Journal of Mathematical Psychology 61 (2014) 19–24

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

# Language evolution, coalescent processes, and the consensus problem on a social network

Hariharan Narayanan [a,*], Partha Niyogi [b]

[a] Departments of Statistics and Mathematics, University of Washington, United States
[b] Department of Computer Science, University of Chicago, United States

## HIGHLIGHTS

- We prove that agents arrive at a common lexicon in a new model of language evolution in a social network.
- The rate of convergence depends on the "spectral gap" of the graph.
- The proof hinges upon a novel relation to coalescent processes, usually seen in the context of population genetics.

## ARTICLE INFO

*Article history:*
Received 3 January 2012
Received in revised form
23 April 2014
Available online 16 September 2014

*Keywords:*
Cognitive science
Neuroscience
Learning theory

## ABSTRACT

In recent times, there has been an increased interest in theories of language evolution that have an applicability to the study of dialect formation, linguistic change, creolization, the origin of language, and animal and robot communication systems in general. One particular question that has attracted some interest has the following general form: *how might a group of linguistic agents arrive at a shared communication system purely through local patterns of interaction and without any global agency enforcing uniformity?* In this paper, we consider a natural model of language (or more precisely, word) evolution on a social network, prove several theoretical properties, and establish connections to related phenomena in biology, social sciences, and physics.



## 1. Introduction

In recent times, there has been an increased interest in theories of language evolution that have an applicability to the study of dialect formation, linguistic change, creolization, the origin of language, and animal and robot communication systems in general (see Hauser, 1997, Kirby, 1999, Niyogi, 2006 and references therein). One particular question that has attracted some interest has the following general form: *how might a group of linguistic agents arrive at a shared communication system purely through local patterns of interaction and without any global agency enforcing uniformity*? The linguistic agents in question might be humans, animals, or machines in a multi-agent society. For an example of interesting simulations that suggest how a shared vocabulary might emerge in a population, see Baronchelli, Felici, Caglioti, Loreto, and Steels (2006), Baronchelli, Loreto, and Steels (2008), de Boer (2005), Steels (1999), Steels and Mcintyre (1999) among others. In this paper, we consider a generalization of Liberman's model, prove several theoretical properties, and establish connections to related phenomena in biology, social sciences, and physics.

Our model is as follows. For simplicity, we consider how a common word for a particular concept might emerge through local interactions even though the agents had different initial beliefs about the word for this concept. For example agents might use the phonological forms "dog", "kukur", "farama" etc. to describe the concept of a canine animal. Thus we imagine a situation where every time an event in the world occurs that requires the agents to use a word to describe this event, they may start out by using different words based on their initial belief about the word for this event or object. By observing the linguistic behavior of their neighbors agents might update their beliefs. The question is—will they eventually arrive at a common word and if so how fast.

### 1.1. Model

1. Let **W** be a set of words (phonological forms, codes, signals, etc.) that may be used to denote a certain concept (meaning or message).

* Corresponding author.
*E-mail addresses:* harius80@gmail.com, harin@uw.edu (H. Narayanan).

http://dx.doi.org/10.1016/j.jmp.2014.07.002
0022-2496/© 2014 Published by Elsevier Inc.

2. Let each agent hold a belief that is a probability measure on **W**. At time $t$, we denote the belief of agent $i$ to be $b_i^{(t)}$.
3. Agents are on a communication network which we model as a weighted directed graph where vertices correspond to agents. We further assume that the weight of each directed edge is positive and that there exists a directed path from any node to any other. An agent (say $i$) can only observe the linguistic actions of its out-neighbors, i.e. nodes to which a directed edge points from $i$. We denote weight of the edge from $i$ to $j$ by $A_{ij}$.
4. The update protocol for the $b_i^{(t)}$ as a function of time is as follows:
   (a) At each time $t$, each agent $i$ chooses a word $w = w_i^{(t)} \in$ **W** (randomly from its current belief $b_i^{(t)}$) and produces it. Let $X_i^{(t)}$, denote the probability measure concentrated at $w_i^{(t)}$. Since $w_i^{(t)}$ is a random word $X_i^{(t)}$ is correspondingly a random measure.
   (b) At every point in time, each agent can observe the words that their neighbors produce but they have no access to the private beliefs of these same neighbors.
   (c) Let $P$ be the matrix whose $ij$th entry satisfies
   $$P_{ij} = \frac{A_{ij}}{\sum_{k=1}^{n} A_{ik}}.$$
   At every time step, every agent updates its belief by a weighted combination of its current belief and the words it has just heard, i.e.,
   $$b_i^{(t+1)} = (1-\alpha)b_i^{(t)} + \alpha \sum_{j=1}^{n} P_{ij} X_j^{(t)},$$
   where $\alpha$ is a fixed real number in the interval (0, 1). We assume $A_{ii} = 0$ for each $i$.

At a time $t$, let the beliefs of the agents be represented by a vector

$$b^{(t)} := (b_1^{(t)}, \ldots, b_n^{(t)})^T.$$

Similarly, let the point measures on words $X_i^{(t)}$ be organized into a vector

$$X^{(t)} := (X_1^{(t)}, \ldots, X_n^{(t)})^T.$$

Then the reassignment of beliefs can be expressed succinctly in matrix form where the entries in the vectors involved are measures rather than numbers as

$$b^{(t+1)} = (1-\alpha)b^{(t)} + \alpha P X^{(t)}. \quad (1)$$

### 1.2. Remarks:

1. If beliefs were directly observable and agents updated based on a weighted combination of their beliefs and that of their neighbors,
   $$b^{(t+1)} = (1-\alpha)b^{(t)} + \alpha P b^{(t)}, \quad (2)$$
   the system has a simple linear dynamics, where all beliefs converge to a weighted average of the initial beliefs. Thus eventually, everyone has the same belief (see Bertsekas & Tsitsiklis, 1997 for pioneering work and Jackson, 2008 for a recent elaboration in an economic context).
2. Our focus in this paper is on the situation where the beliefs are *not observable* but only the linguistic actions $X_i^{(t)}$ are (and only to the immediate neighbors). Therefore, the corresponding dynamics follows a Markov chain. The state space of this chain (defined by Eq. (1)) is the set of all $n$-tuples of belief vectors. Since this is continuous, the standard mixing results with finite state spaces do not apply directly.

### 1.3. Results:

Our main results are summarized below.

1. With probability 1 (w.p.1), as time tends to infinity, the belief of each agent converges in total variation distance to one supported on a single word, common to all agents.
2. w.p.1, there is a finite time $T$ such that for all times $t > T$, all agents produce the same fixed word.
3. The rate at which beliefs converge depends upon the mixing properties of the Markov chain whose transition matrix is $P$.
4. The rate of convergence is *independent* of the size of **W**. One might think that a population where every agent has one of two words for the concept would arrive at a shared word faster than one in which every agent had a different word for the concept. This intuition turns out to be incorrect.
5. The proof of these results exposes a natural connection with coalescent processes and has a parallel in population genetics.
6. Our analysis brings out two different interpretations of the behavior of a linguistic agent. In the most direct interpretation, the agent's linguistic knowledge of the word is internally encoded in terms of a belief vector. This belief vector is updated with experience. In a second interpretation an agent's representation of its linguistic knowledge is in terms of a memory stack in which it literally stores every single word it has heard weighted by how long ago it heard it and the importance of the person it heard it from. Such an interpretation is consistent with exemplar theory. An external observer looking at this agent's linguistic actions will not be able to distinguish between these two different internal representations that the agent may have.

## 2. Convergence to a shared belief: quantitative results

We will define an auxiliary Markov Chain to model the exemplar based view of the evolution of the memory stack. We require the original $n$ states $S$ corresponding to agents and an additional $n$ states $\hat{S}$ to model whether a word was uttered at time $t$, or was embedded in the memory of some agent at that time. Fig. 1 depicts the evolution of a memory stack.

Let $\tilde{P}$ be the transition matrix on the state space $\tilde{S} = S \cup \hat{S}$, where for $i, j \in S := \{1, \ldots, n\}$ and $\hat{S} = \{\hat{1}, \ldots, \hat{n}\}$.

$$\tilde{P}(i \to j) = \tilde{P}(\hat{i} \to j) = \alpha P_{ij},$$
$$\tilde{P}(i \to \hat{i}) = \tilde{P}(\hat{i} \to \hat{i}) = 1 - \alpha$$

work

**Definition 1.** Let $T_{\mathrm{mix}}(\epsilon)$ denote the mixing time of $\tilde{P}$, defined as the smallest $t$ for which, for each specific choice of $v \in \tilde{S}$,

$$\sum_{u \in \tilde{S}} |\tilde{P}^{(t)}(v \to u) - \tilde{\pi}(u)| < \epsilon.$$

Here $\tilde{P}^{(t)}(b \to c)$ denotes the probability that a Markov Chain governed by $\tilde{P}$ starting in $b$ lands in $c$ at the $t$th time step. Also, the stationary distribution at $i$ is denoted $\tilde{\pi}(i)$.

The following is the main result of this paper.

**Theorem 1.** 1. *The probability that all agents produce the same word at times* $T, T+1, \ldots$ *tends to* 1 *as* $T$ *tends to* $\infty$. *More precisely, if*

$$\tau = (4n/\alpha^2) T_{\mathrm{mix}}\left(\frac{\alpha}{4}\right)$$

$$M = e,$$

*then*

$$\mathbb{P}[\forall_{\substack{t \geq T \\ u \in S}} X_u^t = X_1^T] > 1 - \frac{MnTe^{-\frac{T}{\tau}}}{1 - e^{-\frac{T}{\tau}}}. \quad (3)$$