Journal of Theoretical Biology 429 (2017) 142–148

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Parent oriented teacher selection causes language diversity

Ibrahim Cimentepe, Haluk O. Bingol*

Department of Computer Engineering, Bogazici University, TR-34342, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 20 February 2017
Revised 20 May 2017
Accepted 23 June 2017
Available online 27 June 2017

Keywords:
Language diversity
Evolution of language
Language learning model

ABSTRACT

An evolutionary model for emergence of diversity in language is developed. We investigated the effects of two real life observations, namely, people prefer people that they communicate with well, and people interact with people that are physically close to each other. Clearly these groups are relatively small compared to the entire population. We restrict selection of the teachers from such small groups, called imitation sets, around parents. Then the child learns language from a teacher selected within the imitation set of her parent. As a result, there are subcommunities with their own languages developed. Within subcommunity comprehension is found to be high. The number of languages is related to the relative size of imitation set by a power law.

1. Introduction

Language remains mystery in many aspects including how it emerged, how it evolved, and how it is learned (Bickerton, 2007; Bolhuis et al., 2014; Fitch, 2007; Gong and Shuai, 2012; Hauser et al., 2014; Lieberman et al., 2007; Nowak et al., 2002; Pagel et al., 2007; Pinker and Bloom, 1990). This is partly due to no agreed definition of language. A group of scientists, including Chomsky, believe that "communication cannot be equated with language" (Bolhuis et al., 2014). Yet another group considers language as a means to transfer meanings between individuals through signaling structures (Gong and Shuai, 2012; Kirby et al., 2007; Nowak et al., 2002). Assuming that language provides an evolutionary advantage, some evolutionary models are proposed (Cangelosi and Parisi, 1998; Krakauer, 2001; Nowak and Komarova, 2001; Nowak and Krakauer, 1999; Nowak et al., 2000; 1999; Pinker and Bloom, 1990; Plotkin and Nowak, 2000; Tzafestas, 2008), some of which are game theoretical (Mitchener and Nowak, 2004). Information theoretical approaches predict that not only symbols but word formation is necessary in order to have efficient communication, which leads to basic grammatical rules (Nowak and Krakauer, 1999; Plotkin and Nowak, 2000). There are also empirical approaches to language evolution (Kirby et al., 2008; Lieberman et al., 2007; Pagel et al., 2007).

It is believed that language evolves within generation and while it is transferred from generation to generation. One of the critical issues, which includes rich discussions on universal grammar, is how language is learned by the new generation (Bolhuis et al., 2014; Hauser et al., 2014; Lieberman et al., 2007; Niyogi and Berwick, 1996; Nowak et al., 2001; Pagel et al., 2007; Pinker and Bloom, 1990).

Individuals may imitate each other or prefer to imitate experienced members in population (McEwen, 2007). It may be the case that one learns language by means of imitation. If it is so, who should serve as teachers in community for the next generation? And which imitation strategies can be applied to the population that lead to emergence of language that is shared locally or across population?

1.1. Motivation

In this study, we use and extend the mathematical framework that is already established (Hurford, 1989; Nowak and Krakauer, 1999; Nowak et al., 1999). Our extension leads us to emergence of diversity in language. Language diversity is very popular indeed; it is even addressed in the well-known story of the Tower of Babel. According to the story, the people who speak the same language were once scattered all around the world so that they could no longer understand each other.

One expects that a child can learn language from her neighbors in the society. The neighborhood includes her parents, her kinship network, territoriality, and labor roles (Krakauer, 2001). Ref. Nowak et al. (1999) considers language as a culturally transmitted entity where *cultural transmission* is defined to be a type of transmission where socially obtained information is passed on, in form of teaching. Three types of neighborhoods, for child to learn a language, are investigated. (i) In the *parental learning*, asexually produced child learns from her parent. An agent reproduces proportionally to its mutual comprehension, which will be defined shortly, with the

* Corresponding author.
E-mail address: bingol@boun.edu.tr (H.O. Bingol).

http://dx.doi.org/10.1016/j.jtbi.2017.06.032
0022-5193/

rest of the population. Therefore the agent who better fits to the population language-wise has better chances to transfer her language to agents of the next generation. (ii) The *role model learning* is based on reputation. An agent with a higher reputation is imitated more. Therefore it is not important whose offspring it is, a child imitates agents who comprehend better. So the language of an agent with better mutual comprehension is transferred more. In this learning type, T teachers are selected proportional to their mutual comprehension and child learns from them. It is observed that higher values of T produce higher mutual comprehension although it takes longer for system to settle down. (iii) In the *random learning* there is no structure. A child randomly selects an agent in the population as her teacher. That is, mutual comprehension has no role in teacher selection.

In this work, we investigate two new teacher selection strategies. A child is born to her parent. So her teacher has to be related to her parent if not the parent itself. Considering the parent, there are two possible circles of friends. (i) We assume that one is surrounded by people that understand each other well. In the context of language, parent's friends should be the ones that have high mutual comprehension. (ii) Since we all live in a physical environment, our friends should not be physically too far from us. If we assume that agents are located on the nodes of 1D ring lattice, friends should be the ones within close proximity to the parent. In this paper, we modify teacher selection to investigate these two cases.

2. Background

We revisit the language model developed by Ref. Hurford (1989), Nowak and Krakauer (1999) and Nowak et al. (1999) with a slightly modified notation. Then we go over k-means clustering algorithm (Duda et al., 2012; MacKay, 2003; Theodoridis et al., 2010). Finally we adapt k-means to language domain and use it to identify language subcommunities.

2.1. Language model

We model language communication in a very simple way, called *proto-language*, as follows: Let $\mathcal{P}$ be a set of N agents. An agent i thinks of a meaning μ and wants to pass it to agent j. Since she does not have means to pass a meaning in her mind directly to the mind of j, she has to use signals. She selects a signal x, which she thinks as a representation of μ, and passes the signal to j. We assume that there is no noisy channel, i.e., one receives exactly what is sent. Receiving x, j tries to interpret x in his own way. Hopefully j will interpret it as μ.

Clearly, mappings from μ to x at i and from x back to μ at j are very important for a successful communication. We need to specify how association of meaning and signal in sending and receiving ends are done. Suppose every agent has her own statistics $a_{\mu x}$ of how frequently she uses signal x to mean meaning μ. Assuming that there are M meanings and S signals, we have an $M \times S$ *association matrix* $\mathbf{A} = [a_{\mu x}]$, for each agent, from which we can derive encoding and decoding methods. *Encoding matrix*, $\mathbf{E} = [e_{\mu x}]$, is an $M \times S$ matrix where $e_{\mu x}$ is the probability of using signal x for meaning μ. *Decoding matrix*, $\mathbf{D} = [d_{x\mu}]$, on the other hand, is an $S \times M$ matrix where $d_{x\mu}$ is the probability of understanding meaning μ for given signal x.

The encoding and decoding matrices can be obtained from the association matrix as follows:

$$e_{\mu x} = \frac{a_{\mu x}}{\sum_{x'=1}^{S} a_{\mu x'}}, \quad d_{x\mu} = \frac{a_{\mu x}}{\sum_{\mu'=1}^{M} a_{\mu' x}}.$$

We will focus on $\mathbf{A}$ for language learning since $\mathbf{E}$ and $\mathbf{D}$ can be derived from $\mathbf{A}$.

2.1.1. Comprehension

Suppose agent i wants to pass meaning μ to agent j. Probability of doing this correctly is

$$\sum_{x=1}^{S} e_{\mu x}^{(i)} d_{x\mu}^{(j)}$$

where $e_{\mu x}^{(i)}$ and $d_{x\mu}^{(j)}$ are encoding of i and decoding of j, respectively. When we average that over all meanings, we obtain *comprehension* from i to j, that is

$$F(i \to j) = \frac{1}{M} \sum_{\mu=1}^{M} \sum_{x=1}^{S} e_{\mu x}^{(i)} d_{x\mu}^{(j)}.$$

If we want them to communicate both ways, we consider *mutual comprehension*

$$F(i \leftrightarrow j) = \frac{F(i \to j) + F(j \to i)}{2}.$$

Now, let's consider comprehension within a community $\mathcal{C} \subseteq \mathcal{P}$. *Within community comprehension* is defined as the average comprehension in a community $\mathcal{C}$. Thus,

$$W(\mathcal{C}) = \frac{1}{2\binom{|\mathcal{C}|}{2}} \sum_{i \in \mathcal{C}} \sum_{\substack{j \in \mathcal{C} \\ j \neq i}} F(i \leftrightarrow j).$$

Within community comprehension of the entire population, i.e., $W(\mathcal{P})$, is called *overall comprehension*.

2.1.2. Learning model

The evolution of language can happen in two different ways. Language evolves both through agents interacting with each other within a generation, and as it is transferred from one generation to the next by means of learning. We follow the latter form as given in Ref. Nowak et al. (1999).

At each generation, population is replaced with new set of N agents. Agents of new generation have no meaning-signal associations initially. That is, the association matrices of agents are empty. For language to be transferred from the generation of parents to the generation of children, some agents are assumed to be chosen as *teachers*.

In Ref. Nowak et al. (1999), teacher selection is a result of fitness gains. Fitness of an agent is directly related to her ability to communicate with overall population. Specifically, the *fitness* of agent i is defined as

$$F_i = \sum_{j \in \mathcal{P}} F(i \leftrightarrow j).$$

For the next generation, offspring are produced proportional to the fitness of an agent: the chance that a particular agent arises from agent i is proportional to

$$\frac{F_i}{\sum_{j \in \mathcal{P}} F_j}.$$

That is, each child agent selects her teacher according to this probability distribution. Thus, agents who have better fitness are picked more. In Ref. Nowak et al. (1999), it is stated that more than one teacher could be assigned for each child agent. This case is examined as a form of cultural learning, where some elite group of agents is responsible for transition of language. It is reported that since the selection mechanism remains the same, total number of teachers assigned only affects how fast the language emerges in such populations (Nowak et al., 1999).

After teachers of the next generation are assigned, language is transferred from teacher to child. The learning process between the child and her teacher is similar to a naming game (Steels, 1995). Child learns the language of her teacher by sampling their