Modeling phonological similarity effects on the self-organization of vocabularies

Javier Vera (jxvera@gmail.com)
Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez
Santiago, Chile

Abstract

This work describes the formation of a language under phonological similarity constraints. Classical studies on human subjects show that in recalling experiments short-term memory performance was impaired for phonologically similar words versus dissimilar ones. Here, artificial individuals confound phonologically similar words according to a predefined parameter: the more the parameter, the more the confusion. Theoretical results present proofs of convergence for a particular case of the model within a worst-case complexity framework. Computer simulations describe the evolution of an energy function that measures the amount of local agreement between individuals. The main finding is the appearance of sudden consensus changes at critical parameters.

Keywords: Phonological similarity; Short-term memory; Automata networks; Vocabulary.

Introduction

In natural language, a typical linguistic interaction is not a simple sequence of individual actions, but also a form of *joint activity* that involves cooperation and coordination between participants (Tomasello, 2008). What is more, inside the dialogue language users tend to converge in their choice of constructions. This mutual convergence process, or *alignment*, has been extensively studied within computational studies of the formation of language through the Naming Game (Steels, 1995, 2011; Baronchelli, Felici, Caglioti, Loreto, & Steels, 2006; Loreto, Baronchelli, Mukherjee, Puglisi, & Tria, 2011). It attempts to ask how on a population of agents only from local interactions it arises a shared word-meaning association (the simplest version of a *vocabulary*). This model considers a finite population of agents, where each one is endowed with a memory in which it stores an in principle unlimited number of words. At each discrete time step, a pair of agents is selected: one plays the role of *speaker*, one plays the role of *hearer*. First, the speaker refers to an object by using a word. Next, the hearer tries to identify the referent. For this purpose, the hearer inspects its own memory: (1) if the word belongs to the memory of the hearer, both speaker and hearer cancel all the words in their memories, except such word; or (2) if the word does no belong to its memory, the hearer adds the word to its memory. The Naming Game only is possible if pairwise interactions are a kind of joint activity. In a more realistic scenario, both agents interact in a context where the object is located and share focus on the object by means of pointing or eye-gazing. The interaction continues until both agents reach a common word associated to the object, a word-meaning association.

Here, it is assumed that the development of word-meaning associations is founded on self-organization mechanisms arising only from *local* interactions between agents (Steels, 1995,

1996; Baronchelli et al., 2006). Given this self-organized nature, Automata Networks (AN) (Neumann, 1966; Wolfram, 2002) provide the adequate framework to explore alignment from a computational (and mathematical) point of view. AN are extremely simple models where each vertex of a network evolves following a local rule based on the states of "nearby" vertices. Despite of the simplicity of the defining rules, AN exhibit astonishing rich patterns of behavior. Thus, a word-meaning association can be considered as a complex pattern.

What is the relationship between cognitive mechanisms of the individuals and the emergence of language on artificial populations? The question entails the study of computational machines which exhibit limitations and constraints of *human* language users. This work stresses a novel question related to models of the self-organization of language: To what extent do working memory constraints influence the alignment of shared conventions on artificial populations of agents? Language users (and therefore agents playing computational games of the formation of language) suffer limited short-term memory constraints (A. Baddeley & Hitch, 1974; A. Baddeley, 2007). Particularly, *phonological similarity* effects suppose that individuals confound word items sharing large portions of phonological content. A classical work (A. D. Baddeley, 1966) reports experiments where subjects heard sequences of unrelated words and tried to recall in the correct order. The results suggest that memory performance was impaired for phonologically similar words (man, cad, mat, cap, can) versus dissimilar ones (pen, sup, cow, day, hot). This effect is a strong evidence for the existence of the *phonological loop*, understood as part of the short-term (working) memory system (A. Baddeley & Hitch, 1974; A. Baddeley, 2007).

This paper does not attempt to develop a "realistic" model, but rather an abstract symbolic approach that extracts the essential elements of the problem. The simulations described here are based on a parameter that measures the amount of phonological confusion between words (considered as a way to describe the influence of working memory limits): the more the parameter, the more the confusion. The work explores the hypothesis of there being a critical range of the parameter that implies drastic changes in the shared conventions of the entire population. Therefore, the features of language (in particular, the consensus on word-meaning associations) emerge abruptly at some critical range of phonological confusion (Hauser, 1996). This hypothesis is strongly related to previous work on on the absence of stages in the formation and evolution of human languages (Bickerton, 1998; Calvin & Bickerton, 2001; Ferrer-i-Cancho & Solé, 2003).

The work is organized as follows. Section "Description

of the model" explains basic notions on AN, the instrumentalization of phonological similarity and the rules of interaction. Section "Theoretical issues" reports simple mathematical results related to the convergence of particular cases of the model. The next section ("Simulations") describes simulation tasks based on an energy operator, over a parameter that measures the amount of similarity confusion between words. A brief discussion of the results is presented in the final section.

Description of the model

Elements of the AN model

Roughly speaking, the AN model involves the following elements:

- A regular grid graph: vertices represent individuals; edges represent possible communicative interactions. More generally, as in the section "Theoretical results", the graph can be a connected, undirected and simple network.

- Each individual is associated to a *state* that eventually changes along time. This *state* is a way to represent the language of the individual at some time frame.

- A set of *local rules* that define how the system changes. The local rule associated to one individual considers as inputs the states of the nearby individuals (the *neighbors*).

- A function, the *updating scheme*, that indicates the order in which the individuals are updated. Two updating schemes are considered in this work: (1) the *fully-asynchronous* scheme, where at each time step one individual is choosen uniformly at random; and (2) the *sequential* scheme, defined as a permutation of the set of vertices.

In what follows, some of the previous elements will be treated in greater depth.

Basic notions

The set $P = \{1, ..., n\}$ represents the population of individuals, located on the vertices of the regular grid graph $G = (P, E)$, where E is the set of edges. The vertex $u \in P$ only interacts (or "talks") with the set of adjacent vertices $V_u = \{v \in P : (u,v) \in E\}$ (the *neighborhood*). Each individual considers four neighbors: up, down, left and right ones (*Von Neumann* neighborhood). The individual $u \in P$ is endowed with a *memory* set M_u in which it stores words belonging to a finite set W, with p elements.

Let $\Sigma = \{a_1, ..., a_s\}$ be a set of sounds. Each word of W is constructed by a combination of sounds taken from Σ. For instance, $a_3 a_1 a_8 \in W$. The *length* of the word x is its number of sounds. For the sake of simplicity, all the words have the same length L. $x(k)$, with $k \leqslant L$, denotes the k−th sound (or position) of the word x. To explicitly measure the amount of phonological similarity between words, the *Hamming distance* $H(x,y)$ between the words x and y is defined as the number of positions in which they differ. Consider two words $a_4 a_6 a_5$ and $a_7 a_6 a_3$, then $H(a_4 a_6 a_5, a_7 a_6 a_3) = 2$.

Confusion parameter

To explicitly measure the ability to distinguish between words, the *confusion parameter* $\varepsilon \in [0, 1]$ is defined. Suppose that the vertex u faces two words x, y. Then,

if $H(x,y) > \varepsilon L$, u **distinguishes** the words x and y

else u **confounds** the words x and y (or simply "$x = y$")

For instance, the individual $u \in P$ is confronted with the words $x = a_1 a_8 a_6 a_4$ and $y = a_1 a_8 a_6 a_3$. Two values of ε are considered, 0 and 0.5:

- ($\varepsilon = 0$) $H(x,y) = 1 > \varepsilon L = 0 \times 4 = 0$, then u distinguishes the words x and y

- ($\varepsilon = 0.5$) $H(x,y) = 1 < \varepsilon L = 0.5 \times 4 = 2$, then u confounds the words x and y

Local rules

Inspired in the Naming Game (Baronchelli et al., 2006), the *local rule* associated to the individual $u \in P$ is based on two possible actions on the memory M_u:

- u updates its memory M_u by the **addition** of words; or

- u **collapses** its memory if M_u is updated by cancelling all its words, except one of them.

Both actions attempt to take into account *lateral inhibition* strategies (Steels, 1995, 2011) in the alignment process: the individuals add words in order to increase the chance of future agreements (local consensus), and defect the words that do not cooperate with mutual understanding. In a *collapse*, the individuals prefer the *minimal* word, according to the *lexicographic order* over the set of words. The *lexicographic order*, denoted $\prec$, is a generalization of the typical alphabetical order of words on the alphabetical order of their component letters (or sounds). For example, in the dictionary the word "Me" appears before "My" because the letter e comes before the letter y in the alphabet. In some sense, the word "Me" is *lower* than the word "My". Formally, the order $<$ is defined on the set Σ. Two words x and y of length L are considered. Then, $x \prec y$ (x is *lexicographically* lower than y) if the first position in which they differ, say $k \leqslant L$, satisfies $x(k) < y(k)$. For instance, given the set of sounds $\Sigma = \{a, b, c, d\}$, with $a < b < c < d$, the words abc, bcd and cda satisfy $abc < bcd < cda$. Therefore, abc is the *minimal* word or, in other terms, $\min(\{abc, bcd, cda\}) = abc$. Associated to the previous words "Me" and "My", it is possible to write $\min(\{\text{"Me"}, \text{"My"}\}) = \text{"Me"}$.

The preference for the minimal words can be viewed in accordance with the following hypothetical scenario (Nowak & Krakauer, 1999). It is possible to think in a population of early hominids for which leopards represents a higher risk than cows. So, the word "leopard" may be more valuable than "cow". In the terms of this paper, "leopard" can be the *minimal* word.

209

At time step t, the vertex $u \in P$ is selected according to the updating scheme (*fully asynchronous* or *sequential*). Consider a simple population $P = \{1,2,3,4,5\}$ (each number represents one individual). For a *fully asynchronous* scheme, at each time step any individual of P can be selected (for instance, the individual 3). In the next step, the individual 3 can be selected again. For a *sequential* scheme, a permutation of the set P is defined, for instance, the order 5-4-3-2-1. The individual 5 updates first, then the individual 4 updates, taking into account the effects of the changes in the first individual, and so on. At time step 6, the dynamics starts in the same previous way. In other terms, a *sequential* scheme supposes that each individual is updated after that all the other individuals of the population have been updated.

The individual $u \in P$ is completely characterized by its *state* (M_u, x_u), where M_u is the memory to stores words (M_u is a subset of W) and $x_u \in M_u$ is a word that u conveys to the vertices of V_u. The model induces specific communication roles: the vertex u plays the role of "hearer" (it receives the words conveyed by its neighbors); the neighbors of u play the role of "speaker" (they convey words to the vertex u). The set of all words conveyed by the speakers can be re-written as two subsets: N_u and B_u. Roughly speaking, N_u includes the unknown words, and B_u includes the known ones.

The state pair (M_u, x_u) changes according to the following steps, which define the *local rule* of the automata (see Fig. 1):

step 1 the vertex u defines two sets:

$$N_u = \{x_v : (v \in V_u) \wedge (\forall y \in M_u, H(x_v, y) > \varepsilon L)\}$$
$$B_u = \{x_v : (v \in V_u) \wedge (\forall y \in M_u, H(x_v, y) \leqslant \varepsilon L)\}$$

step 2

if $N_u \neq \emptyset$, M_u **adds** the words of N_u

else M_u **collapses** in the word $\bar{x}$, selected at random from the set $\{x \in B_u : H(x, \min(B_u)) \leqslant \varepsilon L\}$ (that is, the new state is $(\{\bar{x}\}, \bar{x})$)

Step 1 comprises (1) the speaker's behavior (the neighbors convey words to the vertex u); and (2) the definition by the hearer of the sets N_u and B_u. Given a conveyed word x_v, $v \in V_u$, the hearer u decides between: either x_v is added to N_u if for all $y \in M_u, H(x_v, y) > \varepsilon L$; or x_v is added to B_u, otherwise.

Step 2 summarizes the behavior of the hearer in order to *align* itself with the speakers. In the case that $N_u \neq \emptyset$, the hearer simply adds to its memory the words of N_u. Otherwise ($N_u = \emptyset$), the hearer collapses its memory in the word $\bar{x} \in B_u$. The word $\bar{x}$ is selected uniformly at random from $\{x \in B_u : H(x, \min(B_u)) \leqslant \varepsilon L\}$. Thus, the preferred word $\min(B_u)$ can be confused by "similar" ones (with $H(\cdot, \min(B_u)) \leqslant \varepsilon L$).

Dynamics of the automata

As initial configuration each individual receives a word constructed by a random combination of L sounds from the set of symbols Σ. Thus, at $t = 0$ each individual is associated

to a state of the form $(\{x\}, x)$, with $x \in W$. Each discrete time step $t \geqslant 0$ supposes that one individual, say u, is selected uniformly at random (the *fully-asynchronous* scheme) or according to a permutation of the set of vertices (the *sequential* scheme). The individual receives the words conveyed by its neighbors: it plays the role of hearer, the neighbors play the role of speaker. Regarding to the conveyed words the individual follows the two defined steps in order to decide possible changes in its own language state:

step 1 the individual defines the sets N_u (unknown words) and B_u (known words).

step 2 the individual either **adds** words if N_u is non empty or **collapses** its memory in the minimal word of B_u, if N_u is empty.

Both steps involve the possibility of confusion between *similar* words. In the next time step $t + 1$ another individual (possibly the same one) is selected, and so on.

A *fixed point* is a configuration invariant under the application of local rules, which can be interpreted as a final consensus configuration, where all individuals agree about some linguistic convention.

Theoretical issues

Convergence under $\varepsilon = 0$

An interesting problem related to the formation of consensus on linguistic conventions is to propose a proof of convergence. Given the mathematical framework of this paper, the problem becomes to count (in the *worst* case) the number of simulation steps whom the dynamics needs to stop, that is, to prove that after a finite number of time steps the population reaches a shared word-meaning association. Despite that other works have solved related tasks (DeVylder & Tuyls, 2006), the novelty of the rest of this section is to develop a convergence proof based on the worst-case complexity, which measures the amount of resources (running time) needed by the system if it is considered as an *algorithm*. Running time, defined as the number of steps until the entire population reaches a global consensus language, indicates the largest dynamics performed by the automata given the size n of the population (denoted $O(f(n))$, where $f(n)$ is a function of n). For instance, $O(n^2)$ means that in the worst-case the running time has a growth rate scaling as n^2.

In this section, individuals are located on a general undirected and connected network (not necessarily a regular *grid*), and they do not confound words, that is, $\varepsilon = 0$.

It is straightforward to notice that at $\varepsilon = 0$ ($\forall y \in M_u, H(x_v, y) > \varepsilon L$) is equivalent to ($x_v \notin M_u$). Roughly speaking, the expression ($\forall y \in M_u, H(x_v, y) > \varepsilon L$) means that x_v is "different" to every word in M_u and, therefore, $x_v \notin M_u$. As a consequence, the two steps of the rule take a simpler form:

step 1 the vertex u defines two sets:
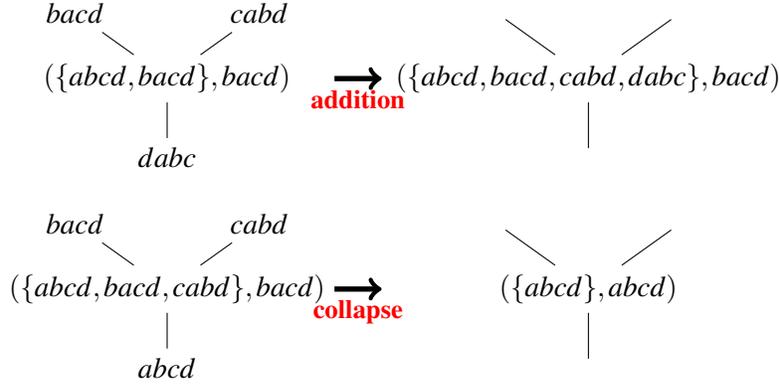
$$N_u = \{x_v : (v \in V_u) \wedge (x_v \notin M_u)\}$$

Figure 1: **Example of the two actions at $\varepsilon = 0$.** The order $a < b < c < d$ is defined on the set $\Sigma = \{a,b,c,d\}$. $W = \{abcd, bacd, cabd, dabc\}$. The *lexicographic* order establishes $abcd \prec bacd \prec cabd \prec dabc$. The confusion parameter is set to $\varepsilon = 0$ (that is, the individuals do not confound words). The central vertex ($u \in P$) has been choosen. Four individuals participate in the interaction: the central vertex u and its three neighbors of V_u. Two different configurations are showed. First row (**addition**): $B_u = \{bacd\}$ and $N_u = \{cabd, dabc\}$. Second row (**collapse**): $B_u = \{abcd, bacd, cabd\}$ and $N_u = \emptyset$.

$$B_u = \{x_v : (v \in V_u) \wedge (x_v \in M_u)\}$$

step 2

  if $N_u \neq \emptyset$, M_u **adds** the words of N_u

  else M_u **collapses** in the word $\min(B_u)$ (the new state is $(\{\min(B_u)\}, \min(B_u))$)

First of all, a sequential updating scheme is considered. This means that a *permutation* of the vertices is defined. As previously noted, each individual is updated after that all the other individuals of the population have been updated. In the worst case, each time step supposes that one individual adds one word. This process ends when some individual has all possible words, that is, after $p - 1$ steps (there are p words). Since the population has size n, after $n(p-1)$ steps the individuals must collapse their memories. Then, at step $t^* = n(p-1)$ all individuals have been collapsed at least once. As the next theorem shows in detail, the minimum conveyed word at t^* propagates in the system until it reaches a fixed point where all individuals convey this minimum word.

Theorem 1 *(Goles, Montealegre, & Vera, 2016)*

  Consider a population of n individuals playing the automata model with the set of p words W and confusion parameter $\varepsilon = 0$. Then,

  (I) for the sequential scheme, the system converges to fixed points in at most $O(n^2 p)$ steps;

  (II) for the fully-asynchronous scheme, the system converges to a fixed point in expected time $O(n^2 p \log(n))$.

Proof 1 (I) Initially there are p words. Then, in at most $p-1$ updates a vertex has collapsed for the first time (in the worst case the vertex must add every possible word, one at a time). This implies that in $n(p-1)$ steps ($p-1$ updates of each vertex) all vertices have collapsed at least one time. Let m

be the minimum conveyed word at step $t^* = n(p-1)$, and let u be a vertex such that $x_u = m$ (in more precise terms, $m = \min(\{x_u\}_{u \in P})$).

  Since u has collapsed at least one time, the updating scheme is sequential, and m is the minimum word, then u must have another neighbor $v \in V_u$ conveying m. In consequence, after t^* both vertices u and v will remain conveying m, and at each time a neighbor of any of these two vertices will necessarily collapse in the word m. The graph has a diameter of $O(n)$ and each np steps it occurs a collapse. Therefore, in at most $O(n^2 p)$ steps the system converges to a fixed point where all vertices convey the same word m.

  (II) In the fully-asynchronous scheme, at each time step a single vertex is picked independently and uniformly at random. The expected convergence time grows by a factor of $O(\log(n))$ with respect to the sequential scheme. Notice that in the proof of the part (I) it is shown that after updating [$O(np)$ times] [each vertex], a fixed point is reached. From the well known coupon collector's problem (see, for instance, (Grimmett & Stirzaker, 2001)), the expected number of steps required to update [at least once][every vertex in the graph (or pick every coupon)] is $O(n \log(n))$. Since in a fully asynchronous updating scheme each step is independent to the others, the result follows.

A brief note on dynamics under $\varepsilon = 1$

At $\varepsilon = 1$, individuals confound any pair of words $x, y \in W$. Indeed, for all $x, y \in W$ $H(x,y) \leqslant \varepsilon L \leqslant L$. A simple result says:

Proposition 1 *Consider the automata model with confusion parameter $\varepsilon = 1$. Suppose that the individual $u \in P$ has been selected at some time step. Then,*

  1. $N_u = \emptyset$; and

  2. $B_u \neq \emptyset$

A relevant consequence of Proposition 1 is that **additions** are not allowed. This fact requires further work.

Simulations

Protocol

To describe the amount of agreement between individuals, an "energy" operator is defined. This energy-based approach arises from a "physicist" interpretation, related to (Regnault, Schabanel, & Thierry, 2009): the energy measures the amount of local unstability of the system. At each neighborhood V_u, the function $\sum_{v \in V_u} H(x_u, x_v)$ is defined, which measures the *Hamming* distance between the word x_u and the words conveyed by the neighbors of the individual u. This function is bounded by 0 (in case of *agreement*, that is, the individual u and its neighbors convey the same word) and $4L$ (*disagreement*, which means that the individual u and its neighbors convey radically different words). Summing over all individuals defines the total energy function at some time step t:

$$E(t) = \frac{1}{4Ln} \sum_{u \in P} \sum_{v \in V_u} H(x_u, x_v) \qquad (1)$$

The function $E(t)$ is bounded: $0 \leqslant E(t) \leqslant 1$. Thus, the dynamics of the AN model can be understood as the trajectory between initial configurations associated to large amounts of unstability (with $E(0) \sim 1$) and final consensus configurations where $E(t) \sim 0$, or equivalently, all individuals convey similar -in the sense defined by the *Hamming* distance- words.

The analysis focuses on two-dimensional periodic lattices of size $n = 128^2$ with Von Neumann neighborhood. The simulations describe $500n$ steps of the energy function $E(t)$. At each time step, one vertex (playing the role of hearer) is selected uniformly at random (*fully-asynchronous* scheme). The plots show average values over 20 initial conditions. An initial condition is defined as follows: each individual receives uniformly at random a word constructed by a random combination of L sounds from the set of symbols Σ, $|\Sigma| = 10$. Word-length L varies from: $\{2, 4, 8, 16, 32, 64\}$. ε is varied from 0 to 1 with an increment of 10%.

Given the worst-case complexity approach to theoretical aspects of convergence, it is important to notice that on low-dimensional lattices it seems hard that one individual adds $O(np)$ words after it collapses, because lattices are low-connected. This intuition suggests that running times on lattices will be lower than theoretical bounds ($O(n^2 p \log(n))$) for the *fully-asynchronous* scheme).

Results

There are several remarkable aspects, as shown in Fig. 2 and Fig. 3. First of all, $E(t = 500n)$ versus ε exhibits at $\varepsilon = 1$ a maximum which is close to 0.5 for all values of L (Fig. 2). Secondly, to the extent L grows, $E(t = 500n)$ versus ε evolves more "smoothly": $L \leqslant 8$ supposes "ladder" steps which mean that different values of the confusion parameter ε lead to the same energy. Finally, focusing the description on $L = 32$ (Fig.

3), it is interesting to notice that at $\varepsilon = 0.7$ the average value of $E(t)$ stops approximately after $t = 200n$ steps. This fact exhibits that only a few set of runs does not converge to the global minimum of the $E(t) = 0$. This strongly suggests the appearance of a critical parameter $\varepsilon^* \sim 0.7$ which clearly defines two phases in the evolution of $E(t = 500n)$ versus ε: (1) $\varepsilon < \varepsilon^*$ implies the convergence to the global minimum $E(t) = 0$, where all individuals convey the same word; and (2) for $\varepsilon \geqslant \varepsilon^*$ the dynamics changes drastically until it reaches local minima of the energy function ($E(t) \to 0.5$).
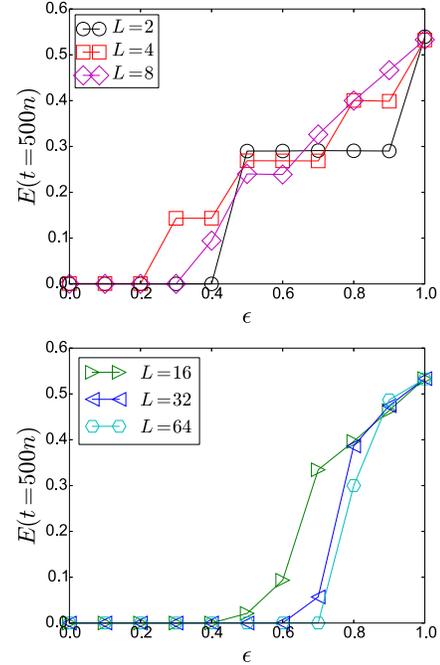

Figure 2: $E(t = 500n)$ **versus** ε. On a two dimensional grid of size $n = 128^2$, the figure shows the final value of the energy function versus the parameter ε, after $500n$ steps or until they reach the global minimum $E(t) = 0$. The plots show averages over 20 initial conditions: for an initial condition, each vertex receives a word constructed by a random combination of L sounds from the set of symbols Σ, $|\Sigma| = 10$. Word-length varies from: $\{2, 4, 8\}$ (top); and $\{16, 32, 64\}$ (bottom). ε is varied from 0 to 1 with an increment of 10%.

Discussion

This work summarizes an AN approach to the formation of linguistic conventions under phonological similarity (and, in general, short-term memory) mechanisms. The paper presents the evolution of an energy functional, defined as a word "confusion" average, during the alignment game. Two aspects are remarkable. On the one hand, the appearance of drastic transitions can be related to previous works that focus on the absence of stages in the formation and evolution of human languages (see, for instance, (Bickerton, 1998; Calvin & Bickerton, 2001; Ferrer-i-Cancho & Solé, 2003)).
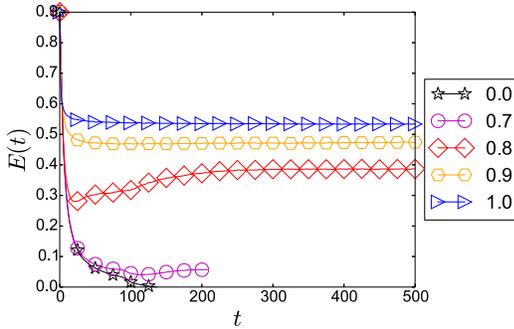
Figure 3: $E(t)$ **versus** t, $L = 32$**, for different values of** ε. On a two dimensional grid of size $n = 128^2$, the figure shows evolution of the energy function $E(t)$ versus t, for different values of the parameter ε. Simulations run $500n$ steps or until it reaches the global minimum $E(t) = 0$. The plot shows averages over 20 initial conditions: for an initial condition, each vertex receives a word constructed by a random combination of L sounds from the set of symbols Σ, $|Σ| = 10$. $L = 32$ and ε is varied from $\{0, 0.7, 0.8, 0.9, 1\}$.

On the other hand, the proposed model becomes an alternative (mathematical) framework for agent-based studies on language formation.

As a first approach to the convergence of the formation of linguistic conventions, this paper presents within an AN account simple results of the number of steps needed to reach fixed points. As the main tool, the proofs are based on the *worst* case of convergence.

Many extensions of the proposed model should be studied with the purpose to describe the role of cognitive constraints on the formation of word-meaning associations. First, within a mathematical point of view it seems interesting to explore convergence bounds on regular lattices. Also, a comparison between theoretical and numerical convergence times should be carried out. Second, AN allow to study new aspects of the formation of linguistic conventions. Indeed, the model can be extended to other *updating schemes*, for example, the *synchronous* one, where at each time step all individuals are updated. Third, the results should be compared with larger word lengths (L) and several number of symbols (Σ). Finally, future work should involve more realistic ways to measure the amount of phonological confusion and its effects on the formation of conventions.

Acknowledgments

References

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford University Press.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, p. 47-90). Academic Press.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*(4), 362-365. (PMID: 5956080)

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

Bickerton, D. (1998). Catastrophic evolution: The case for a single step from protolanguage to full human language. In J. R. Hurford, S. M. Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 341–358). Cambridge: Cambridge University Press.

Calvin, W., & Bickerton, D. (2001). *Lingua ex machina: Reconciling darwin and chomsky with the human brain*. MIT Press.

DeVylder, B., & Tuyls, K. (2006). How to reach linguistic consensus: A proof of convergence for the naming game. *The Journal of Theoretical Biology*, *242(4)*, 818–831.

Ferrer-i-Cancho, R., & Solé, R. V. (2003). Least Effort and the Origins of Scaling in the Human Language. *Proceedings of the National Academy of Science (USA)*, *100*, 788–791.

Goles, E., Montealegre, P., & Vera, J. (2016). Naming game automata networks. *Journal of Cellular Automata*. (accepted for publication)

Grimmett, G., & Stirzaker, D. (2001). *Probability and random processes*. OUP Oxford.

Hauser, M. (1996). *The evolution of communication*. MIT Press.

Loreto, V., Baronchelli, A., Mukherjee, A., Puglisi, A., & Tria, F. (2011). Statistical physics of language dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, *2011*(04), P04006.

Neumann, J. von. (1966). *Theory of self-reproducing automata*. Champain, IL: University of Illinois Press.

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, *96*(14), 8028-8033.

Regnault, D., Schabanel, N., & Thierry, E. (2009). Progresses in the analysis of stochastic 2d cellular automata: A study of asynchronous 2d minority. *Theoretical Computer Science*, *410*(4749), 4844 - 4855.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life*, *2*(3), 319-332.

Steels, L. (1996). Self-organizing vocabularies. In *Proceedings of artificial life v, nara, japan* (pp. 179–184). Nara, Japan.

Steels, L. (2011, December). Modeling the cultural evolution of language. *Physics of Life Reviews*, *8*(4), 339-356.

Tomasello, M. (2008). *The origins of human communication*. MIT Press.

Wolfram, S. (2002). *A new kind of science*. Wolfram Media.