Cogn Comput (2016) 8:462–466
DOI 10.1007/s12559-015-9371-7

Automata Networks for Memory Loss Effects in the Formation of Linguistic Conventions

Javier Vera[1] · **Eric Goles**[1]

Received: 14 July 2015 / Accepted: 20 November 2015 / Published online: 16 December 2015

Abstract This work attempts to give new theoretical insights into the absence of intermediate stages in the evolution of language. In particular, a mathematical model, based on automata networks, is proposed with the purpose to answer a crucial question: How a population of language users can reach agreement on linguistic conventions? To describe the appearance of drastic transitions in the development of language, an extremely simple model of working memory is adopted: at each time step, language users simply lose part of their word memories according to a forgetfulness parameter. Through computer simulations on low-dimensional lattices, sharp transitions at critical values of the parameter are described.

Keywords Automata networks · Linguistic conventions · Working memory · Sharp transition

Introduction

Contrarily to the extended view on language evolution [3, 7] which proposes a gradual transition through successive stages between a "protolanguage" (in simple terms a modern language minus syntax) and modern languages, recent works have suggested the absence of intermediate stages. For instance, in [5] it is suggested the appearance of phase transitions (scaling relations close to the *Zipf's law*) in the emergence of vocabularies under *least effort* constraints.

This work attempts to give new theoretical insights into the absence of intermediate stages in the evolution of language. The startpoint is a crucial question: How a population of language users can reach agreement on a linguistic convention? [2, 9–11]. Surprisingly, language users collectively reach shared communication systems without any kind of central control influencing the formation of language, and only from *local* conversations between few participants. The solution is partly based on two opposite *alignment* preferences, which guide the behavior of language users by the selection of the words that give the highest chance of communicative success and the removal of the words not involved in agreement [9, 10]. As a natural consequence, these strategies will lead to the self-organization of shared conventions.

To describe the appearance of drastic transitions in language formation, an extremely simple model of working memory is adopted [1], understood as a temporal finite memory involved in online tasks and, specially, in language production and comprehension. At each time step, language users simply lose part of their word memories according to a predefined forgetfulness parameter. What is more, it is hypothesized that the features of language (in particular, the consensus on linguistic conventions) emerge drastically at some critical memory loss capacity [6].

This work proposes a mathematical model based on automata networks [12, 14]. Automata networks are attractive models for systems that exhibit self-organization. From extreme simplified rules of local interactions inspired in real phenomena, automata networks exhibit astonishingly rich patterns of behavior. The essential feature of the adopted framework is *locality*: only from local

✉ Javier Vera
jxvera@gmail.com

Eric Goles
eric.chacc@uai.cl

[1] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Avda. Diagonal Las Torres 2640, Peñalolén, Santiago, Chile

communicative interactions, the emergence of complex language patterns will be described.

The work proceeds by introducing basic definitions and the rules of the automata ("The Model" section). This is followed in "Methods" and "Sharp Transitions on Two-Dimensional Lattices" sections by experiments based on an energy function that measures the amount of local agreement between individuals. Finally, a brief discussion about sharp transitions in language formation is presented.

The Model

Basic Notions

$\mathcal{G} = (P, I)$ is a connected and undirected graph with vertex set $P = \{1, \ldots, n\}$ and edge set I. The set P represents the finite population of individuals, whereas I is the set of possible interactions between individuals. This work indistinctly speaks of "individual" and "vertex". A crucial element of the model is that the interactions (defined by I) are *local*. The individual $u \in P$ participates in communicative interactions only with "nearby" neighbors. A positive integer parameter r (the *radius*) is introduced. The *neighborhood* of radius r of u is the set $V_u^r = \{v \in P : 0 < d(u, v) \leqslant r\}$, where d is the usual distance on $\mathcal{G}$ (the length of the shortest path between two vertices). Communicative interactions occur then between an individual $u \in P$ and its associated set of neighbors located on V_u^r.

W is a finite set of words. Each individual $u \in P$ is characterized by its *state* pair (M_u, x_u), where M_u is the memory to store words and $x_u \in M_u$ is a word that u conveys to the neighbors of V_u^r. In this context, within a communicative interaction (between a vertex and its neighbors), the "central" vertex plays the role of "hearer" and the neighbors play the role of "speaker". Indeed, the central vertex receives the words conveyed by its neighbors. This set of conveyed words is called W_u, $u \in P$. Some conveyed words are known, and some of these words are unknown by the central vertex u. Two sets are defined: $B_u = \{x \in W_u : x \in M_u\}$, the set of *known* words, and $N_u = \{x \in W_u : x \notin M_u\}$, the set of *unknown* words.

Automata Networks

On the graph $\mathcal{G}$, the **naming automata** is defined as the tuple $\mathcal{A} = (\mathcal{G}, Q, (f_u : u \in P), \phi)$, where

- Q is the set of all possible states $\mathcal{P}(W) \times W$ ($\mathcal{P}$ denotes the set of subsets of W). So, the state associated with the vertex $u \in P$, (M_u, x_u), is an element of Q.
- $(f_u : u \in P)$ is the set of *local rules*. Each cell is associated with the same local rule. Roughly speaking, the rule takes as inputs the words of W_u (in particular, the sets B_u and N_u) and it gives as output the new state of the vertex u.
- ϕ is a function, the *updating scheme*, that defines the order in which the vertices are updated. Traditionally, automata networks suppose the existence of a global "clock" that establishes that all vertices are updated at the same time. In this work, a *fully asynchronous* scheme is considered. This updating scheme implies that at each time step, one vertex is selected uniformly at random. The purpose of considering a fully asynchronous scheme arises from the typical updating order of the Naming Game: at each time step, two vertices (speaker and hearer) are chosen at random.

The configuration $X(t)$ at time step t is the set $\{(M_u, x_u)\}_{u \in P}$, where each individual $u \in P$ is associated with a state pair (M_u, x_u). The vertex $u \in P$ is chosen according to the fully asynchronous scheme. The dynamics evolves in a simple way: the configuration at step $t + 1$, $X(t + 1)$, is obtained by updating through the local rule f_u the state of the vertex u. The configuration X' is a *fixed point* of the dynamics if $X'(t) = X'(t + 1)$ for any vertex update.

Local Rules

The local rules $(f_u : u \in P)$ are based on the alignment strategies of the Naming Game [2]. Suppose that at time step t the vertex u has been selected. u and its neighbors of V_u^r define a communicative interaction, in which the vertex u plays the role of "hearer" and the neighbors of V_u^r play the role of "speaker". The vertex u faces two possible actions: (1) M_u is updated by adding the words of N_u [**addition (A)**] in order to increase the chance of future successful interactions; or (2) M_u is updated by eliminating the words [**collapse (C)**] that do not participate in agreement. Here, a particular collapse action is considered. Suppose that each agent is endowed with an internal total order for the set of words (equivalently, if $W \subseteq \mathbb{Z}$ the agents are endowed with the order $<$). Every agent chooses to collapse in the minimum word presented in the neighborhood. This rule represents, for example, the situation that the words differ according to their degree of relevance related to specific linguistic contexts [13].

To measure the amount of memory loss, a third action, **forgetfulness (F)**, is introduced. Let $p \in [0, 1]$ be a parameter. In simple terms, to the extent that p increases, the amount of *memory loss* increases. P_u is the subset of $M_u \setminus \{x_u\}$[1] formed by $\lfloor p(|M_u| - 1) \rfloor$[2] words, selected at

[1] $M_u \setminus \{x_u\}$ denotes the set M_u without the element x_u.

[2] $\lfloor p(|M_u| - 1) \rfloor$ means the largest integer lower than $p(|M_u| - 1)$.

random without replacement from $M_u \setminus \{x_u\}$. Then, the local rules read

$$f_u = \begin{cases} \text{if } \emptyset \neq \mathrm{N_u}, & \begin{cases} (\mathbf{F})(M_u \setminus P_u, x_u) \\ (\mathbf{A})(M_u \cup N_u, x_u) \end{cases} \\ \text{if } \emptyset = \mathrm{N_u}, & (\mathbf{C})(\{\min(B_u)\}, \min(B_u)) \end{cases} \tag{1}$$

In other words, in the case that $\emptyset \neq N_u$ the local rule acts following two steps: first, by the *forgetfulness* action and, second, by the *addition* action (along these two steps the set W_u do not change) (see Fig. 1).

Methods

To explicitly describe the amount of local agreement between individuals, a function, called the "energy," is defined (for a similar function, see [8]). This energy-based approach arises from a physicist interpretation. The energy measures the amount of local unstability of the configuration. Large values of energy imply the evolution until reach *ordered* low-energy configurations. At each neighborhood V_u^r, it is defined the function $\delta(x_u, x_v)$, $v \in V_u^r$, which is 1 in the case that $x_u = x_v$ (*agreement* between the vertices u and v), and 0 otherwise (*disagreement*). Thus, it measures the amount of local agreement of the neighborhood $\sum_{v \in V_u^r} \delta(x_u, x_v)$; summing this quantity over all vertices defines the total energy of the configuration at that time:

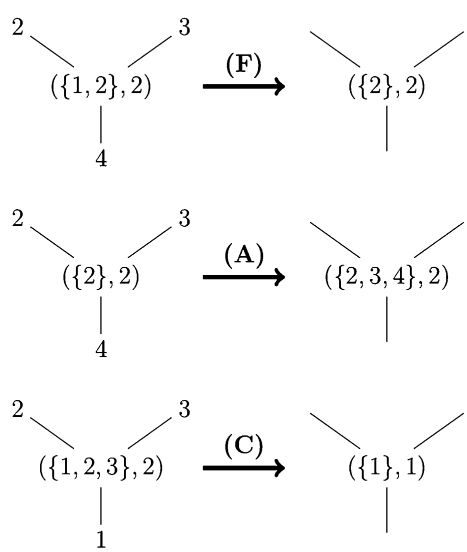

Fig. 1 Example of **forgetfulness (F)** and **addition (A)** actions. $W = \{1, 2, 3, 4\}$ is the set of words. Suppose that at some time step the central vertex ($u \in P$) has been chosen. Four individuals participate in the interaction: the central vertex u (the "hearer") and the three neighbors of V_u (the "speakers"). It is assumed that $p = 0.5$. In the *first row* (F), u "forgets" the word "1," and then $(\{1, 2\}, 2)$ is updated to $(\{2\}, 2)$. In the *second row* (A), $(\{2\}, 2)$ is updated to $(\{2, 3, 4\}, 2)$. In the *third row* (C), u cancels of the words in its memory, except the word "1". Then, $(\{1, 2, 3\}, 2)$ is updated to $(\{1\}, 1)$

$$E(t) = -\frac{1}{n} \sum_{u \in P} \frac{1}{|V_u^r|} \sum_{v \in V_u^r} \delta(x_u, x_v) \tag{2}$$

The function $E(t)$ is bounded by two extreme *agreement* cases: $E(t) = 0$ if each individual conveys a different word, and $E(t) = -1$ if all individuals convey the same word. Global agreement coincides with the final absorbing state of the Naming Game, where there is one unique shared word.

The analysis is focused on a two-dimensional periodic lattice of size $n = 128^2 = 16384$ with von Neumann neighborhood. The final value $E(t)$ (after $200n$ time steps or until reach $E(t) = -1$) is described for several values of p and r: p varies from 0 to 1 with an increment of 10%, and $r = \{1, 2, 3, 4\}$ (respectively, 4, 12, 24 and 40 neighbors). In general, a von Neumann neighborhood of radius r supposes $2r(r + 1)$ neighbors. Even though the radius $r = 4$

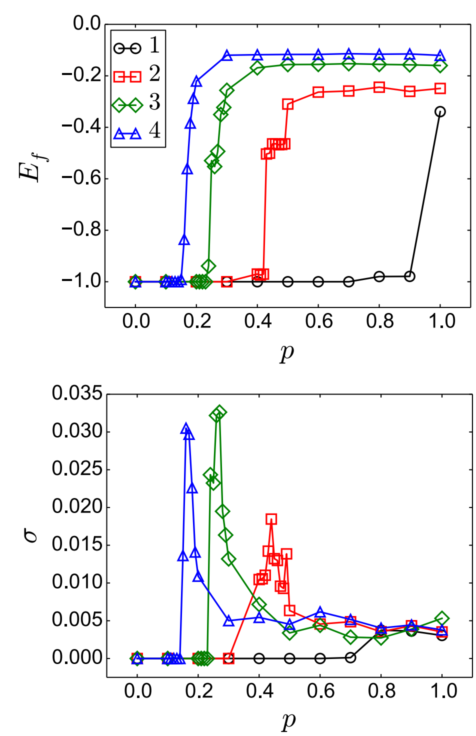

Fig. 2 E_f versus p on two-dimensional lattices. (*top*) On a two-dimensional lattice of size $n = 128^2$, it is showed the final value of the energy function, E_f, versus the parameter p, after $200n$ steps or until reach the global minimum $E(t) = -1$. Averages over 100 initial conditions ($r = 1, 2$) and ten initial conditions ($r = 3, 4$), with $|W| = n$. First, $r \in \{1, 2, 3, 4\}$ and p is varied from 0 to 1 with an increment of 10 %. With these parameters, $p_c^2 \in (0.4, 0.5)$, $p_c^3 \in (0.2, 0.3)$ and $p_c^4 \in (0.1, 0.2)$. New simulations run over the previous critical zones. Averages over ten initial conditions. At each critical zone, p is varied with an increment of 10 %. The critical parameter p_c^r, $r > 1$, is then defined as the lower value of p in the critical zone so that the energy function clearly does not converge to the global minimum $E(t) = -1$. (*bottom*) Standard deviation of the data for $r = 1, 2, 3, 4$ and all values of the parameter p

supposes 40 neighbors, there is no loss of *locality*. Indeed, $\frac{40}{n} \sim 2\,\%$ of the population of individuals.

Sharp Transitions on Two-Dimensional Lattices

Several aspects are remarkable in the behavior of E_f versus p, as shown in Fig. 2 (top). The parameters $r = 1$ and $p < 1$ imply that the dynamics reaches the global agreement configuration ($E(t) = -1$). For $r = 2, 3, 4$, the behavior of E_f versus p exhibits three clear domains. First, E_f reaches the minimum -1 for $p < p_c^r$. This critical parameter depends on r: $p_c^2 \approx 0.43$, $p_c^3 \approx 0.25$, $p_c^4 \approx 0.17$. In general, it is noticed that an increasing in the radius $r > 1$ implies a decreasing in the critical parameter p_c^r. Second, a drastic change is found at $p = p_c^r$. It is lost the convergence to the global minimum $E_f = -1$. Finally, for $p > p_c^r$ the dynamics seems to reach a stationary value $E_f > -1$ which increases to the extent r grows ($E_f \to 0$).

Standard deviation σ of the data versus p, as shown in Fig. 2 (bottom), confirms the previous observations. Three aspects are remarkable. First, σ takes small values in all cases (< 0.035). Second, for $p < p_c^r$, $r > 1$, the standard deviation is close to 0. Third, it is observed a peak in σ close to the critical parameters p_c^r, $r > 1$.

Discussion

The sudden changes observed on two-dimensional lattices, the peaks in standard deviation, as shown in Fig. 2, and the presence of power laws (Fig. 3) suggest the appearance of sharp (phase) transitions at $p = p_c^r$, for $r > 1$ [4, 5]. It was noticed that $r = 4$ exhibits the most drastic sharp transition, at $p_c^r \approx 0.17$. Critical forgetfulness parameters are strongly related to scaling relations, as shown in Fig. 3: low words (the first-ranked ones) are associated with multiple individuals, whereas several words are related to one-to-one individual-word associations. Despite of the fact that different radius, $r > 1$, exhibit similar slopes, the scaling relations for $r = 4$ differ from $r = 2, 3$, as shown in Fig. 3. More precisely, at $r = 4$ the frequency-rank k is associated with small frequencies in comparison with $r = 2, 3$.

The simple approach of this paper to forgetfulness (and, in general, to working memory) introduces a novel framework to study the influence of minimal cognitive mechanisms on the formation and evolution of languages.

Future work could involve the study of the dynamics on general topologies (for instance random graphs or long-range connections), more complex cognitive mechanisms of memory capacities, and the influence of large radius r (for instance, $r \sim \sqrt{n}$) on the appearance of sharp transitions.

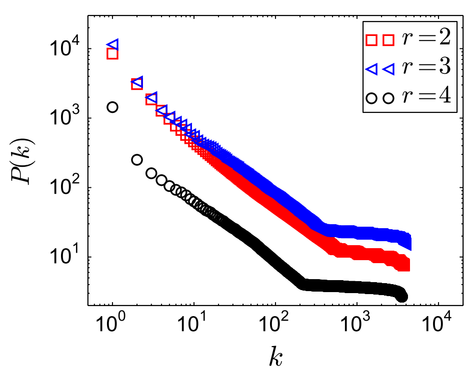

Fig. 3 $P(k)$ versus k on two-dimensional lattices, for the critical values $p_c^2 = 0.43$; $p_c^3 = 0.25$; and $p_c^4 = 0.17$. On a two-dimensional lattice of size $n = 128^2$, it is exhibited after $200n$ time steps the global distribution of the number of individuals showing the $k-$ranked word, $P(k)$, versus k (log − log plot). Averages over ten initial conditions

Acknowledgments The authors like to thank CONICYT-Chile under the Doctoral scholarship 21140288 (J.V.) and the Grants FONDECYT 11400090 (E.G.), ECOS BASAL-CMM (DIM, U. Chile) (E.G.), ECOS C12E05 (E.G.). Finally, the authors acknowledge to anonymous referees for useful commentaries.

Compliance with Ethical Standards

Conflict of Interest Javier Vera and Eric Goles declare that they have no conflict of interest.

Informed Consent All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5). Additional informed consent was obtained from all patients for which identifying information is included in this article.

Human and Animal Rights This article does not contain any studies with human or animal subjects performed by the any of the authors.

References

1. Baddeley A. Working memory, thought, and action. Oxford psychology series. Oxford: Oxford University Press; 2007.
2. Baronchelli A, Felici M, Caglioti E, Loreto V, Steels L. Sharp transition towards shared vocabularies in multi-agent systems. J Stat Mech. 2006;P06014
3. Bickerton D. Language and species. Chicago: University of Chicago Press; 1990.
4. Fatès N. Asynchronism induces second-order phase transitions in elementary cellular automata. J Cell Autom. 2009;4(1):21–38.
5. Ferrer-i-Cancho R, Solé RV. Least effort and the origins of scaling in the human language. Proc Natl Acad Sci USA. 2003;100:788–91.
6. Hauser M. The evolution of communication. Bradford Books. MIT Press, Cambridge, MA (1996). https://books.google.cl/books?id=QbunyscCJBoC
7. Jackendoff R. Possible stages in the evolution of the language capacity. Trends Cogn Sci. 1999;3(7):272–9.

8. Regnault D, Schabanel N, Thierry E. Progresses in the analysis of stochastic 2d cellular automata: a study of asynchronous 2d minority. Theor Comput Sci. 2009;410(4749):4844–55. doi:10.1016/j.tcs.2009.06.024.
9. Steels L. A self-organizing spatial vocabulary. Artif Life. 1995;2(3):319–32.
10. Steels L. Modeling the cultural evolution of language. Phys Life Rev. 2011;8(4):339–56.
11. Steels L. Self-organizing vocabularies. In: Proceedings of artificial life V, Nara, Japan. Japan: Nara; 1996. p. 179–84
12. von Neumann J. Theory of self-reproducing automata. Champain: University of Illinois Press; 1966.
13. Wilson D, Sperber D. Relevance theory. In: Horn LR, Ward G, editors. Handbook of pragmatics. Oxford: Blackwell; 2004. p. 607–632.
14. Wolfram S. A new kind of science. UK: Wolfram Media; 2002.