M., Schwefel, H.-P., Yao, X. (eds.) Parallel Problem Solving from Nature-PPSN VI. LNCS, vol. 1917, pp. 807–816. Springer, Heidelberg (2000)
11. Tereshko, V., Lee, T.: How information-mapping patterns determine foraging behaviour of a honey bee colony. Open Systems and Information Dynamics 9, 181–193 (2002)
12. Tereshko, V., Loengarov, A.: Collective decision making in honey-bee foraging dynamics. Computing and Information System Journal 9, 1–7 (2005)
13. Mikhailov, A.S.: Foundations of Synergetics I. Springer, Berlin (1990)
14. Eigen, M., Schuster, P.: The Hypercycle: A Principle of Natural Self-Organization. Springer, Berlin (1979)