Viability of an elementary syntactic structure in a population playing naming games

Edgardo Brigatti*

Instituto de Física, Universidade Federal do Rio de Janeiro, Av. Athos da Silveira Ramos, 149, Cidade Universitária, 21941-972, Rio de Janeiro, RJ, Brazil

(Received 6 February 2012; published 13 August 2012)

We explore how the social dynamics of communication and learning can bring about the rise of a syntactic communication in a population of speakers. Our study is developed starting from a version of the Naming Game model in which an elementary syntactic structure is introduced. This analysis shows how the transition from nonsyntactic to syntactic communication is socially favored in communities which need to exchange a large number of concepts.

DOI: 10.1103/PhysRevE.86.026107 PACS number(s): 89.65.−s, 05.65.+b, 89.75.Fb

I. INTRODUCTION

The study of the evolution of languages and their structures has generated a very rich debate spanning different disciplines and approaches. The ideas developed in the linguistic community, which introduced the thesis of considering some linguistic structures as innate with some specific properties genetically encoded in a language module or organ [1], have been very prolific. They bootstrapped the development of many works where pure evolutionary perspectives are introduced to explain the generation of languages. Here, the dominant paradigm is the Darwinian evolution of biological systems. The description of language evolution is based on a biological dynamics constructed above the concepts of natural selection, mutation, and fitness, elaborated in terms of communication success [2]. This approach is particularly well suited to describe evolution from a functionalist perspective, where the category of utility is the one that drives the dynamics. In any event, many linguistic properties appear to be so highly abstract as to even hinder communication [3]. This means that they are quite difficult from those being introduced on a purely functional basis, and they cannot be explained merely in terms of communicative effectiveness or cognitive constraints. Moreover, it is hard to explain how shifts from learned linguistic conventions can be fixed into genetically encoded principles necessary to evolve a language module. Cultural conventions change much more rapidly than genes, and the Baldwin effect, a possible Darwinian solution to this challenge, cannot be the solution to this puzzle [4]. Biological models can be seen more as a powerful metaphor for studying the effects of random copying and selection, but more specific mechanisms, typically related to cultural transmission, should be considered.

Recently, attention was paid to defining specific cultural dynamics, directly related to linguistic ability. The mechanisms which define the dynamics of the evolution of language are different from those underlying biological evolution. Language is transmitted among people through learning and not DNA. It is shaped by processes of cultural transmission across generations of language learners. In this view, linguistic constructions are not innate, but rather they are acquired through some form of probabilistic learning. This learning process, articulated on the use of cognition-general principles, has become the central issue governing language evolution. In fact, learning defines the dynamics of linguistic variants, and the differences among language learnability control these dynamics. Learnability is quantified measuring the learner's capacity to recover a complete description of a linguistic construction to which she/he has been exposed sufficiently [5]. Several works have studied the differential learnability of competing linguistic variants [6,7] and also their dynamics in the absence of selection [8].

Biological dynamics and cultural dynamics of learning are two central issues which determine the creation of linguistic structures, but they are not the only ones. Language is constructed for communication. It is not only the basis for social relationships, but it is also based on social relationships. Individual learning is just one aspect of a more general and collective process. The fixation of the linguistic conventions among a population of speakers is another dynamics related to the linguistic definition and the structures which appear to be learnable at an individual level must be socially fixable. These social dynamics cause a pressure on language which shapes a shared communication system. This is a form of collective learning. It is important to perceive how on a social level even completely arbitrary linguistic properties can succeed. In fact, if the same convention is adopted by all members of a community, this convention can work and finally it becomes fixed. The only important fact is that everyone adopts the same set of culturally mediated conventions. Even different conventions, if equally effective, may serve equally well if there are no costs or no conflicting functional pressures. The fixation of structures is not driven by a fitness or a learnability advantage, but rather by the mechanisms which generate consensus about the linguistic elements used by speakers. This process is not necessarily a functional process and is not only driven by utility. A lot of works related to this social aspect of language appeared recently, following a seminal paper by Steels [9]. These ideas were first used to describe the birth of neologisms, and they have been tested by an artificial experiment in which embodied software agents bootstrap a shared lexicon without any external intervention [10]. Robots concretize a *language game* developing a vocabulary throughout a self-organized process called the Naming Game. Recently, these studies have also attracted the interest of the statistical physics community. An initial study in this

*edgardo@if.ufrj.br

direction [11] was inspired by experiments conducted with the use of robots [9]. In that work, each player is characterized by an inventory of words which can all be used to name the same object. At each time step, two players, randomly chosen, interact following some simple rules. These dynamics force the system to undergo a disorder/order transition toward an absorbing state characterized by a common word used by all the players. These ideas have been developed to describe other phenomena, such as the emergence of universality in color naming patterns [12] and the self-organization of a hierarchical category structure in a population of individuals [13].

The ideas and considerations we have exposed can be applied across different linguistic levels: lexicon, phonology, morphology, and syntax. In this work, we will focus our attention on syntax. Syntax is a process to combine progressively symbolic units in an ordered output which falls within the quite narrow bounds that delimit human language. This is obtained by merging words into larger units and superimposing algorithms that determine the reference of items that might otherwise be ambiguous or misleading [14]. As proposed in the minimalist program [15], the basic syntax-creating process is Merge, a process that takes two units (words, phrases, clauses) and forms them into a single one satisfying some constraints. This means that Merge has many restrictions on the items to be merged, and there is a consistent way of merging them. As we are interested in the transition from nonsyntactic to syntactic communication, it is reasonable to look for the simplest advance from the pre-syntactic (one-word) stage, even if it does not specifically correspond to the syntax of some present-day language. This first step can be identified with the most basic (proto)syntactic combination, namely a flat concatenation of two symbols, where all the possible combinations are functional [16]. This correspond to a purely linear bead-stringing process, a practice which underlies protolanguages, such as the one used by speakers of a pidgin language [17].

In our work, we are interested in exploring the transition from nonsyntactic to syntactic communication from a social dynamics point of view. Directly following the ideas of Nowak *et al.* [18], the example that we are going to explore can be stated in this way. Let us consider the situation in which a speaker is interested in communicating some concepts. If she/he uses a nonsyntactic language, a symbol (word) is used for each concept. In the case of a syntactic language, a combination of two symbols, for example one for the object i and one for the action j, can be used to communicate the concept C_{ij} [18]. In the following, we will consider the simplified situation in which the number of objects and actions exchanged in the communication are the same (S). Moreover, all the possible combinations of these symbols can occur and correspond to a meaningful concept. It follows that, in this model, the possible combinations can be represented by a matrix $C_{ij} = S \times S$ (see Table I). For example, as a particular situation we can think that the line elements represent nouns and the column elements represent verbs. A population of individuals coevolves this system of symbols, with or without syntax, by playing elementary language games (the Naming Game) analogous to the ones introduced in [11]. In this way, we can analyze the differences between syntactic and nonsyntactic communication, and we can distinguish when the transition from nonsyntactic to syntactic communication is socially favored.

TABLE I. In a Naming Game with syntax for the communication of nine concepts, we represent these concepts using the matrix $C = 3X3$. Each concept is specified by the couple formed by two different possible words contained in two different inventories among the six inventories a,b,c,d,e,f. For example, in a concrete situation, a,b,c can stand for an object (noun) and d,e,f for an action (verb).

	a	b	c
d	$a+d$	$b+d$	$c+d$
e	$a+e$	$b+e$	$c+e$
f	$a+f$	$b+f$	$c+f$

The paper is organized as follows. Section II A introduces a version of the basic Naming Game model for the communication of one concept, and Sec. II B illustrates the model with one concept and syntax. Section II C describes the generalization of the model for many different concepts in the case of syntactic or nonsyntactic communications. In Sec. III A, we show the numerical results obtained for a one-concept model using syntax, or not, along the communications. Section III B is devoted to illustrating what happens with the introduction of syntax in a many-concept game. Conclusions are reported in Sec. IV.

II. THE MODEL

A. The basic model for one concept

The Naming Game is played by P agents who try to reach consensus in naming a single concept. An inventory, which contains an arbitrary number of words, represents each agent. Population starts with empty inventories. At each time step, two agents are randomly selected; the first one assumes the role of speaker, the second one of hearer. Then, the following microscopic rules [11] control their actions:

(i) The speaker retrieves a word from its inventory or, if its inventory is empty, invents a new word.

(ii) The speaker transmits the selected word to the hearer.

(iiia) If the hearer's inventory contains such a word, the communication is a success. The two agents update their inventories so as to keep only the word involved in the interaction.

(iiib) Otherwise the communication is a failure. The hearer learns the word communicated by the speaker.

The players invent new words choosing from among 32 possible words with equal probability. In contrast with the classical implementation of the model [11], we work with a fixed maximum number of possible different words. This is the only difference between our implementation and the classical one. We introduce this simplification with the goal of implementing a light model that can be easily generalized for the description of the naming process for more than one concept. For this reason, we need a fast algorithm. This is obtained by using boolean programming techniques, which causes a veritable improvement in the computational times. An example of these dynamics is represented in Fig. 1.

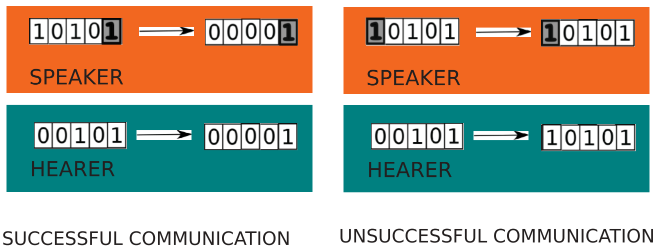

FIG. 1. (Color online) Example of the dynamics of the inventories for a one-concept model without syntax. On the left we present a successful game, on the right a failed one. We represent the 32 possible symbols with only five slots for the sake of clarity. When a symbol is not present in the inventory, the corresponding bit is set to 0; otherwise, when the symbol is present, the bit is set to 1. The shadowed elements are the ones transmitted by the speaker.

B. The syntax model for one concept

Now we consider a context in which one single concept is exchanged with the use of compositionality, a rudimental form of syntax. In this situation, the concept is represented by a couple of symbols $\alpha_x + \beta_y$, each one sorted from a different inventory K_x and K_y. It follows that each agent is characterized by two different inventories K_x and K_y. At each time step, the following microscopic rules control the communication:

(ia) The speaker retrieves a word (α_x) from its inventory K_x; if its inventory is empty, the speaker invents a new word.

(ib) The speaker retrieves a word (β_y) from its inventory K_y; if its inventory is empty, the speaker invents a new word.

(ii) The speaker transmits the selected pair of words to the hearer.

(iiia) If the hearer's inventories K_x and K_y contain the pair of words (α_x, β_y), the communication is a success. The two agents update their inventories so as to keep, in each one, only the correspondent words involved in the interaction (α_x in K_x and β_y in K_y).

(iiib) Otherwise the communication is a failure. The hearer adds the words he does not know (one or two) to the corresponding inventory (inventories); the speaker does nothing.

Here we hypothesize that the learning of segmented elements of the utterance is possible even if the communication is a failure. This idea is supported by the fact that it is not necessary to have any positive feedback to identify the components of a speech. In fact, some popular experiments shown how very young infants can achieve the task of word segmentation of an utterance with only minimal exposure, just by exploiting the transitional probabilities between syllables [19]. Even so, the ability to use exclusively statistical information coming from a passive exposure to process a given language stream seems to be confined to the individuation of the segments of a stream, but not to acquiring the generalization correspondent to a syntax structure [20]. For this reason, we hypothesize that the fixation of the structured element $\alpha_x + \beta_y$ is possible only if there is a communication success, i.e., on the basis of exposure to a positive feedback.

C. Many-concepts games

In this situation, agents develop communications which can exchange C different concepts. If no syntax structures are present, each concept is represented by one symbol. If syntax is introduced, a combination of two symbols, $\alpha_x + \beta_y$, each one picked up from a different inventory, represents each concept. As in the basic model, each symbol is represented by 1 of 32 different possible words.

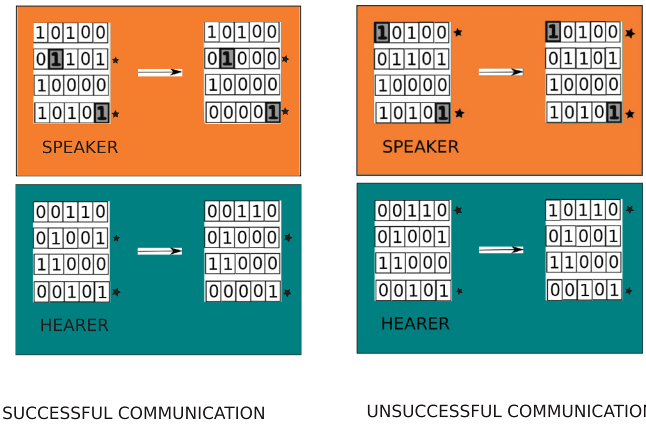

FIG. 2. (Color online) An example of the structure and evolution of the inventories for the model with syntax when four concepts ($S = 2$) are exchanged. In this particular example, the number of inventories ($2S$) is equal to the number of concepts. We can think that the first two inventories represent nouns and the third and fourth stand for verbs. The starred inventories are the ones corresponding to the concept sorted out in a specific communication event. The shadowed elements are the ones transmitted by the speaker.

It follows that for a C concepts game, if there is no syntax, each agent is represented by $K_z, z = 1, 2, \ldots, C$ inventories, each one containing no more than 32 words. These words are exchanged in the same way as in a single-concept Naming Game without syntax.

In the case of a syntactic communication, as explained in the Introduction, the concepts C_{ij} are represented by the elements, generated by the combination of two symbols, of a matrix $C_{ij} = S \times S$. It follows that a C concepts game with syntax is obtained by introducing agents represented by K_z, $z = 2, \ldots, 2S$ inventories, each one containing no more than 32 words. In Table I, we give an example of a game with nine concepts. At each time step, a concept is chosen determining the two inventories that represent it (for example, K_a and K_e, from which we represent the concept with the couple $\alpha_a + \beta_e$). The dynamics of each single communication is the same as the one-concept game with syntax (an example of a four-concept game is represented in Fig. 2).

III. RESULTS AND DISCUSSION

A. One-concept model

We describe the time evolution of our system on the basis of some of the usual global quantities [11,21,22], namely the total number of words (N_{tot}) present in the population and the success rate (R_S), which measures the average rate of success of communications. This is obtained by evaluating the frequency of successful communications in a given time interval.

The basic model is a simplified version of the original Naming Game, where the number of different words introduced in

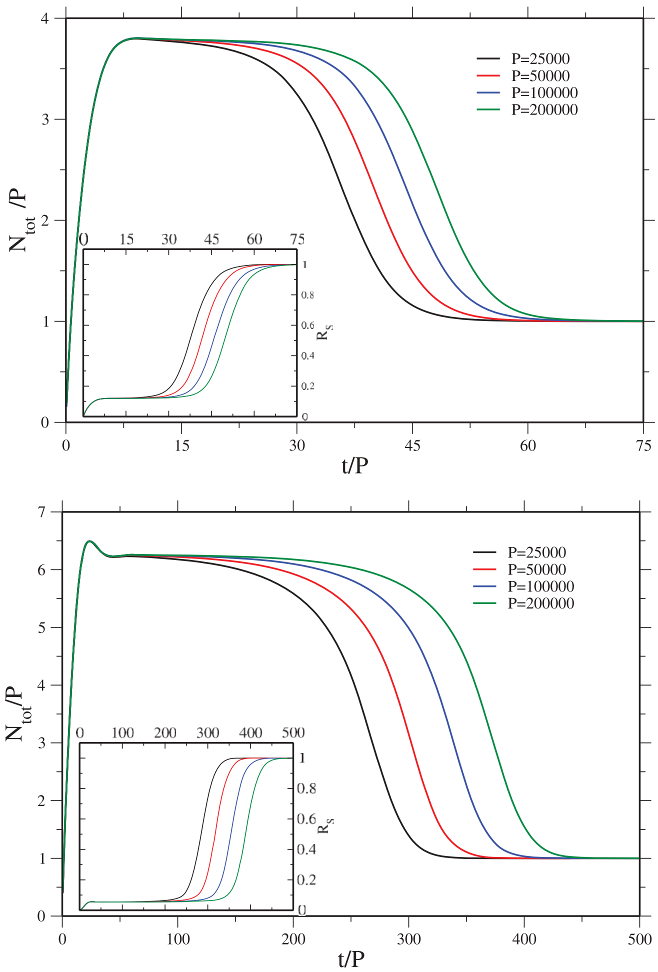

FIG. 3. (Color online) One-concept model. Top: model without syntax. We present the temporal evolution for the total number of words divided by the total population. The inset shows the success rate $R_S(t)$. Bottom: the same data for the syntactic model. All data are averaged over 1000 simulations.

the system is limited and sorted among a cache of 32. As a result, a homologous behavior occurs with some limited differences (see also [21] for similar results). In our version of the model, the maximum number of words for each agent is constant and it does not scale with the square root of the population. This important fact is responsible for remodeling the temporal scaling behavior. It is important to point out that, as the mean number of words for each player is always well below 32, the model maintains all the basic features of the original one, and the possibility of the invention of new words is not affected.

All agents start with an empty inventory, and an initial transient exists which corresponds to the rise of $N_{\text{tot}}(t)$. This stage finishes when this quantity attains a maximum value ($\max[N_{\text{tot}}] \approx 3.8P$), which is maintained along a plateau. When the redundancy of words reaches a sufficiently high level, the number of successful plays increases. The curve $N_{\text{tot}}(t)$ begins a decay toward the consensus state, corresponding to one common word for all the players, reached at time T. In Fig. 3, we report the temporal evolution for $N_{\text{tot}}(t)$ and $R_S(t)$.

As shown in [11], it is possible to estimate the maximum number of total words using some simple analytical considerations. If we represent the mean total number of words for an agent, at time step t, with $n(t)$, and the mean total number of different words with $D(t)$, we obtain

$$n(t+1) - n(t) = \frac{1}{n(t)}\left(1 - \frac{n(t)}{D(t)}\right)\frac{1}{2} - \frac{1}{n(t)}\frac{n(t)}{D(t)}[n(t) - 1]. \tag{1}$$

We are considering that the probability for the speaker to communicate a specific word is $\frac{1}{n(t)}$ and the probability for the hearer to own that word is $\frac{n(t)}{D(t)}$. It follows that the first term represents the gain term for a failed communication [which increases $n(t)$ by 1/2], and the second term represents the loss term [which decreases $n(t)$ by $n(t) - 1$]. We can use this equation for describing the P dependence. If we assume that at the plateau, where we can consider $n(t+1) - n(t) = 0$, $n(t)$ scales as αP^β, and that $D \approx 32$, we can write

$$\frac{1}{2\alpha P^\beta}\left(1 - \frac{\alpha P^\beta}{32}\right) = \frac{1}{32}(\alpha P^\beta - 1). \tag{2}$$

This equation reduces to $\frac{1}{P^\beta} \propto P^\beta$, which forces $\beta = 0$. This fact implies that the number of total words for each player is not dependent on P. It follows that $\max[N_{\text{tot}}] \propto P$, as can be seen in Fig. 3. Equation (2) can also be used to evaluate the exact numerical value of the plateau. It is sufficient to consider αP^β as a constant, and the corresponding value is 4.25. If we take into consideration that our equations are a mean field approximation that does not account for the correlations built up between the individuals' inventories, the value is comparable with the result 3.8 obtained by the simulations.

We explored the behavior of the convergence time T^1_{ns} (the index 1 stays for a one-concept play). As stated before, this is the time at which the system reaches the consensus state, corresponding to one shared word for all the players. We studied the dependence of T^1_{ns} on P averaging over different simulations, obtaining, throughout a regression, the following dependence: $T^1_{ns}(P) \approx -22.1P + 7.6P \ln P$ (see Fig. 4). This result corresponds to the average convergence time over different simulations, which is obviously different from the convergence time of the mean simulation (the one presented in Fig. 3). These analyses, as well as the following ones, are consistent for sufficiently large populations.

We can support these numerical results with some analytical considerations analogous to the ones presented in [23]. During the time evolution, we can distinguish two periods. (i) A first interval, between $t = 0$ and the time when the system reaches the maximum number of words ($t_{\max}$), which clearly scales linearly with the population size (see Fig. 3). (ii) A second interval, between $t_{\max}$ and T^1_{ns}, which is governed by the following dynamics. To reach convergence, the mean number of words for each individual, which does not depend on P, has to decrease to 1. As at each play the loss term does not depend on P, from the definition of the dynamics of our model a necessary condition for convergence is that each agent must win at least once. For this reason, near convergence, the number of agents which did not have a successful interaction (P^*) should be finite. We can estimate this number. In fact,

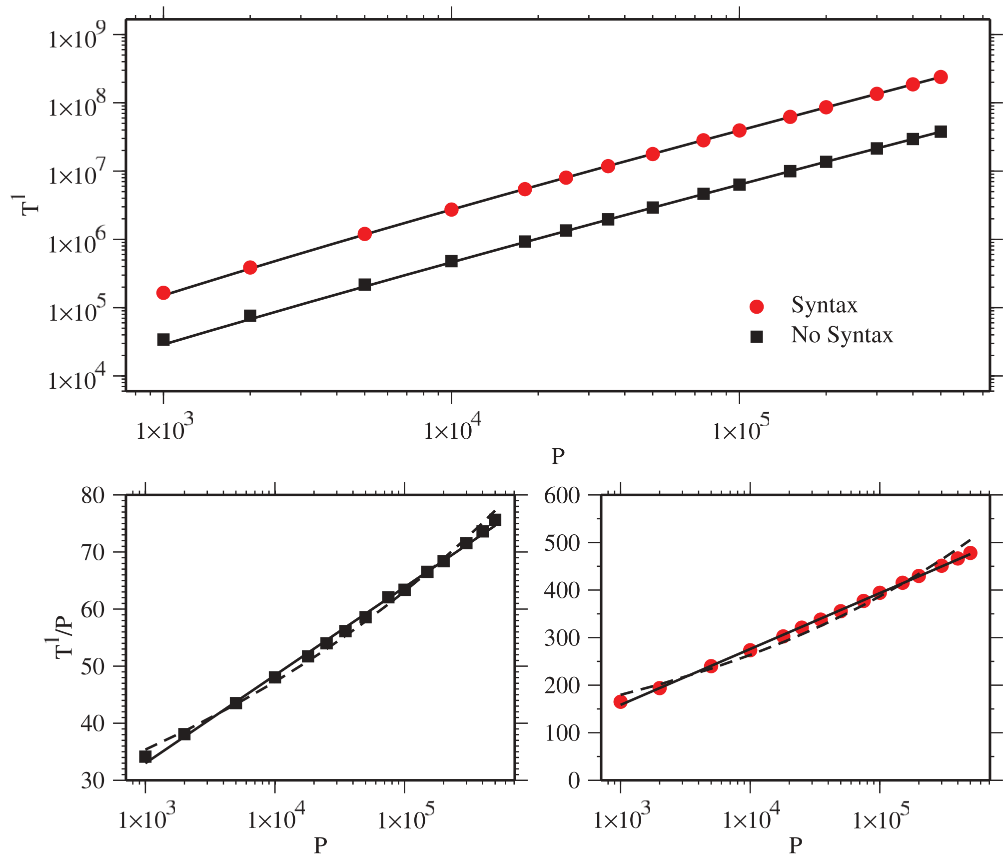

FIG. 4. (Color online) We present the convergence time as a function of the population size for the one-concept model. Top: the model with no syntax is well fitted by the depicted relation $T_{ns}^1 \approx -22.1P + 7.6P \ln P$. The model with syntax is well fitted by the relation: $T_s^1 \approx -213.2P + 52.7P \ln P$. Bottom: rescaled convergence times for the model without syntax (on the left) and with syntax (on the right). The rescaled times are well fitted by the function $a_1 + a_2 \ln P$ (continuous lines). A power law fit of these data ($b_1 P^{b_2}$), represented by the dashed lines, turns out to be moderately less accurate than the previous one, which can be derived from some theoretical considerations.

$P^*(t) = P[1 - R_S(t)/P]^t$, where $1/P$ is the probability of selecting an agent and $R_S(t)$ corresponds to the probability of a success. As can be seen in the inset of Fig. 3, $R_S(t)$ does not depend on P and it is practically constant for a long time after $t_{\max}$. It follows that $(1 - 1/P)^{t_{\text{diff}}} \propto 1/P$, where $t_{\text{diff}} = T_{ns}^1 - t_{\max}$. For large P, we obtain $t_{\text{diff}} \propto P \log(P)$. This condition turns out to be sufficient when confronted with the numerical data. In fact, T_{ns}^1 turns out to be very well fitted by a function of the type $c_1 P + c_2 P \ln P$.

We can perform a similar analysis for the syntactic model with one concept. In this case, we are implementing a game where each agent is represented by two inventories, which follows the rules presented in the previous section. From the results of our simulations, we can observe that the crucial features that determine the scaling properties of the convergence times are also maintained in this scenario. In fact, as can be seen in Fig. 3, $\max[N_{\text{tot}}] \propto P$, $t_{\max}$ scales linearly with P and $R_S(t)$ continues independent of P. As the same arguments produced for estimating T_{ns}^1 continue to be valid, we can expect the T_s^1 dependence on P to have the same functional form of T_{ns}^1. Fitting our numerical data with such a function (see Fig. 4), we obtain the following scaling relation for the convergence time for a syntactic model with one concept: $T_s^1(P) \approx -213.2P + 52.7P \ln P$.

B. Many-concepts model

First, it is useful to introduce a quantity, the total convergence time T^{tot}. This is the time when the system reaches the consensus state, corresponding to one shared expression for each concept and for all the players. Equivalently, it is the time when every communication event, relative to any one of the possible concepts, is a success.

Given an experiment set up so that C concepts are exchanged, if there is no syntax, the behavior described for the basic game with no syntax is reproduced for each of the concepts, and so it is easily generalized. In fact, the dynamics of each concept is obviously independent of those of the other concepts. This fact implies that if there is no syntax, $T_{ns}^{\text{tot}} = C T_{ns}^1(P) = S^2 T_{ns}^1(P)$.

This is not the case if we introduce syntax. In this case, as we stated before, we consider the situation in which the matrix $C_{ij} = S \times S$ represents all the exchanged concepts. We explore numerically the behavior of the total convergence time as a function of the number of concepts. We assume as a null hypothesis that T_s^{tot} depends linearly on the dimension of the matrix C_{ij}. As presented in Fig. 5, the analysis of the data from simulations suggests that this scaling relation is satisfied. Moreover, for $S \geqslant 2$, we can express the regression coefficients for different population sizes with a simple ex-

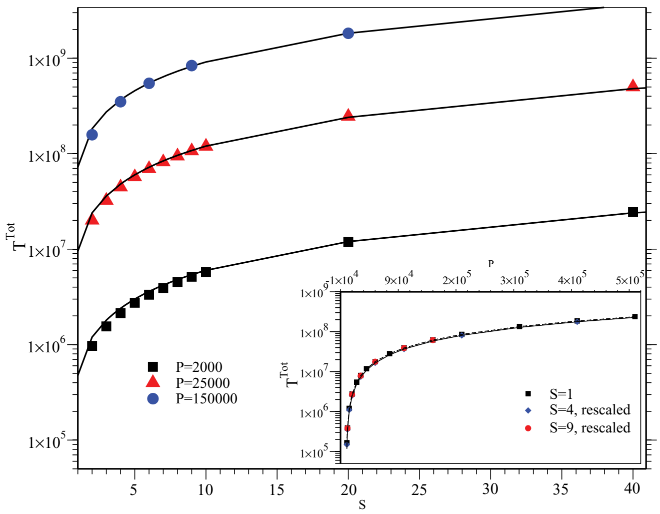

FIG. 5. (Color online) The convergence time as a function of S for different populations. The linear regressions are very well approximated by the general expression $T_s^{\text{tot}} \approx 1.5ST_s^1(P)$ (the continuous lines). The inset shows the convergence time as a function of P for different S values. The convergence times for $S = 4{,}9$ are rescaled by the factor $1.5S$. The continuous lines are the rescaled curves $c_1P + c_2P\ln P$ obtained from the regressions. The rescaling transformation produces a good collapse of data and curves. Data are averaged over up to 1000 simulations.

pression that uses $T_s^1(P)$: $T_s^{\text{tot}} \approx 1.5ST_s^1(P)$. This expression is derived from a numerical analysis of the data obtained from simulations. Some results supporting this scaling are reported in Fig. 5.

Starting from the scaling relation for the convergence time in dependence on P and S, we can determine the different behavior generated by the introduction of syntax. Quantifying this convergence time allows us to determine the strategy that enables a more effective communication. In fact, reaching consensus at a collective level corresponds to an efficient communication at an individual level. Depending on the number of concepts exchanged and on the population size, we can determine if the total convergence time for a model with syntax is shorter than that for a model without syntax. This situation is attained if $1.5ST_s^1(P) < S^2T_{ns}^1(P)$. Using this estimation, we are able to determine a critical value of S for which the emergence of syntax is viable: $S > 1.5T_s^1(P)/T_{ns}^1(P)$.

From this relation it follows that, if the number of exchanged concepts is sufficiently large in relation to the population size, the syntactic model is able to generate a faster convergence toward consensus. The dependence on the population size is very weak. For a population of 2000 individuals, $S = 8$ is sufficient for the conventionalization of syntax, and for a population 100 times larger, it is sufficient to select $S = 10$. So, from an empirical point of view, for typical populations, the relevant factor is simply the number of exchanged concepts, an interesting fact that enhances the possibility of syntax to emerge as an auto-organized process. The results obtained using our approximation were confirmed by different numerical simulations. In Fig. 6, we present an example for $P = 25\,000$. In this case, the matrix dimension should be bigger than 8, and effectively, from our simulation, if we exchange 81 concepts ($S = 9$), the syntactic model clearly performs better than the nonsyntactic model. In others words, the introduction of syntax generates a social communicative advantage when language must cope with a lot of concepts and when it is employed in smaller communities. In this context, the transition from nonsyntactic to syntactic communication is socially favored.

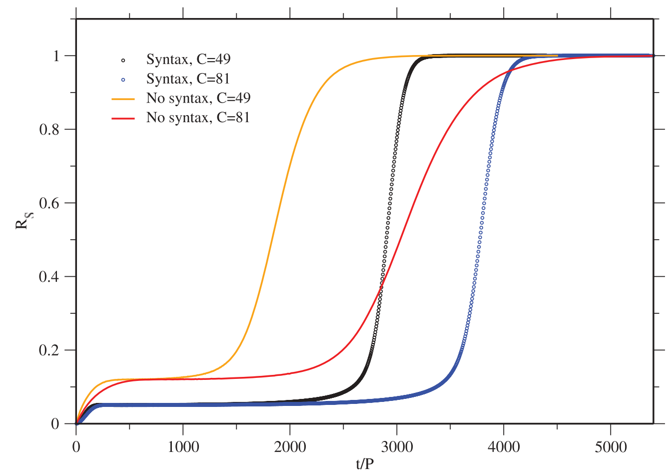

FIG. 6. (Color online) Temporal evolution of the success rate for a different number of exchanged concepts (C) for a model without syntax or with syntax. For $C \geqslant 81$, $T_s^{\text{tot}} < CT_{ns}^1$, which means a faster convergence for population using syntax. $P = 25\,000$

IV. CONCLUSIONS

We formulate a simple framework to explore the possibility of the emergence of an elementary syntactic structure fixed by the social dynamics defined by communication. We start with a version of the Naming Game model generalized for exchanging many different concepts. A simple syntactic structure is introduced in the form of a binary combination process, and the algorithm for fixing this structure is inspired by some known results relative to the individual learnability of linguistic structures. In this way, we can analyze the transition between syntactic and nonsyntactic communication on the basis of the social communicative potential of a linguistic structure, and not on the basis of the individual fitness or the velocity of individual learning.

From the analysis of this model, we can show that, under certain conditions, syntactic communication can reach consensus more efficiently than nonsyntactic communication, even if the task of fixing syntactic structure is more difficult. We are able to show some critical values for the number of exchanged concepts in dependence on the population size for which the emergence of syntax is viable.

ACKNOWLEDGMENTS

I am grateful to Marcio Argollo de Menezes and David Eduardo Zambrano Ramírez for fruitful discussions. I thank the anonymous referees for constructive and helpful comments.

[1] N. Chomsky, *Rules and Representations* (Blackwell, Oxford, 1980).
[2] N. L. Komarova and M. A. Nowak, Proc. R. Soc. London, Ser. B **268**, 1189 (2001); M. A. Nowak, N. L. Komarova, and P. Niyogi, Nature (London) **417**, 611 (2002).
[3] N. Chomsky, Ling. Inq. **36**, 1 (2005).
[4] N. Chater, F. Reali, and M. H. Christiansen, Proc. Natl. Acad. Sci. USA **106**, 1015 (2009).
[5] A. S. Hsu, N. Chater, and P. Vitányi, Cognition **120**, 380 (2011).
[6] L. Pearl and A. Weinberg, Lang. Learn. Dev. **3**, 43 (2007).
[7] M. H. Christiansen and N. Chater, Behav. Brain Sci. **31**, 489 (2008).
[8] F. Reali and T. L. Griffiths, Proc. R. Soc. London, Ser. B **277**, 429 (2010).
[9] L. Steels, Artificial Life **2**, 319 (1995); Trends. Cogn. Sci. **7**, 308 (2003).
[10] L. Steels, *The Talking Heads Experiment* (Laboratorium, Antwerpen, 1999), Vol. 1.
[11] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, and L. Steels, J. Stat. Mech. (2006) P06014.
[12] A. Baronchelli, T. Gong, A. Puglisi, and V. Loreto, Proc. Natl. Acad. Sci. USA **107**, 2403 (2010).
[13] A. Puglisi, A. Baronchelli, and V. Loreto, Proc. Natl. Acad. Sci. USA **105**, 7936 (2008).
[14] D. Bickerton and E. Szathmáry, *Biological Foundations and Origin of Syntax* (MIT Press, Cambridge, MA, 2009).
[15] N. Chomsky, *The Minimalist Program* (MIT Press, Cambridge, MA, 1995).
[16] L. Progovac and J. L. Locke, Biolinguistics **3**, 337 (2009).
[17] D. Bickerton, *Bastard Tongues* (Hill & Wang, New York, 2008).
[18] M. A. Nowak, J. B. Plotkin, and V. A. A. Jansen, Nature (London) **404**, 495 (2000).
[19] J. R. Safran *et al*., Science **274**, 1926 (1996).
[20] M. Peña, L. L. Bonatti, M. Nespor, and J. Mehler, Science **298**, 604 (2002).
[21] E. Brigatti and I. Roditi, New J. Phys. **11**, 023018 (2009).
[22] E. Brigatti, Phys. Rev. E **78**, 046108 (2008).
[23] A. Baronchelli, V. Loreto, and L. Steels, Int. J. Mod. Phys. C **19**, 785 (2008).