Emergence of Communication and Language,
C. Lyon, A. Cangelosi & C. L. Nehaniv (Eds.),
Springer, pp. 237–252 (2007)

Language Change among Memoryless Learners Simulated in Language Dynamics Equations

Makoto Nakamura[1], Takashi Hashimoto[2], and Satoshi Tojo[1]

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

Abstract. Language change is considered as a transition of a user population among languages. Language dynamics equations represent such a transition of population. Our purpose in this paper is to develop a new formalism of language dynamics based on a realistic situation of multiple language contact. We assume a situation where *memoryless learners* are exposed to a number of languages. Our experiments show that contact with other language speakers during the acquisition of a first language reduces learning accuracy and prevents the emergence of a dominant language. We suppose there is a special *communicative language* which has a higher similarity to some languages than others; when learners are frequently exposed to a variety of languages, these similar languages attract a relatively higher proportion of the population. We discuss the simulation results from the viewpoint of the language bioprogram hypothesis.

1 Introduction

In general, all human beings can learn any human language in their first language acquisition. One of the functions of language use is to communicate with others. In the work described here we investigate situations in which learners are exposed to more than one language. We make the assumption that the language learners come to acquire one of the languages that is optimal for communication, which would vary according to the environment. It is postulated that the most preferable language in the community would eventually survive and become dominant in competition with other languages, depending on how large a proportion of the people speak it. Accordingly, language change can be represented by population dynamics, examples of which include an agent-based model of language acquisition proposed by Briscoe et al. [1] and a mathematical framework by Nowak et al. [2], who elegantly presented an evolutionary dynamics of grammar acquisition in a differential equation, called the *language dynamics equation*.

One of the main factors of language change can be considered as the interaction between different language groups [3]. Introducing this factor to the language dynamics equation, we can provide a more realistic situation for language change than the existing language dynamics model. Thus, our purpose of this study is to develop a new formalism of language dynamics which deals with language contact among some number of languages, and then to investigate the relationship between the language contact and language change.

For representing the first language acquisition, two extreme learning algorithms have been proposed, called *memoryless* and *batch* learning algorithms [7]. Both memoryless and batch learners receive training examples as language input. While the batch learners guess a grammar after hearing a batch of language input, the memoryless learners do not need to store training examples for learning, changing their assumption of grammar whenever they receive an input that is inconsistent with their assumption. Komarova et al. [4] adopted those two kinds of learners into their model, comparing conditions of the two models for the emergence of a dominant language. In this paper, introducing a new transition probability for a memoryless learner exposed to a variety of languages, we compare the behavior of the dynamics with that of Komarova et al. [4].

Thus far, we have revised the model of Nowak et al. [2] in order to study the emergence of creole [5] in the context of population dynamics [6]. For the purpose of modeling the process of creolization, we claimed that infants during language acquisition had contact not only with their parents but also with other language speakers. To meet this condition, we revised the transition probability between languages to be sensitive to the distribution of languages in the population at each generation. A new control parameter, the *exposure rate*, is introduced to determine the degree of influence from other languages during acquisition. Namely, focusing on language learners, we have given a more precise environment of language acquisition than Nowak et al. [2]. In other words, introducing the exposure rate, we have regarded their model as a specific case of ours in language acquisition. Therefore, these revisions enable us to deal not only with the emergence of creole but also with other phenomena of language change. We investigate the relationship between the exposure rate and the emergence of a dominant language.

In Section 2, we propose a modified language dynamics equation and a new transition matrix for the memoryless learning algorithm. We describe our experiments in Section 3. We discuss the experimental results in Section 4. Finally, we conclude this paper in Section 5.

2 Learning Accuracy of Memoryless Learners

2.1 Outline of the Language Dynamics Equation

In this section, we explain the outline of the language dynamics equation proposed by Nowak et al. [2]. In their model, based on the principles of a universal grammar, the search space for candidate grammars is assumed to be finite, that is $\{G_1, \ldots, G_n\}^3$. The language dynamics equation is given by the following differential equations:

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} x_j f_j Q_{ji} - \phi x_i \ \ (i = 1, \ldots, n), \tag{1}$$

where

x_i : the proportion of the population that speak G_i, where $\sum_{j=1}^{n} x_j = 1$,
$Q = \{Q_{ij}\}$: the transition probability between grammars that a child of G_i speaker comes to acquire G_j,
f_i : fitness of G_i, which determines the number of children individuals reproduce, where $f_i = \sum_{j=1}^{n}(s_{ij} + s_{ji})x_j/2$,
$S = \{s_{ij}\}$: the similarity between languages, which denotes the probability that a G_i speaker utters a sentence consistent with G_j, and
ϕ : the average fitness or *grammatical coherence* of the population, where $\phi = \sum_i x_i f_i$.

The language dynamics equations are mainly composed of (i) the similarity between languages given by the matrix $S = \{s_{ij}\}$ and (ii) the probability that children fail to acquire their parental languages by the matrix $Q = \{Q_{ij}\}$.

As a similarity matrix, in this paper, we mainly deal with a special case such that:

$$s_{ii} = 1, \quad s_{ij} = a \quad (i \neq j) \ , \tag{2}$$

where $0 \leq a \leq 1$. In accordance, the transition probability comes to:

$$Q_{ii} = q, \quad Q_{ij} = \frac{1-q}{n-1} \quad (i \neq j) \ , \tag{3}$$

where q is the probability of learning the correct grammar or the *learning accuracy* of grammar acquisition. The accuracy of language acquisition depends on the search space $\{G_1, \ldots, G_n\}$, the learning algorithm, and the number of input sentences, w, during language acquisition.

2.2 Modified Language Dynamics Equation

In a situation of language contact, a child may learn a language not only from his parents but also from other language speakers who speak a different language from his parental one. In order to incorporate this possibility in a language dynamics equation, we divide the language input into two categories; one is from the parents and the other is from other language speakers. We name the ratio of the latter to the former an *exposure rate* α. This α is subdivided into smaller ratios corresponding to the distribution of all language speakers. An example distribution of languages is shown in Fig. 1. Suppose a child has parents who speak G_p, he receives input sentences from G_p on the percentage of the

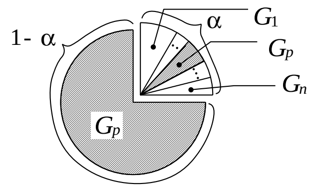


Fig. 1. The exposure rate α

shaded part, $\alpha x_p + (1-\alpha)$, and from non-parental languages G_j $(j \neq p)$ on the percentage, αx_j.

Introducing the exposure rate α, we can represent the proportion of each language to which a child is exposed during the acquisition period. Hence, assuming a total number of sentences for language acquisition, we can calculate the number of sentences the child hears for each language. We make the assumption that the language input is all in sentential form. Here, let us consider a probability of accepting with a grammar a sentence that a learner receives. If the learner presuming G_j hears a sentence only from one teacher speaking G_i, an element s_{ij} in the S matrix predefines the probability of accepting a sentence derived from G_i with G_j. In another case that the learner whose parents speak G_p is exposed to a number of languages, the learner presuming G_j accepts a sentence with such a probability, U_{pj}, that:

$$U_{pj} = \alpha \sum_{k=1}^{n} s_{kj} x_k + (1-\alpha) s_{pj} \quad . \tag{4}$$

For the special case where Eqn (2) is assumed, it is transformed to:

$$U_{pj} = \begin{cases} 1 - \alpha(1-a)(1-x_j) & (p = j) \\ a + \alpha(1-a)x_j & (p \neq j) \end{cases} \quad . \tag{5}$$

When a learning algorithm is expanded to allow language learners to be exposed to a number of languages, the matrix $U = \{U_{ij}\}$ corresponds to $S = \{s_{ij}\}$ in terms of a probability of accepting a sentence with a learner's grammar. Then, the Q matrix depends on the U matrix and the U matrix on the distribution of languages in the population, $X = \{x_i\}$. Since the distribution of population changes in time, the Q matrix comes to include a time parameter t, that is, Q is redefined as $\overline{Q}(t) = \{\overline{Q}_{ij}(t)\}$. Thus, the new language dynamics equation is expressed by:

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^{n} x_j(t) f_j(t) \overline{Q}_{ji}(t) - \phi(t) x_i(t) \qquad (i = 1, \ldots, n). \tag{6}$$

We call it the *modified language dynamics equation.*

[3] In this paper we assume that a grammar is equivalent to a language.

2.3 Memoryless Learning Algorithm

Niyogi [7] presented two extreme learning algorithms called the batch learning algorithm and the memoryless learning algorithm, in which the former is considered as the most sophisticated algorithm within a range of reasonable possibilities, and the latter as the simplest mechanism. Because the memoryless learning algorithm is easy to remodel with our proposal, we will use it and compare the behavior of the dynamics with that of Komarova et al. [4]. In this section, we explain the learning accuracy of the memoryless learning algorithm, which is derived from a Markov process.

The memoryless learning algorithm describes the interaction between a child learner and language speakers, who are assumed to speak one language each. Namely, the learner hears a set of sentences in a particular language during the acquisition period. The learner starts presuming a grammar by randomly choosing one of the n grammars as an initial state. When the learner hears a sentence from the teacher, he tries to apply his temporary grammar to accept it. If the sentence is consistent with the learner's grammar, no action is taken; otherwise the learner changes his hypothesis about the grammar to the next one randomly picked up from the other grammars. This series of learning is repeated until the learner receives w sentences.

If we consider only one teacher (the learner's parent), the learner hears only one language. In this case, the algorithm is presented by the following expressions. Let us consider a probability distribution of grammar acquisition, denoted by $\boldsymbol{p}^{(w)} = (p_1, \ldots, p_n)^T$[4], where p_i represents a probability that the learner acquires the i-th grammar after hearing w sentences. The initial probability distribution of the learner is uniform:

$$\boldsymbol{p}^{(0)} = (1/n, \ldots, 1/n)^T \ , \tag{7}$$

i.e., each of the grammars has the same chance to be picked at the initial state. If the teacher's grammar is G_k and the child hears a sentence from the teacher, the transition process from G_i to G_j in the child's mind is expressed by a Markov process with such a transition matrix $M(k)$ that:

$$M(k)_{ij} = \begin{cases} s_{ki} & (i = j) \\ \dfrac{1 - s_{ki}}{n-1} & (i \neq j) \end{cases} \ . \tag{8}$$

After receiving w sentences, the child will acquire a grammar with a probability distribution $\boldsymbol{p}^{(w)}$. Therefore, the probability that a child of a G_i speaker acquires G_j after w sentences is expressed by:

$$Q_{ij} = [(\boldsymbol{p}^{(0)})^T M(i)^w]_j \ . \tag{9}$$

The transition probability of the memoryless learning algorithm depends on the S matrix. For instance, if the condition of Eqn (2) is satisfied, the off-diagonal elements of the Q matrix are also equal to each other, and Eqn (3)

[4] A^T denotes the transposed matrix of A.

holds. Therefore, $q = Q_{ii}$ $(i = 1, \ldots, n)$ is derived as follows:

$$q = 1 - \left(1 - \frac{1-a}{n-1}\right)^w \frac{n-1}{n} \quad . \tag{10}$$

This is the learning accuracy of memoryless learners, the probability of learning the correct grammar.

Once a memoryless learner achieves his parental grammar, he will never change his hypothesis. Suppose there exist only two grammars, then the memoryless learner has two states in a Markov process, that is, a state for the hypothesis of his parental grammar, G_{parent}, and a state for the other grammar, G_{other}. The transition probability between the states is expressed by a Markov matrix $M = \{m_{ij}\}$ such that (See Fig. 2(a) as the corresponding state transition diagram):

$$M = \begin{pmatrix} 1 & 0 \\ 1-a & a \end{pmatrix} \quad , \tag{11}$$

where

m_{11}: the probability that a child who correctly guesses his parental grammar maintains the same grammar,
m_{12}: the probability that a child who correctly guesses his parental grammar changes his presumed grammar to another,
m_{21}: the probability that a child whose grammar is different from his parents' comes to presume his parental grammar, and
m_{22}: the probability that a child whose grammar is different from his parents' keeps the same grammar by accepting a sentence[5].

Komarova et al. [4] have analyzed the language dynamics equation Eqn (1), and deduced the following results: (i) When the learning accuracy is high enough, most of the people use the same language, that is, there exists a dominant language. Otherwise, all languages appear at roughly similar frequencies. (ii) The learning accuracy is calculated from a learning algorithm. Receiving input sentences, a memoryless learner enhances his learning accuracy.

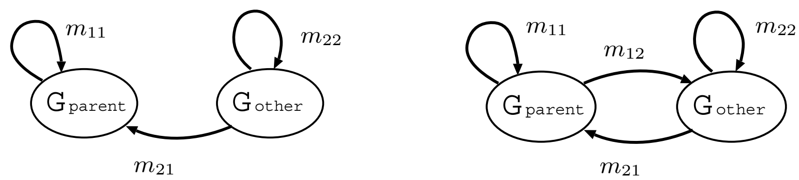
2.4 Memoryless Learners Exposed to a Number of Languages

We define a transition matrix, $\overline{Q}(t) = \{\overline{Q}_{ij}(t)\}$, of memoryless learners exposed to a number of languages during the acquisition period. For a child whose parents speak G_p, the transition matrix of a Markov process is defined by:

$$M(p)_{ij} = \begin{cases} U_{pi} & (i = j) \\ \dfrac{1-U_{pi}}{n-1} & (i \neq j) \end{cases} \quad . \tag{12}$$

[5] If the memoryless learner is able to choose the refused grammar again with a uniform probability when he failed to accept the sentence, the Markov matrix is replaced by:

$$M = \begin{pmatrix} 1 & 0 \\ (1-a)/2 & a + (1-a)/2 \end{pmatrix} \quad .$$



(a) The case in which a child hears sentences only from his parents

(b) The case in which a child hears sentences in a number of languages

Fig. 2. Markov processes for the memoryless learning algorithm

The learning accuracy is derived by substituting Eqn (12) for Eqn (9) instead of Eqn (8). Because U_{ij} varies according to the distribution of population of grammars, even in the special case where Eqn (2) is satisfied the learning accuracy of each grammar is different from each other[6]. In other words, there are n values of the learning accuracy for each grammar. The Markov matrix in Eqn (12) becomes equivalent to Eqn (8) at $\alpha = 0$. Thus, the transition probability with the exposure rate α is regarded as a natural extension of that of Komarova et al. [4].

For a learner exposed to a variety of languages, the most important difference from a non-exposed learner is that even when the learner presumes his parental grammar G_p, a received sentence may not be accepted by the grammar with the probability $1 - U_{pp}$. In this case he chooses one of the non-parental grammars randomly with a uniform probability. In a two-grammars case, for example, the Markov matrix of this process is expressed by the following equation:

$$M(p) = \begin{pmatrix} U_{p1} & 1 - U_{p1} \\ 1 - U_{p2} & U_{p2} \end{pmatrix} \quad . \tag{13}$$

We show in Fig. 2(b) the corresponding state transition diagram of a memoryless learner exposed to a number of languages, which differs from Fig. 2(a) in that for learners at a state G_p it is possible to move to another state.

In the next section, we examine how a memoryless learner is influenced by a variety of languages, and how a dominant language appears dependent on the initial conditions. Especially, we will look into the relationship between the exposure rate and the occurrence of a dominant language.

[6] For example, suppose there are two grammars, G_1 and G_2, and the number of input sentences is $w = 1$. Then, the learning accuracy of G_1 is $q_{11} = 1 - a/2 - \alpha(1-a)(1 - x_1 + x_2)/2$, while $q_{22} = 1 - a/2 - \alpha(1-a)(1 + x_1 - x_2)/2$ for G_2. When $\alpha = 0$, $q_{11} = q_{22}$.

3 Experiments

In this section, we show that the behavior of our model with the memoryless learning algorithm depends on the exposure rate α. We set the number of grammars, $n = 10$, throughout the experiments. Firstly, comparing the dynamics of the model with that of Komarova et al. [4], we examine how the exposure rate α works in our model. Secondly, we observe the behavior of the dynamics, when we suppose there is a communicative language which has a higher similarity to some languages than others have. We take the term *communicative language* to mean a special language, the speakers of which can communicate with other language speakers more easily than speakers of those languages which are not termed *communicative*. This is reflected in the similarity between the special language and other languages.

3.1 Exposure and Learning Accuracy

In this section, we compare the behavior of our model with analytical solutions of Komarova et al. [4], and with the behavior of their model by memoryless learners, which is equivalent to that of our model at $\alpha = 0$. We set the similarity between two languages, $a = 0.1$ in Eqn (2), and the number of input sentences w within the range from 10 to 50.

Komarova et al. [4] have analytically solved Eqn (1) for which Eqn (2) and Eqn (3) are substituted. The solutions of the model are derived by setting an arbitrary initial condition of the distribution of population, affected by the learning accuracy. We show in Fig. 3 the proportion of the population that speak the most prevalent grammar in the community, $\hat{x}$, versus the learning accuracy, q, by which children correctly acquire the grammar of their parents.

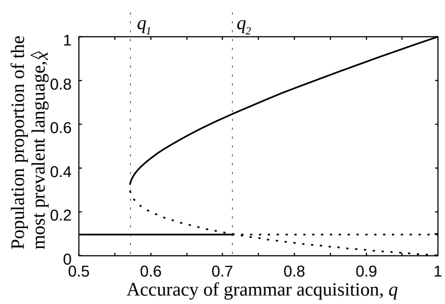
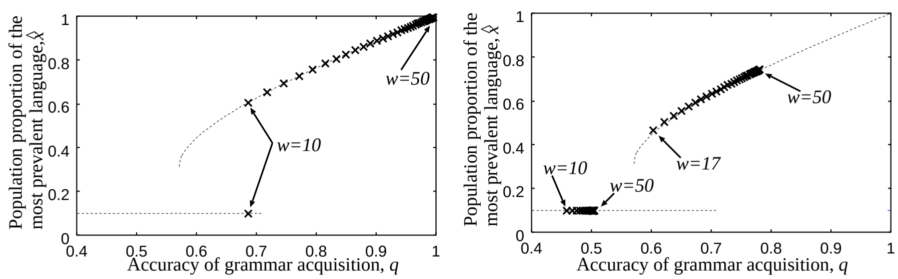


Fig. 3. Analytical solutions of Eqn (1) with Eqn (2) and Eqn (3) ($n = 10, a = 0.1$)

There are two types of solutions; one is that only one of the grammars attracts a certain proportion of the population whereas the others are given the rest divided equally. Which of the languages would be dominant depends on the initial condition. The other is that the solutions take the uniform distribution among grammars. Therefore, there are two thresholds, q_1 and q_2, in terms of the learning accuracy. When $q < q_1$, the population of each language would be uniform. When $q > q_2$, there would be one prevalent language in the community. Thus, q_1 is the necessary condition for the existence of a prevalent language and q_2 is the sufficient condition. When $q_1 < q < q_2$, the supremacy of one language depends on the initial distribution of the population.

Here, we examined our model with memoryless learners at $\alpha = 0$, which is equivalent to that of Komarova et al. [4]. Because the learning accuracy, q, depends on the number of input sentences, w, the $q - \hat{x}$ relation is discretely represented by integer numbers of w. At $\alpha = 0$, the relation must be identified with the analytical solutions, depicted in Fig. 3. The result is shown in Fig. 4(a), in which a cross ($\times$) denotes the $q - \hat{x}$ relation for a given w, and dotted lines are that of analytical solutions (copied from Fig. 3). As the result, we observed that the $q - \hat{x}$ relation of the model with memoryless learners exactly corresponds to that of the analytical solutions.



(a) $\alpha = 0$ (b) $\alpha = 0.12$

Fig. 4. Solutions by memoryless learning ($a = 0.1, w = 10, \ldots, 50$)

Next, we experimented with different values of α in the memoryless learning by w. In our model, although the transition probability $\overline{Q}_{ij}(t)$ varies depending on the distribution of the population by language at each generation, the value of $\overline{Q}_{ij}(t)$ becomes stable as the distribution of the population approaches to the solution, and vice versa. Therefore, we can observe the $q - \hat{x}$ relation as well. We expected that because of the variable transition matrix $\overline{Q}(t)$, the $q - \hat{x}$ relation underwent a change from that of the base model along with the increase of α. However, as is shown in Fig. 4(b) where $\alpha = 0.12$, the relation becomes the same as the one in Fig. 3. Instead, we can easily observe that the increase of α produces a deterioration in q in regard to w. Additionally, the solutions of q seem to be separated into two groups. We drew the graph with several patterns of the initial distribution of population. As a result, some values of α seem to derive a bifurcation of q values which depend on the initial population distribution.

In order to observe the influence of α on q, we show $\alpha - q$ relation in Fig. 5, where two lines are represented for each of $w = 10$ and 50. The number of q values is determined according to α. At $w = 50$, when α is between the dashed lines in the figure, there exist two solutions of q which depend on the initial distribution of population. Accordingly, two solutions of $\hat{x}$ are derived at $\alpha = 0.12$ and $w = 50$, as shown in Fig. 4(b). Although the $\alpha - q$ relation varies along with w, the learning accuracy, q, monotonously decreases depending on α, in common with any w. Therefore, the increase of α produces a deterioration of q in regard to a common value of w.

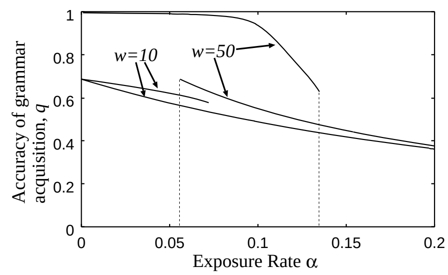


Fig. 5. Exposure rate α versus learning accuracy q ($w = 10, 50$)

In our model, q varies from generation to generation, while Komarova et al. [4] gave a constant value to q fixed by a learning algorithm. We showed that q would be stable for given α and thus x also would be stable. Apparently $q - x$ relation is similar to that of the analytical solutions, regardless of the exposure rate. At this stage, we may well conclude that the increase of α would just decrease the accuracy of learning, and would not affect $q - x$ relation, when the algorithm is memoryless and the language similarity is uniform.

3.2 Communicative Language

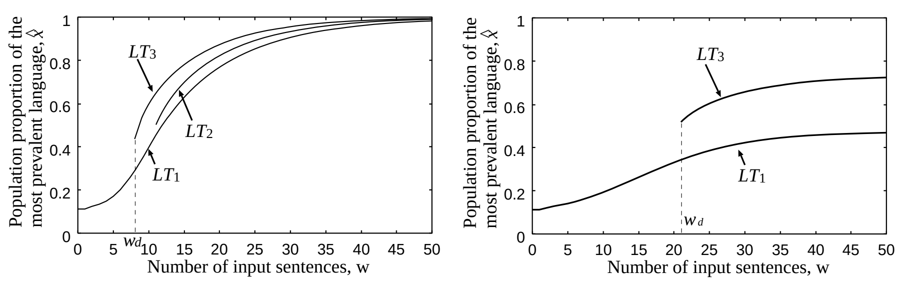
In the previous section, assuming a set of languages with a uniform similarity matrix, we succeeded in observing the characteristic behaviors of our model. Toward the investigation of the model with the general case of a similarity matrix, that is nonuniform, we consider to introduce a special communicative language, the speakers of which are easier to communicate with people speaking other languages than the others.

In terms of similarity, the special language, say G_1, has a higher similarity with a subset of languages, say G_2 and G_3, than the rest. Namely, the S matrix is expressed by:

$$S = \begin{pmatrix} 1 & b & b & & \\ b & 1 & a & & a \\ b & a & 1 & & \\ & a & & \ddots & \\ & & & & 1 \end{pmatrix} , \tag{14}$$

where $0 \leq a < b \leq 1$. We set $a = 0.1$ and $b = 0.5$ in the following experiments. Accordingly, languages are classified into three categories in terms of similarity. For simplicity, we call them LT_1, LT_2 and LT_3, which respectively contain the communicative language (G_1), the similar languages to G_1 (G_2 and G_3) and the others ($G_4 \ldots G_{10}$).

In order to observe how the exposure of children to a number of languages affects the most prevalent language, we draw diagrams of the proportion of the population that speak the most prevalent language, $\hat{x}$, versus the number of input sentences, w, at particular points of α (see Fig. 6). Although which language obtains the highest population depends on the initial distribution of the population, the proportion of the population speaking the most prevalent language is determined by its language type. For example, when the number of input sentences is $w_d = 8$ in Fig. 6(a), only G_1 or one of languages belonging to LT_3 can be the most prevalent language, while none of LT_2 can be predominant. When G_1 obtains the corresponding population speaking the most used language, that is $\hat{x}$, the rest of the languages $\{G_2, \ldots G_{10}\}$ share the rest of the population proportion, that is $1 - \hat{x}$.



(a) $\alpha = 0$ (b) $\alpha = 0.12$

Fig. 6. Number of input sentences, w, versus the proportion of the population that speak the most prevalent language, $\hat{x}$

In Fig. 6(a), we can see that the greater the number of input sentences is, the higher the population proportion of the most prevalent language exists in stable generations. Although the most prevalent language is spoken by the most of population, the proportion of the population depends on which of the language types the language belongs to. There are three kinds of $w - \hat{x}$ relation in the figure, which correspond to the type of the language (LT_i). Note that in Fig. 6(a), $LT_1 < LT_2 < LT_3$. In the language dynamics equation, the more similar two languages are to each other, the easier it is for the population to flow out to each other. In this case, G_1 has two similar languages belonging to LT_2, while each of LT_2 is similar to only one language, that is G_1, and none of LT_3 has any similar language. Thus, LT_1 is the easiest for the population to flow out. This is because the highest proportion of the population speaking the most prevalent language G_1 in LT_1 is less than that of LT_2, and LT_2 is less than LT_3.

If w is smaller than a certain number, G_1 becomes the most prevalent at any initial distribution of population. Otherwise, one of the other languages might supersede G_1 depending on the initial condition. Here, we define a threshold w_d as the smallest number of input sentences in which a language other than G_1 could become the most prevalent language. When $\alpha = 0$, the threshold w_d is 8.

We show in Fig. 6(b) a diagram of $\hat{x}$ versus w at $\alpha = 0.12$. The threshold w_d is boosted to 21, and none of LT_2 reaches the enough population to become the most prevalent language at $w < 50$. As was mentioned in Section 3.1, the increase of the exposure rate makes the learning accuracy low. For the memoryless learning algorithm, the learning accuracy, q, increases with the number of input sentences, w. The increase of w keeps the same quality of learning accuracy in response to α. Accordingly, w_d increases along with the exposure rate α.

We suggested in Fig. 6 that the larger the exposure rate α was, the greater the threshold w_d was. It is expected when language learners are exposed to a number of languages, one of the languages other than G_1 may stand out as long as the learners hear the proper quantity of language input. The minimum quantity is w_d in Fig. 6. However, human beings have an acquisition period in which an appropriate grammar is estimated from their language input [8]. If the possible number of input sentences to be heard during the acquisition period was settled at a specific value, then we could draw a diagram concerned with the influence of the exposure rate, α, on the proportion of the population who speak the most prevalent language, $\hat{x}$. We show an example of the diagram for $w = 30$ in Fig. 7.

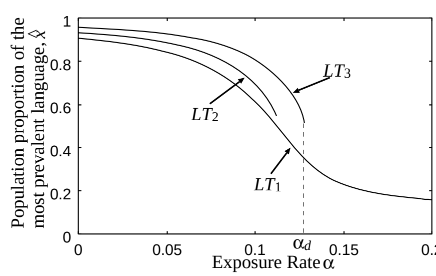


Fig. 7. Influence of the exposure rate, α, on the population proportion of the most prevalent language, $\hat{x}$ ($w = 30$)

We define α_d as the highest value of the exposure rate at which one of the languages other than G_1 could become the most prevalent depending on the initial distribution. When $w = 30$, $\alpha_d \simeq 0.128$. It is easily conceivable that the greater the number of the input sentences is, the larger the threshold α_d is.

Thus far, we have observed the smallest number of input sentences for the appearance of the most prevalent language other than G_1, that is w_d, at particular values of α. On the other hand, we saw the highest value of the exposure rate for the appearance of the most prevalent language other than G_1, that is α_d, at a particular number of input sentences. These two values have a functional relationship as shown in Fig. 8. This figure represents the relationship between w and α for the most prevalent language other than G_1. The necessary number of input sentences rapidly increases along with the exposure rate. Learners need to receive 222 sentences at $\alpha = 0.13$, though only 34 sentences at $\alpha = 0.129$.

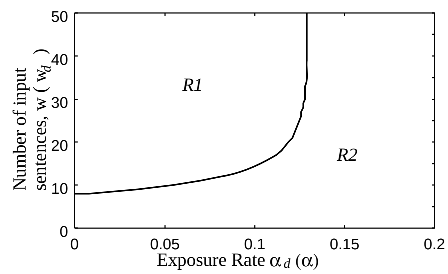


Fig. 8. The relationship between two thresholds, α_d and w_d

This series of experiments shows that the communicative language may be the most prevalent, regardless of the exposure rate α or the number of input sentences w. We discuss the communicative language in the next section.

4 Discussion

4.1 Possibility of Language Change

In this paper, we consider the change of language as the transition of language users. In other words, the change of language is a phenomenon that the proportion of the population who speak a language at the stable generation exceeds that of the most used language at the initial condition. Here, we discuss the possibility of language change, based on the experimental result shown in Fig. 8.

The line in the diagram Fig. 8 can be recognized as a boundary between the following two regions:

R1: All of the languages have a possibility of being predominant. The language change hardly occurs.

R2: Only the communicative language attracts a certain proportion of the population in any initial conditions. The language change is likely to occur.

Language learners developing under the condition of $R1$ hear enough language input to acquire their parental languages with high learning accuracy. One of the languages may predominate in the community, depending on the initial distribution of the population. In most cases, the language used by most speakers at the initial state tends to keep the predominance.

In the area of $R2$, the most populous language comes nothing but G_1, although the proportion of the population speaking G_1 at the stable generation is quite low in comparison with that of the most prevalent languages in $R1$. Even if no one spoke G_1 at the initial state, G_1 eventually comes to be the most used language. Because G_1 definitely exceeds the other languages in population, it is considered as the change of the predominant language.

4.2 Communicative Language and the Bioprogram Hypothesis

In Section 3.2, we assumed that there is a communicative language G_1, which is more similar to two particular languages than the others. Let us consider what the language corresponds to in the real world. We suggest that it is considered as a language that Bickerton [9] supposed in the *Language Bioprogram Hypothesis*. Kegl et al. [10] briefly outline the features of the hypothesis as follows:

> Bickerton [9] proposed the Language Bioprogram Hypothesis. This hypothesis claims that a child exposed to nonoptimal or insufficient language input, such as a pidgin, will fall back on an innate language capacity to flesh out the acquisition process, subsequently creating a creole. This is argued to account for the striking similarities among creoles throughout the world.
>
> Kegl et al. [10]

The communicative language has something in common with the bioprogrammed language, the innate language in the passage above, with regard to the condition of emergence. It appears when learners are exposed to other languages so frequently that any dominant language does not appear, or when they are not given sufficient language input. The communicative language would emerge as a creole, since from the viewpoint of population dynamics, a creole is a language which no one spoke at the initial state but comes to obtain a significant population after generations [11].

If we recognize that the communicative language is consistent with the language bioprogram hypothesis, does its reverse still keep true? Namely, are the bioprogrammed languages in the real world such as creoles more communicative with other languages than the others? We cannot examine in the real world whether the creoles are more similar to some particular languages or not. In order to answer the question, we further need to associate the languages given in our experiments with actual languages. Namely, if we introduced to embed some linguistic features into the equation, the *creole* which emerged in our experiments could be compared with actual ones.

4.3 Applicability of the Modified Language Dynamics Equation

Let us consider what further aspects of language could be modeled in our simulation. In both the models of Komarova et al. [4] and ourselves, it is necessary to introduce a method of representing the similarity of languages. If we take some aspects of language in a real situation into the model, we need to abstract a similarity measure from the target languages. In other words, these models could be applied to whatever the underlying similarity of the target feature is calculated, and thus the model could be extended to investigate whether the emerging *creoles* resembled each other, as predicted by the bioprogram hypothesis.

5 Conclusion

Contact between different language groups has been considered as one of the main factors in language change. We modeled the language contact by introducing the *exposure rate* to the language dynamics equation proposed by Nowak et al. [2]. The exposure rate is the rate of influence of languages other than the parental one on language acquisition. We assess the accuracy of parental language acquisition in the memoryless learning algorithm. The exposure to other languages made it possible that the language learner refuted his presumed grammar even though he once acquired his parental grammar. We revised a new transition probability that changes in accordance with the distribution of users of each language, which is a different feature from Nowak et al. [2].

As the experimental result showed, the emergence of a dominant language depends not only on the similarities between languages but also on the amount of contact between users of different languages. We compared our result with Komarova et al. [4] in Section 3.1. First, when the similarity was uniform, we found that the introduction of the exposure rate only reduced the accuracy of the target language acquisition. And then, we confirmed that no dominant language emerges when the exposure rate is sufficiently high.

In the next experiment in Section 3.2, we assumed that there is a special language called the communicative language, the speakers of which are easier to communicate with users of other languages, among the multiple language communities. The result suggests the following conclusions. If language learners hear enough language input to estimate their parental language, one of the languages other than the communicative language would be dominant. However, when language learners are frequently exposed to a variety of languages, the communicative language attracts a significant proportion of the population regardless of the number of input sentences. This characteristic behavior suggests that a bioprogrammed language as hypothesized by Bickerton [9] will develop. The experimental result shown in Fig. 8 suggests that creole will emerge when language learners are exposed to a variety of languages at a certain rate.

Overall, we observed that language change is affected by the interaction between multiple languages in a rather convincing way through our experiments. Our contribution in this study can be of practical use in investigations into the relationship between the environment of language learning and language change.

Acknowledgment

We would like to thank Dr. Caroline Lyon, who is one of the editors, for reading the draft and making a number of helpful suggestions.

References

1. Briscoe, E.J.: Grammatical acquisition and linguistic selection. In Briscoe, T., ed.: Linguistic Evolution through Language Acquisition: Formal and Computational Models. Cambridge University Press (2002)
2. Nowak, M.A., Komarova, N.L., Niyogi, P.: Evolution of universal grammar. Science **291** (2001) 114–118
3. Sebba, M.: Contact Languages: Pidgins and Creoles. Macmillan, London (1997)
4. Komarova, N.L., Niyogi, P., Nowak, M.A.: The evolutionary dynamics of grammar acquisition. Journal of Theoretical Biology **209**(1) (2001) 43–59
5. Arends, J., Muysken, P., Smith, N., eds.: Pidgins and Creoles. John Benjamins Publishing Co., Amsterdam (1994)
6. Nakamura, M., Hashimoto, T., Tojo, S.: The language dynamics equations of population-based transition – a scenario for creolization. In Arabnia, H.R., ed.: Proc. of the IC-AI'03, CSREA Press (2003) 689–695
7. Niyogi, P.: The Informational Complexity of Learning. Kluwer, Boston (1998)
8. Lenneberg, E.H.: Biological Foundations of Language. John Wiley & Sons, Inc., New York (1967)
9. Bickerton, D.: The Language Bioprogram Hypothesis. Behavioral and Brain Sciences **7**(2) (1984) 173–222
10. Kegl, J., Senghas, A., Coppola, M.: Creation through contact: Sign language emergence and sign language change in Nicaragua. In DeGraff, M., ed.: Language Creation and Language Change. The MIT Press, Cambridge, MA (1999)
11. Nakamura, M., Hashimoto, T. and Tojo, S.: Creole viewed from population dynamics. Proc. of the Workshop on Language Evolution and Computation in ESSLLI, (2003) 95–104