

Prediction of Creole Emergence in Spatial Language Dynamics

Makoto Nakamura¹, Takashi Hashimoto², and Satoshi Tojo¹

¹ School of Information Science,
Japan Advanced Institute of Science and Technology
1-1, Asahidai, Nomi, Ishikawa, 923-1292, Japan

² School of Knowledge Science,
Japan Advanced Institute of Science and Technology
1-1, Asahidai, Nomi, Ishikawa, 923-1292, Japan
{mnakamur,hash,tojo}@jaist.ac.jp

Abstract. Creole is a new born language emerging in most cases where language contact takes place. Simulating behaviors that creole communities are formed in some environments, we could contribute to actual proof of some linguistic theories concerning language acquisition. Thus far, a simulation study of the emergence of creoles has been reported in the mathematical framework. In this paper we introduce a spatial structure to the framework. We show that local creole communities are organized, and creolization may occur when language learners learn often from non-parental language speakers, in contrast to the non-spatial model. The quantitative analysis of the result tells us that emergence of local colonies at the early stage tends to induce the full creolization.

1 Introduction

Computer simulation of diachronic change in human languages has widely been reported in the study of language evolution [1,2], where interactions among individuals affect language spoken throughout the community, dependent upon the abilities of individuals or the learning environment. Among those simulations, the emergence of pidgins and creoles is one of the most interesting phenomena in language change [3,4,5].

Pidgins are simplified tentative languages spoken in multilingual communities, which come into being where people need to communicate but do not have a language in common. On the other hand, creoles are full-fledged new languages based on the pidgins in later generations. For example, Hawaiian Creole English emerged among plantation workers coming from Hawaii, China, the Philippines, Japan, Korea, Portugal, Puerto Rico and so on in the 19th century to the beginning of the 20th century. Since they needed to communicate with farm owners, they first formed Hawaiian pidgin based on English; later their offspring immersed with the pidgin had developed the language to one with its own grammatical structure. In general, grammar of a creole is different from any contact languages, although its vocabulary is often borrowed from them. Our goal in this paper is to discover specific conditions under which creoles emerge.

Thus far, we proposed a mathematical framework for the emergence of creoles [6] based on the language dynamics equation by Nowak et al. [7], showing that creoles become dominant under specific conditions of similarity among languages and linguistic environment of language learners. Our purpose in the present study is to introduce a *spatial structure* to our model, in order to observe self-organization process of creole community. Especially, in this paper we compare behaviors of the two models. A related work for introducing a spatial structure into a mathematical model of language change has been done by Castelló et al. [8], who have analyzed a spatial version of a mathematical framework by Abrams et al. [9]. Different from Abrams-Strogatz's model, our model [6] is well-defined in terms of learning algorithm and a learning environment.

Introducing a spatial structure to a mathematical framework, we expect to observe a process of creolization and then to obtain more precise conditions from the model more similar to the environment where actual language phenomena took place. We recognize this study to fill the gap between the study of multi-agent models and mathematical models.

In Section 2, we describe the modified language dynamics model and a learning algorithm, and in Section 3 we define a creole in population dynamics. Section 4 reports our experiments, and we conclude in Section 5.

2 Population Dynamics for the Emergence of Creole

In this section, we briefly explain how to divert a mathematical model proposed by Nakamura et al. [6] to the one with a spatial structure.

The most remarkable point in the model of Nakamura et al. [6] is to introduce an *exposure ratio* α , which determines how often language learners are exposed to a variety of language speakers other than their parents. They modified the learning algorithm of Nowak et al. [7], taking the exposure ratio into account to model the emergence of creole community. Nakamura et al. [6] have shown that a certain range of α is necessary for a creole to emerge.

2.1 Language Dynamics Equation for the Emergence of Creole

In response to the language dynamics equation by Nowak et al. [7], Nakamura et al. [10] assumed that any language could be classified into one of a certain number (n) of grammars. Thus, the population of language speakers is distributed to $\{G_1 \dots G_n\}$. Let x_i be the proportion of speakers of G_i within the total population. Then, the language dynamics is modeled by an equation governing the transition of language population.

Because Nowak et al. [7] assumed that language speakers bore offspring in proportion to their successful communication, they embedded a fitness term in their model which determined the birth rate of each language group. The model for creolization has excluded the biological fitness, on the assumption that in the real world creoles did not emerge because creole speakers had more offspring than speakers of other pre-existing languages, that is:

$$\frac{dx_j(t)}{dt} = \sum_{i=1}^n \bar{q}_{ij}(t)x_i(t) - x_j(t) . \quad (1)$$

In the language dynamics equations, the similarity matrix S and the transition matrix $\bar{Q}(t)$ play important roles: the similarity matrix $S = \{s_{ij}\}$ is defined as a probability that a sentence of G_i is accepted also by G_j . Children learn a language in accordance with a learning algorithm, in which the accuracy varies depending on the similarity among languages. The transition matrix $\bar{Q}(t) = \{\bar{q}_{ij}(t)\}$ is defined as a probability that a child of G_i speaker acquires G_j , and is calculated based on the learning algorithm. Being different from the definition by Nowak et al. [7], the definition of $\bar{Q}(t)$ depends on the generation parameter t , as well as the S matrix and a learning algorithm.

2.2 Introducing Spatial Structure

In the spatial model, we use the language distribution in neighbors surrounding each agent to calculate the local transition probability \bar{Q} , by which each agent acquires a language, while a child is exposed to the whole population in a non-spatial model.

Hereafter, replacing x_i for $x_i^{(l)}$ as a population rate of G_i speakers surrounding an agent at location l , equations are applicable to the spatial structure, too. We calculate $\bar{Q}^{(l)}(t)$ for each agent every generation.

2.3 Learning Algorithm

In some communities, a child learns language not only from his/her parents but also from other adults, whose language may be different from the parental one. In such a situation, the child is assumed to be exposed to other languages, and thus may acquire the grammar most efficient in accepting multiple language input. In order to assess how often the child is exposed to other languages, we divide the language input into two categories: one is from his/her parents, and the other is from other language speakers. We name the ratio of the latter to the total amount of language input an *exposure ratio* α . This α is subdivided into smaller ratios corresponding to those other languages, where each ratio is in proportion to the population of the language speakers. An example distribution of languages is shown in Figure 1(a). Suppose a child has parents who speak G_p , s/he receives input sentences from G_p on the percentage of the shaded part, $\alpha x_p + (1 - \alpha)$, and from non-parental languages $G_i (i \neq p)$ on the percentage, αx_i .

We have adopted a batch learning algorithm, which resolves Niyogi [11]'s problem regarding an unrealistic Markov structure which implies that some children cannot learn certain kinds of language. From the viewpoint of universal grammar, that all conceivable grammars of human beings are restricted to a finite set [12], language learning is considered as a choice of a plausible grammar from them. The following algorithm realizes such learning as: 1) In a child's memory, there is supposed to be a score table of grammars. 2) The child receives a sentence uttered by an adult. 3) The acceptability of the sentence is tested using

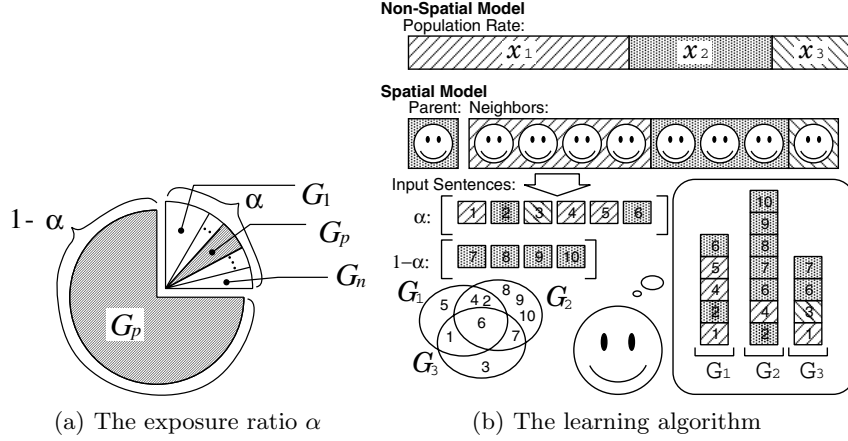


Fig. 1. The learning algorithm including the exposure ratio α

each grammar. The grammar which accepts the sentence scores one point. 4) Steps 2) and 3) are repeated until the child receives a fixed number (w) of sentences, which is regarded as sufficient for the decision of the grammar selection. 5) The child adopts the grammar with the highest score.

The child is exposed to utterances of adult speakers of each language, the percentage of which is determined by the distribution of population and the exposure ratio α , while the S matrix determines the acceptability of a sentence. In Figure 1(b), we show an example where a child of G_2 speaker obtains G_2 after exposure to a variety of languages. The child receives sentences, which are boxes numbered from 1 to 10. The input sentences are divided into two sets according to the exposure ratio α . One of the sets consists of sentences of all grammars. The number of the sentences of each language is proportional to the population share of the language speakers. For example, the child hears sentences 1, 4 and 5 uttered by G_1 speakers. The other consists of sentences of his/her parents. Therefore, these sentences are acceptable by a particular grammar. Because his/her parental grammar is G_2 , for example, the sentences 7 to 10 are randomly chosen from the language of G_2 . The child counts acceptable sentences for each grammar. The sentence 1 can be accepted by G_3 as well as G_1 , while it is uttered by a G_1 speaker. The Venn diagram in Figure 1(b) represents that each language shares sentences with others. In this case, because the sentence 1 is acceptable both by G_1 and by G_3 , the child adds 1 to both of the counters in his/her mind.

2.4 Revised Transition Probability

Suppose that children hear sentences from adult speakers depending on the exposure ratio and on the distribution of population. A probability that a child whose parents speak G_i accepts a sentence by G_j is expressed by:

$$U_{ij} = \alpha \sum_{k=1}^n s_{kj} x_k + (1 - \alpha) s_{ij} . \quad (2)$$

After receiving a sufficient number of sentences for language acquisition, the child will adopt the most plausible grammar, as estimated by counting the number of sentences accepted by each grammar. This learning algorithm is simply represented in the following equation. Exposed to a variety of languages in proportion to the population share of adult speakers, children whose parents speak G_i will adopt G_{j^*} by:

$$j^* = \underset{j}{\operatorname{argmax}} \{U_{ij}\} . \quad (3)$$

When the children hear w sentences, a probability that a child of G_i speaker accepts r sentences with G_j is given by a binomial distribution,

$$g_{ij}(r) = \binom{w}{r} (U_{ij})^r (1 - U_{ij})^{w-r} . \quad (4)$$

On the other hand, a probability that the child accepts less than r sentences with G_j is

$$h_{ij}(r) = \sum_{k=0}^{r-1} \binom{w}{k} (U_{ij})^k (1 - U_{ij})^{w-k} . \quad (5)$$

From these two probability distributions, the probability that a child of G_i speaker accepts k sentences with G_j , while less than $k - 1$ sentences with the other grammars, comes to $g_{ij}(k) \prod_{l=1, l \neq j}^n h_{il}(k)$. For a child of G_i speaker to acquire G_j after hearing w sentences, G_j must be the most efficient grammar among n grammars; viz., G_j must accept at least $\lceil \frac{w}{n} \rceil$ sentences. Thus, the probability \bar{q}_{ij} becomes the sum of the probabilities that G_j accepts $w, w-1, \dots, \lceil \frac{w}{n} \rceil$ sentences. Because each of the sentences is uttered by a speaker and is accepted by at least one grammar, there must be a grammar which accepts $\lceil \frac{w}{n} \rceil$ or more out of w sentences. Thus, if G_j accepts less than $\lceil \frac{w}{n} \rceil$ sentences, the child does not acquire G_j . Therefore, \bar{q}_{ij} becomes:

$$\bar{q}_{ij}(t) = \frac{\sum_{k=\lceil \frac{w}{n} \rceil}^w \left\{ g_{ij}(k) \prod_{\substack{l=1 \\ l \neq j}}^n h_{il}(k) + R(k, n) \right\}}{\sum_{m=1}^n \left[\sum_{k=\lceil \frac{w}{n} \rceil}^w \left\{ g_{im}(k) \prod_{\substack{l=1 \\ l \neq m}}^n h_{il}(k) + R(k, n) \right\} \right]} , \quad (6)$$

where $R(k, n)$ is the sum total of the probabilities that the child would choose G_j when one or more other grammars accept the same number of sentences as G_j . When there are m candidate grammars including G_j , the probability becomes one divided by m . The following expression is an example when $n = 3$.

$$\begin{aligned} R_{ij}(k, 3) = & \frac{1}{3} \{ g_{ij}(k) g_{ij_2}(k) g_{ij_3}(k) \} \\ & + \frac{1}{2} \{ g_{ij}(k) g_{ij_2}(k) h_{ij_3}(k) + g_{ij}(k) h_{ij_2}(k) g_{ij_3}(k) \} \\ & (j_2, j_3 \in \{1, 2, 3\}, j \neq j_2, j_3, j_2 \neq j_3) \end{aligned} \quad (7)$$

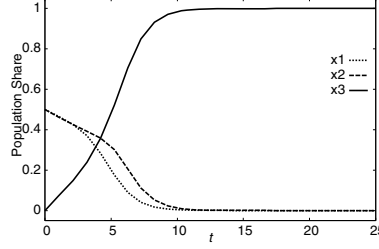


Fig. 2. Example of creolization $((a, b, c) = (0, 0.3, 0.4), w = 10, \alpha = 0.7)$

3 Creole in Population Dynamics

Creoles are considered as new languages. From the viewpoint of population dynamics, we define a creole as a transition of population of language speakers. A creole is a language which no one spoke in the initial state, but most people have come to speak at a stable generation. Therefore, creole is represented by G_c such that: $x_c(0) = 0, x_c(t) > \theta_c$, where $x_c(t)$ denotes the population share of G_c at a convergent time t , and θ_c is a certain threshold to be regarded as a dominant language. We set $\theta_c = 0.9$ through the experiments.

For convenience, we have mainly observed the behavior of the model using three grammars. The similarity matrix can be expressed as a symmetric matrix such that:

$$S = \begin{pmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{pmatrix}. \quad (8)$$

Here, we regard G_3 as a creole grammar, giving the initial condition as $(x_1(0), x_2(0), x_3(0)) = (0.5, 0.5, 0)$. Therefore, the element a denotes the similarity between two pre-existing languages, and b and c are the similarities between G_1 and the creole, and between G_2 and the creole, respectively.

We show an example of creolization in Figure 2. The parameters were set to $(a, b, c) = (0, 0.3, 0.4)$, $w = 10$, and $\alpha = 0.7$. Note that the conditions in the parameter space for dominant creoles are limited [13].

4 Experiments and Results

The spatial structure is a toroidal 50-by-50 square grid. Each agent has 8 neighbors. Each agent chooses one of three languages every generation, two of which, G_1 and G_2 , are pre-existing and randomly distributed with the same total number at the initial state. The remaining language, G_3 , is a creole, having a certain similarity between two languages. The similarity means the probability that a sentence uttered by a G_i speaker is accepted by G_j . In this paper, We take the following values: $(a, b, c) = (0, 0.3, 0.4)$, and $w = 10$ for the number of input sentences.

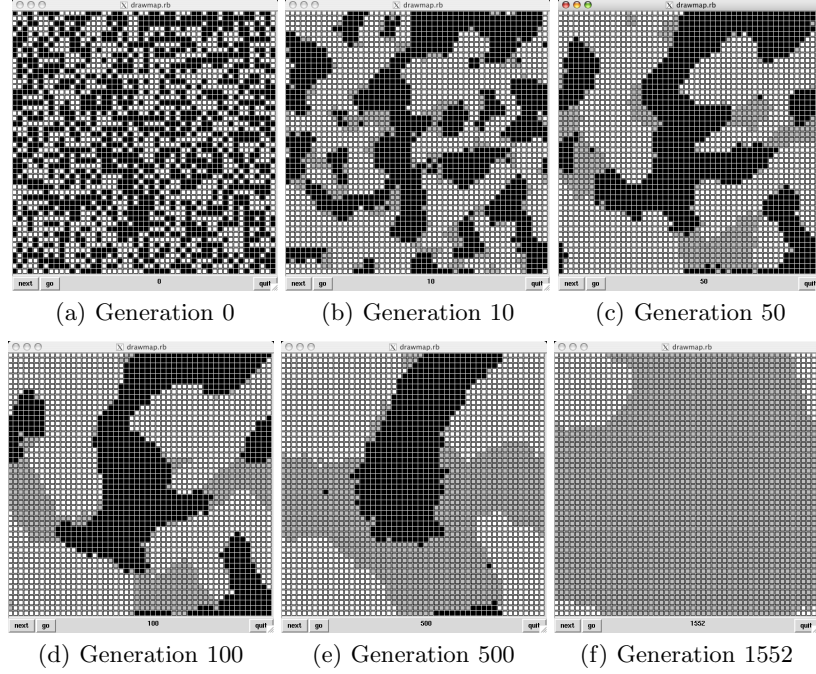


Fig. 3. Example of the spatial dynamics (white: G_1 , black: G_2 , gray: G_3 ; $\alpha = 0.7$)

4.1 Behaviors of Spatial Dynamics

We show an example of the spatial dynamics in Figure 3; (a) Only G_1 and G_2 are distributed at the initial stage. (b,c) Some local communities (hereafter colonies) of creole are organized at the early stage. (d,e) Both G_1 , G_2 and creole coexist at a quasi-stable stage. (f) In this trial, the creole eventually becomes dominant at Generation 1552. Agents surrounded by both G_1 and G_2 neighbors are likely to acquire the creole. In fact, creole speakers often appear on the border between communities. This is because the large value of α makes the agents to be exposed to both languages, and the creole is the most efficient for accepting input utterances from both languages.

In general, learners tend to form a colony, regardless of the languages, affected from its neighbors during learning acquisition. However, the smaller the value of α , the slower the forming colonies. This is because the learners are hardly affected by neighbors, hearing their mother tongue from their parents. Note that even if $\alpha = 0$, it is possible for a learner to acquire a language other than his/her mother tongue due to the similarity among languages. The number of input sentences is also relevant to forming a colony. Learners hearing a lot of language input become conservative in terms of changing his/her language. That is, once a small colony has been formed at a generation, inhabitants in the colony come to choose the same language as the previous generation. On the contrary, if learners choose a

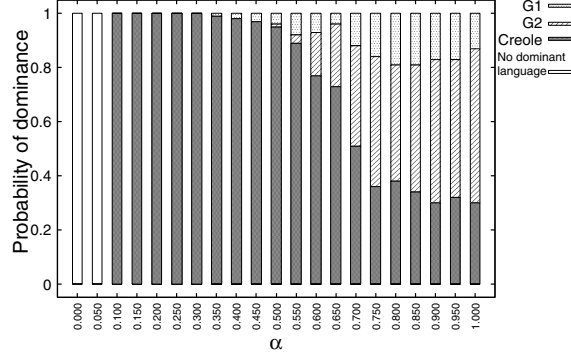


Fig. 4. Probability of dominant language in the spatial model

language with less language input, they tend to be undetermined in choosing a language. As a result, it is difficult to form a colony.

4.2 Comparing the Two Models with and without a Spatial Structure

We examine the probability of dominance for each language (Figure 4). Note that the spatial model is based on a stochastic dynamics. This graph is the result of 100 runs for 1,000,000 generations at each α value. The corresponding result in the non-spatial model is the population distribution at the stable generation, shown in Figure 5¹, since the non-spatial model is based on the deterministic dynamics. This parameter set makes creole dominant at the range $0.1 \lesssim \alpha \lesssim 0.8$. In the spatial model, the probability that the creole is dominant gradually decreases from $\alpha > 0.3$, and it becomes 0.3 around $\alpha > 0.8$.

In general, the larger the value of α , the more prominent the transition of population becomes, and in some cases, the transition leads to creolization. In Figure 5, however, the dominant language changes between the creole and G_2 at $\alpha \simeq 0.8$, and it is not always true that creoles are more likely to become dominant at the larger value of α (See [6]). Since children of a G_1 speaker become more exposed to G_2 in the larger value of α , it is possible for them to acquire G_2 directly instead of the creole, and vice versa. Therefore, the population of the creole remains small and the children of creole speakers are likely to acquire G_2 rather than G_1 . Thus, the dominant language changes to G_2 at $\alpha \gtrsim 0.8$ ².

These differences between the results shown in Figure 4 and Figure 5 can be understood by considering local interaction and stochastic dynamics. The pre-existing language may be able to form a colony due to stochasticity. Once a colony with certain size is formed, agents in the colony are surrounded by the

¹ In other words, the result of the population distribution at the stable generation in Figure 2 is plotted at $\alpha = 0.7$ in Figure 5.

² In the experiments, we have chosen this parameter set with which creoles tend to appear in the wide range of α .

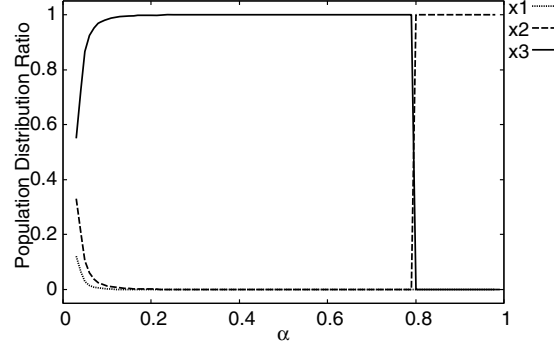


Fig. 5. Stable population distribution in the non-spatial model

same language and the exposure ratio effectively comes to $\alpha = 0$. This situation is hard for the creole speakers to organize a colony. Thus, the probability to be dominant is restrained by the pre-existing language at the middle-high range of α . At the higher range of α , the creole can organize a colony at the early stage with certain ratio through random migration. The colony can grow to the whole space.

4.3 Quantitative Analysis in Forming Creole Communities

Observing behaviors of the spatial dynamics, we realized that forming creole colonies at early stages plays a key role for creolization. Large creole colonies are difficult to vanish and are able to encroach upon a territory of other languages. In other words, if creole speakers fail to form a certain size of colony before the quasi-stable stage, it is difficult for creole to become dominant. Therefore, in this section we try to predict whether creolization takes place or not at the early stage, observing creole colonies quantitatively.

For a quantitative description of the emergence and dynamics of linguistic spatial domains we use the ensemble average interface density $\langle \rho \rangle$ as an order parameter, following a precedent work [8]. This is defined as the density of links joining nodes in the network which are in different states [14]. For associating with the population distribution, we use the inverse value as $\langle \bar{\rho} \rangle = \langle 1 - \rho \rangle$. The ensemble average, indicated as $\langle \cdot \rangle$, denotes average over realizations of the stochastic dynamics starting from different random distributions of initial conditions. As the time proceeds, the increase of $\bar{\rho}$ from its initial value describes the ordering dynamics, where linguistic spatial domains, in which agents are in the same state, grow in time. The maximum value $\bar{\rho} = 1$ corresponds to a stationary configuration in which all the agents belong to the same linguistic community. In addition, we defined $\bar{\rho}_i$ as the density of G_i speakers' neighbors which are in the same states.

We analyzed difference of the behaviors at the early stage among 100 trials every α , shown in Figure 4, classifying the trials into three, each of which denotes the corresponding language eventually becomes dominant. We recognize an early

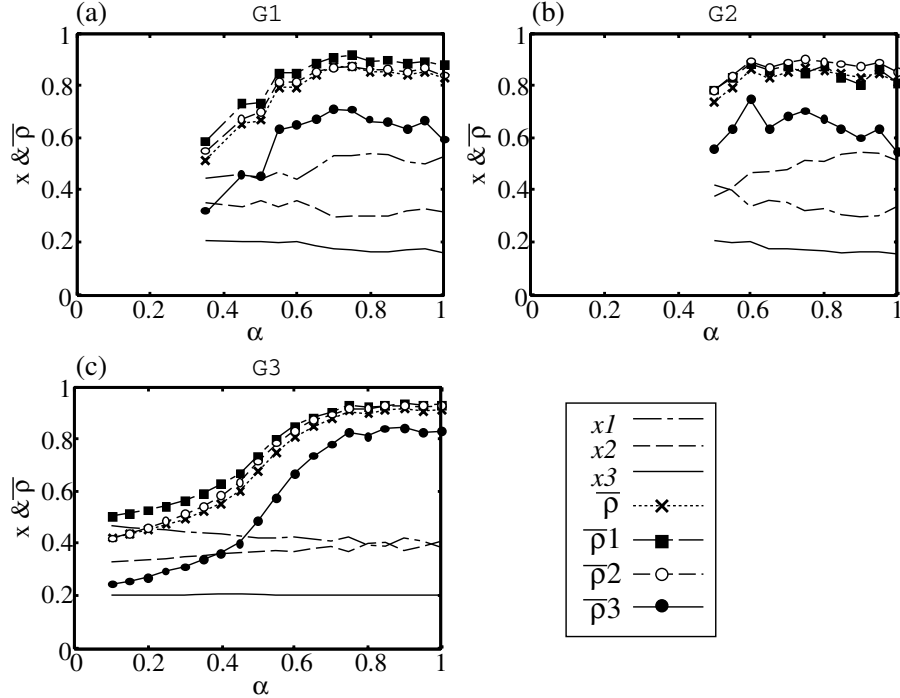


Fig. 6. Difference of x_i and $\langle \bar{\rho}_i \rangle$ at the early stage ($x_3 \simeq 0.2$) between results in which G_1 , G_2 , and G_3 eventually become dominant respectively

stage as the earliest generation at which $x_3(t)$ exceeds 0.2 in each trial. Figure 6 shows the difference of the average $\langle \bar{\rho} \rangle$, $\langle \bar{\rho}_i \rangle$, and x_i every α at the early stage between G_1 , G_2 and G_3 becoming dominant at the stable generation. At some values of α in Figure 6(a) and (b), the values of x_3 are partially plotted less than 0.2, because in some trials x_3 never exceeded 0.2. In this case, we took the values at the stable generation for the average calculation. At some values of α where no data are plotted, there was no trial in which the corresponding language became dominant. Therefore, Figure 6 corresponds to the frequency distribution of dominance, shown in Figure 4.

In Figure 6, we can see that $\langle \bar{\rho}_3 \rangle$ is lower than other densities at any value of α . Note that the average density of G_i denoted by $\langle \bar{\rho}_i \rangle$ is affected not only by a degree of forming colonies but also by its population³. Therefore, it is natural that only 20 percent of the population obtains the density smaller than others.

As was mentioned in Section 4.2, creolization takes place even at the large value of α ($0.8 \lesssim \alpha$), while G_2 dominates the community in the non-spatial model (Figure 5). We can see that the values of $\langle \bar{\rho}_3 \rangle$ in Figure 6(c) are higher than that of (a) and (b). This tendency can be seen with other parameters of w

³ Suppose there is only a colony of G_i forming a square in the space. If the size of the colony is 5-by-5, $\langle \bar{\rho}_i \rangle = 0.72$, while it is 0.855 for a 10-by-10 colony.

and S . Therefore, we consider that forming creole colonies at the early stage is important for full creolization.

On the contrary, at small values of α , the values of $\langle \bar{\rho}_3 \rangle$ are also small. Because the exposure ratio α determines a probability that a language learner communicates with its neighbors, forming a colony with neighbors is hardly effective in creolization at the small value of α . Rather, the creole becomes dominant due to an advantageous parameter set of w , S , and α . It is clear as evidenced by the result of the non-spatial model.

5 Conclusion

In this paper, we introduced a spatial structure to a mathematical framework of creolization. Observing this process, we discovered that forming colonies was an important factor. We showed that in the spatial language dynamics, creole could be dominant even in the high exposure ratio, different from the non-spatial model.

The quantitative analysis implies that there is a condition of creolization in terms of a combination between the ensemble average density $\langle \bar{\rho}_3 \rangle$ and the exposure ratio α . Through the experiments, we can conclude as follows:

- Creole is easy to dominate the community in a parameter set where creolization takes place in the non-spatial model, regardless of the value of α .
- The value of $\langle \bar{\rho}_3 \rangle$ is probably useful for the prediction of creolization at the early stage at the large values of α .

We need to analyze the behavior through further experiments. Although we used a toroidal 50-by-50 square grid for a spatial structure, it can be expanded to more complicated social networks. There is yet room for improvement in some settings including the initial population distribution.

Acknowledgment

This work was partly supported by Grant-in-Aid for Young Scientists (B) (KAKENHI) No.20700239 from MEXT Japan.

References

1. Cangelosi, A., Parisi, D. (eds.): *Simulating the Evolution of Language*. Springer, London (2002)
2. Briscoe, E.J. (ed.): *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, Cambridge (2002)
3. Arends, J., Muysken, P., Smith, N. (eds.): *Pidgins and Creoles*. John Benjamins Publishing Co., Amsterdam (1994)
4. Bickerton, D.: *Language and Species*. University of Chicago Press (1990)
5. DeGraff, M. (ed.): *Language Creation and Language Change*. MIT Press, Cambridge (1999)

6. Nakamura, M., Hashimoto, T., Tojo, S.: Exposure dependent creolization in language dynamics equation. In: Sakurai, A., Hasida, K., Nitta, K. (eds.) JSAI 2003. LNCS (LNAI), vol. 3609, pp. 295–304. Springer, Heidelberg (2007)
7. Nowak, M.A., Komarova, N.L., Niyogi, P.: Evolution of universal grammar. *Science* 291, 114–118 (2001)
8. Castelló, X., Eguíluz, V.M., Miguel, M.S., Loureiro-Porto, L., Toivonen, R., Saramäki, J., Kaski, K.: Modelling language competition: bilingualism and complex social networks. In: Smith, A., Smith, K., Cancho, R. (eds.) *The Evolution of Language: Proceedings of the 7th International Conference (EVOlang7)*, p. 85. World Scientific Pub. Co. Inc., Singapore (2008)
9. Abrams, D.M., Strogatz, S.H.: Modelling the dynamics of language death. *Nature* 424, 900 (2003)
10. Nakamura, M., Hashimoto, T., Tojo, S.: Creole viewed from population dynamics. In: *Proc. of the Workshop on Language Evolution and Computation in ESSLI, Vienna*, pp. 95–104 (2003)
11. Niyogi, P.: *The Informational Complexity of Learning*. Kluwer, Boston (1998)
12. Chomsky, N.: *Lectures on Government and Binding*. Foris, Dordrecht (1981)
13. Nakamura, M., Hashimoto, T., Tojo, S.: Simulation of common language acquisition by evolutionary dynamics. In: *Proc. of IJCAI 2007 Workshop on Evolutionary Models of Collaboration, Hyderabad*, pp. 21–26 (2007)
14. Miguel, M.S., Eguíluz, V.M., Toral, R., Klemm, K.: Binary and multivariate stochastic models of consensus formation. *Computing in Science and Engg.* 7, 67–73 (2005)