Language Change in Modified Language Dynamics Equation by Memoryless Learners

Makoto Nakamura[⋆] Takashi Hashimoto[†] Satoshi Tojo[⋆]

Japan Advanced Institute of Science and Technology

School of { [⋆]Information , [†]Knowledge } Science

Nomi, Ishikawa, 923-1292, Japan

{mnakamur, hash, tojo}@jaist.ac.jp

Abstract

Language change is considered as a transition of population among languages. The language dynamics equation represents such a transition of population. Our purpose in this paper is to develop a new formalism of language dynamics for a real situation of language contact. We assume a situation that memoryless learners are exposed to a number of languages. We show experimental results, in which contact with other language speakers during acquisition period deteriorates the learning accuracy and prevents the emergence of a dominant language. If we suppose a communicative language, when learners are frequently exposed to a variety of languages, the language earns relatively higher rate of population. We discuss the communicative language from the viewpoint of the language bioprogram hypothesis.

1 Introduction

In general, all human beings can learn any human language in the first language acquisition. One of the main functions of language use is to communicate with others. Therefore, it is easy to consider that the language learners come to obtain a language which they hear most in the community. In other words, the most preferable language in the community would eventually survive and become dominant in competition with other languages, depending on how much ratio of the people speak it. Accordingly, language change can be represented by a population dynamics, examples of which include an agent-based model of language acquisition proposed by Briscoe (2002) and a mathematical framework by Nowak et al. (2001), who elegantly presented an evolutionary dynamics of grammar acquisition in a differential equation, called the *language dynamics equation*.

Our purpose of this study is to develop a new formalism of language dynamics which deals with language contact between language learners and speakers, and then to investigate the relationship between the language contact and language change. Thus far, we have revised the model of Nowak et al. (2001) to be more realistic, in order to study the emergence of creole (DeGraff, 1999) in the context of population dynamics (Nakamura et al., 2003). For the purpose of modeling the process of creolization, we claimed that children during language acquisition should contact not only with their parents but also with other language speakers. To meet this condition, we revised the transition rate between languages to be sensitive to the distribution of languages in the population at each generation. We introduced the *exposure rate* 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. (2001). In other words, introducing the exposure rate, we have regarded the model of Nowak et al. (2001) 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.

In this paper, we aim at examining the behavior of our model in terms of language change. Komarova et al. (2001) adopted two kinds of language learners called *memoryless learners* and *batch learners*, 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. (2001).

In Section 2, we propose a modified language dynamics equation and a new transition matrix of 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

We explain the outline of the language dynamics equation proposed by Nowak et al. (2001). In their model, given the principles in the universal grammar, the search space for candidate grammars is assumed to be finite, that is $\{G_1, \ldots, G_n\}$. The language dynamics equation is given by the following differential equations:

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

where

x_i : the ratio of the population of G_i speakers, 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 by (i) the similarity between languages as the matrix $S = \{s_{ij}\}$ and (ii) the probability that children fail to acquire their parental languages as the matrix $Q = \{Q_{ij}\}$. 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.

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

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

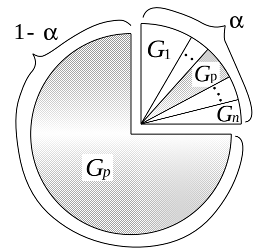


Figure 1: The exposure rate α

where a is a number between 0 and 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.

2.2 Modified Language Dynamics Equation

In a situation of language contact, a child may learn 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 to language dynamics equation, we divide the language input into two categories; one is from his parents and the other is from other language speakers. We name the ratio of the latter an *exposure rate* α. This α is subdivided into the smaller ratios corresponding to the distribution of all language speakers. An example distribution of languages is shown in Fig. 1. The child of G_p speaker is exposed to G_p at the rate of the shaded part, that is $\alpha x_p + (1 - \alpha)$, and the ratio of a non-parental language G_j comes to be αx_j.

Suppose that a child whose parents speak G_p hears sentences from the adult speakers depending on the exposure rate and on the distribution of population. If the child presumes G_j and hears a sentence, it is accepted with such a probability, U_{pj}, that:

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

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

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

When a learning algorithm is expanded into the one which allows language learners to be exposed to a number of languages, the matrix $U = \{U_{ij}\}$ corresponds to $S = \{s_{ij}\}$ in terms of an acceptable probability of a sentence for a child. Then, the Q matrix depends on the U matrix and the U matrix on the population rate. 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*.

2.3 Memoryless Learning Algorithm

Komarova et al. (2001) argue two extreme learning algorithms called the batch learning algorithm and the memoryless learning algorithm (Niyogi, 1998), 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 be remodeled with our proposal, we will use it and compare the behavior of the dynamics with that of Komarova et al. (2001). In this section, we explain the learning accuracy of the memoryless learning algorithm derived from a Markov process.

The memoryless learning algorithm describes the interaction between a child learner and language speakers. Namely, the child hears sentences of a language. 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.

Komarova et al. (2001) supposed there is one teacher (the learner's parent), so that the learner hears only one language. In this case, the algorithm is presented by the following expressions. The initial probability distribution of the learner is uniform: $\boldsymbol{p}^{(0)} = (1/n, \ldots, 1/n)^T$, where A^T is the transposed matrix of A, 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:

$$M(k)_{ij} = \begin{cases} s_{ki} & (i = j) \\ \dfrac{1 - s_{ki}}{n-1} & (i \neq j) \end{cases} . \tag{7}$$

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

$$Q_{ij} = [(\boldsymbol{p}^{(0)})^T M(i)^w]_j \ . \tag{8}$$

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

$$q = 1 - \left(1 - \frac{1-a}{n-1}\right)^w \frac{n-1}{n} \ . \tag{9}$$

This is the learning accuracy of memoryless learner.

If 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:

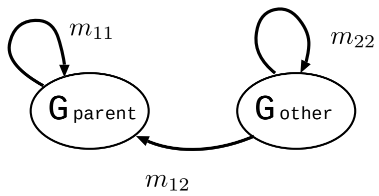
$$M = \begin{pmatrix} 1 & 0 \\ 1-a & a \end{pmatrix} , \tag{10}$$

where

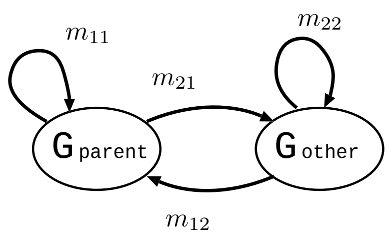
- **m_{11}:** the probability that a child who correctly guesses his parental grammar maintains the same grammar,
- **m_{12}:** the probability that a child who correctly guesses his parental grammar changes his presumed grammar to another,
- **m_{21}:** the probability that a child whose grammar is 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[1].

[1] 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:

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



(a) A case a child hears sentences only from his parents



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

Figure 2: Markov processes for the memoryless learning algorithm

Figure 2(a) shows a state transition diagram.

Komarova et al. (2001) have analyzed the language dynamics equation and deduced the following results:

- 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.

- 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 acquisition period. For a child whose parents speak G_p, the transition matrix of a Markov process is defined by:

$$M(p)_{ij} = \begin{cases} U_{pi} & (i = j) \\ \dfrac{1 - U_{pi}}{n-1} & (i \neq j) \end{cases} . \qquad (11)$$

The learning accuracy is derived by substituting Eqn (11) for Eqn (8). Because U_{ij} varies according to the distribution of population of each grammar, even in the special case where Eqn (2) is satisfied the learning accuracy of each grammar is different from each other. In other words, there are n values of the learning accuracy for each grammar. Expression (11) becomes equivalent to Eqn (7) at $\alpha = 0$. Thus, the transition probability with the exposure rate α is regarded as a natural extension of that of Komarova et al. (2001).

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. Thus, the memoryless learner is likely to refute his hypothesis even if once he acquired his parental grammar. In a two-grammars case, for example, the Markov matrix of this process is expressed by the following equation:

$$M(p) = \begin{pmatrix} U_{p1} & 1 - U_{p1} \\ 1 - U_{p2} & U_{p2} \end{pmatrix} . \qquad (12)$$

Figure 2(b) shows a state transition diagram of a memoryless learner exposed to a number of languages, which differs from Fig. 2(a) in that learners at a state G_p are possibly to move to another state.

In this section, we revised the memoryless learning algorithm in order to model a more real situation of language contact. 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.

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, $n = 10$, through the experiments. Firstly, comparing the dynamics of the model with that of Komarova et al. (2001), we examine how the exposure rate α works in our model. Secondly, we observe the behavior of dynamics, when there is a particular language in terms of the similarity.

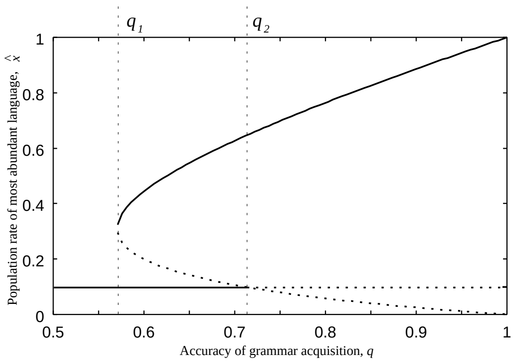


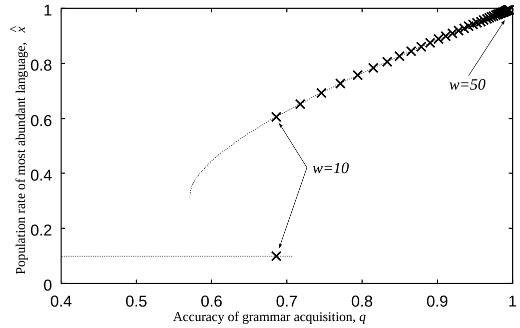
Figure 3: Analytical solutions of Eqn (1) which satisfies Eqn (2) and Eqn (3) ($n = 10, a = 0.1$)

3.1 Exposure and Learning Accuracy

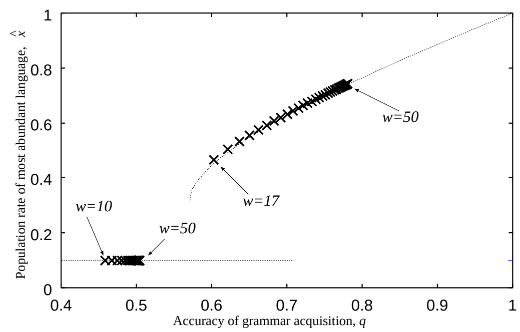
In this section, we observe the behavior of our model especially when Eqn (2) is satisfied. We compare the behavior of our model with analytical solutions of Komarova et al. (2001), and with the behavior of their model by memoryless learners, which is equivalent to that of our model at $\alpha = 0$.

Expression (1) substituted for Eqn (2) and Eqn (3) has analytically been solved by Komarova et al. (2001). The solutions of the model are derived by setting an arbitrary initial condition of the distribution of population, affected by the learning accuracy. Figure 3 shows the population rate of the most prevalent grammar in the community, $\hat{x}$, versus the learning accuracy, q, by which children correctly acquire the grammar of their parents, in case of $a = 0.1$. There are two types of solutions; one is that only one of the grammars earns a certain rate of population whereas the others are given the rest divided equally. Which of 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. 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 the 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 population.

Here, we examined our model with memoryless learners at $\alpha = 0$, which is equivalent to that of Komarova et al. (2001). Because the learning accuracy, q, depends on the number of input sentence, w, the $q - \hat{x}$ relation is discretely represented by integer numbers of w. At $\alpha = 0$, the relation must identify that of the analytical solutions, depicted in Fig. 3.



(a) Solutions by memoryless learning ($\alpha = 0$)



(b) Solutions by memoryless learning ($\alpha = 0.12$)

Figure 4: The behavior of the model depending on the exposure rate α ($a = 0.1, w = 10, \ldots, 50$)

The result is shown in Fig. 4(a), in which the number of sentences, w, was given within the range from 10 to 50. In the figure, 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.

Next, we experimented 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 population rate at each generation, the value of $\overline{Q}_{ij}(t)$ becomes stable as the population rate 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 collapsed 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 deterio-

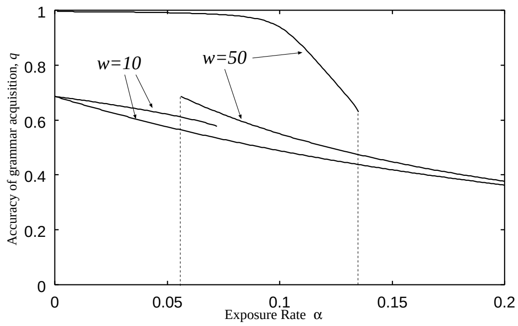


Figure 5: Exposure rate α versus learning accuracy q ($w = 10, 50$)

rates q in regard to w. Additionally, the solutions of q seem to be separated into two groups. We drew the graph with a 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 deteriorates q in regard to a common value of w.

In our model, q varies from generation to generation, while Komarova et al. (2001) 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 Komarova et al. (2001). 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 this section, we assume such a hypothetical language G_1, given G_2 and G_3, that is much similar to G_2 and G_3 than the rest. The S matrix is expressed by:

$$S = \begin{pmatrix} 1 & b & b & & \\ b & 1 & a & & a \\ b & a & 1 & & \\ & a & & \ddots & \\ & & & & 1 \end{pmatrix}, \qquad (13)$$

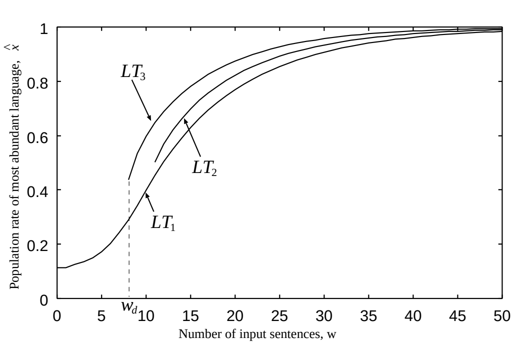
where $0 \leq a < b \leq 1$. We set $a = 0.1$ and $b = 0.5$ for the following experiments. Accordingly, languages are classified into three categories in terms of the similarity. For simplicity, we call them LT_1, LT_2 and LT_3, each of which includes the communicative language (G_1), the similar languages 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 abundant language, we draw diagrams of the population rate of most prevalent language, $\hat{x}$, versus the number of input sentences, w, at particular points of α (see Fig. 6).

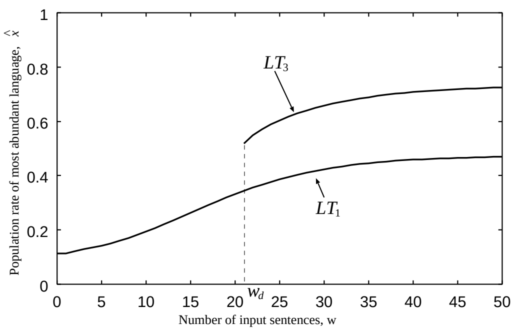
We start from $\alpha = 0$. Figure 6(a) shows that the greater the number of input sentences is, the higher the population rate of the most prevalent language exists in stable generations. The population rate of the most prevalent language depends on which of language types the language belongs to. Therefore, we can see 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), $LT_1 < LT_2 < LT_3$. Komarova et al. (2001) explained the reason as follows; G_1 has a larger intersection with the rest of the languages than the rest of them. When this language becomes preferred, it stands out less than other languages would in its place, i.e., it corresponds to lower values of the population rate. When a language earns the most abundant population rate, the other languages share the rest, so that except for the most abundant language the rate of a language equals to another one which belongs to the same language type.

If w is smaller than a certain number, G_1 becomes the most abundant at any initial distribution of population. Otherwise, one of 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.

Figure 6(b) shows a diagram of $\hat{x}$ versus w at $\alpha = 0.12$. The threshold w_d is boosted to 21, and any of LT_2 does not earn the most abundant rate of population 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



(a) Number of input sentences, w, versus population rate of most abundant language, $\hat{x}$ ($\alpha = 0$)



(b) Number of input sentences, w, versus population rate of most abundant language, $\hat{x}$ ($\alpha = 0.12$)

Figure 6: The behavior of the model with a communicative language

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 showed in Fig. 6 that the larger the exposure rate α was, the greater the threshold w_d was. It is expected that no matter how language learners are exposed to a number of languages, one of languages other than G_1 may stand out as long as the learners hear the proper quantity of language input. The 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 and it is limited in a finite time (Lenneberg, 1967). If the possible number of input sentences to be heard during acquisition period was settled in a specific value, then we could draw a diagram concerned with the influence of the exposure rate, α, on the population rate of most abundant language, $\hat{x}$. Figure 7 is an example of the diagram for $w = 30$.

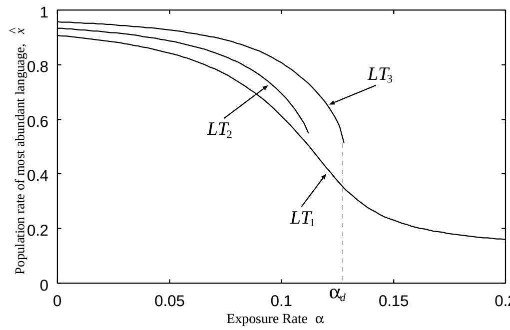


Figure 7: Influence of the exposure rate, α, on the population rate of most abundant language, $\hat{x}$ ($w = 30$)

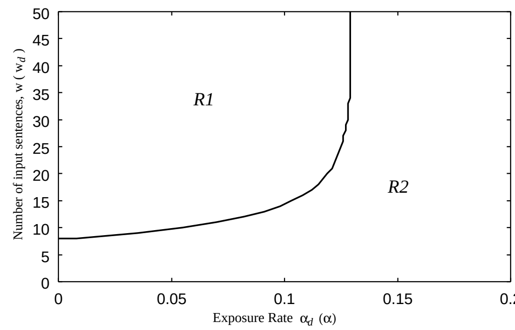


Figure 8: The relationship between two thresholds, α_d and w_d

We define α_d as the highest value of exposure rate at which one of languages other than G_1 could become the most abundant depending on the initial distribution. In case of $w = 30$, it was $\alpha_d \simeq 0.128$. 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 abundant 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 abundant language other than G_1, that is α_d, at a particular number of the input sentence. These two values have a functional relationship as shown in Fig. 8. This figure represents conditions of w and α on the appearance of the most abundant 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 $\alpha = 0.13$, while 34 sentences at $\alpha = 0.129$. Although the $\alpha - w$ relation depends on the S matrix, the figure of the curve is expected to be basically kept at arbitrary distribution of elements in the S matrix.

4 Discussion

4.1 Possibility of Dominant Language

Figure 8 can be recognized as a boundary between the following two regions:

R1: one of languages other than the communicative one may become predominant.

R2: the communicative language obtains a certain rate of population for any initial conditions.

Language learners growing under the condition of $R1$ hear enough language input to acquire their parental languages with high learning accuracy. Although one of the languages may predominate in the community, which of languages becomes predominant depends on the initial distribution of population. Some of them are regarded as a dominant language. In most cases, the most populous language at the initial state tends to take the supremacy.

In the area of $R2$, the most populous language comes nothing but G_1, although it is hard to be regarded as a dominant language because of smaller population rate. Even if no one spoke G_1 at the initial state, G_1 eventually comes to be the most abundant language. Therefore, the change of the predominant language is easy to occur.

It seems that the condition of $R1$ is hardly satisfied for w when α is larger than approximately 0.13. This result suggests that any dominant language never appear as long as language learners are frequently exposed to a variety of languages.

4.2 Communicative Language and Bioprogram Hypothesis

In Section 3.2, we assumed that there is a communicative language, which is more similar to particular two languages than the others, that is G_1. Let us consider what the language corresponds to in the real world. We dare say that it is considered as a language that Bickerton (1984) supposed in the *Language Bioprogram Hypothesis*. Kegl et al. (1999) briefly outline the features of the hypothesis as follows:

> Bickerton (1984) 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. (1999)

The communicative language has something in common with the bioprogrammed language in terms of the condition of existence; it appears when learners are frequently exposed to other languages so that any dominant language does not appear, or when they are not given sufficient language input. Therefore, if no one spoke the communicative language at the initial state, it would emerge as a creole.

If we recognize the communicative language to be consistent with the language bioprogram hypothesis, the bioprogrammed language is more communicative with pre-existing languages than the others. However, we cannot examine whether the creole is more similar to some particular languages or not. To ensure our hypothesis here, we need to embed linguistic features into the equation.

5 Conclusion

Contact of different language groups has been considered as one of main factors in language change. We modeled the contact by introducing the *exposure rate* to the language dynamics equation proposed by Nowak et al. (2001). 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 doubted his hypothetical grammar even though he once acquired his parental grammar. We expressed the acquisition process in a Markov matrix, and then revised a new transition probability that changes in accordance with the distribution of population, which is a different feature from Nowak et al. (2001). In addition, each grammar has a different learning accuracy even in the completely symmetrical similarity matrix of Eqn (2).

As the experimental result showed, the emergence of a dominant language depends not only on the similarities between languages but also on the ratio of contact of multiple languages.

We compared our result with Komarova et al. (2001) in Section 3.1. First, in case the similarity was uniform, we found that the introduction of the exposure rate α only deteriorated the accuracy of the target language acquisition; even though the population ratio versus the learning accuracy was the same, the introduction of α delayed the learning process. For memoryless learners, the failure of communication after achieving their parental grammars is fatal to the acquisition of a correct grammar. Therefore, when children are only exposed to other languages a little, a dominant language disappears. On the contrary, we expect that batch learners are robuster in terms of the noise. Our next target is to show the similar phenomena in the batch learning algorithm.

In the next experiment, we assumed that there is a most communicative language among the multiple language communities. The result suggests the following matters; If language learners hear enough language input to estimate their parental languages, one of languages other than the communicative language would be dominant. However, when language learners are frequently exposed to a variety of languages, the communicative language earns a certain rate of population regardless of the number of input sentences. The characteristic behaviors suggest that a bioprogrammed language hypothesized by Bickerton (1984). 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.

References

D. Bickerton. The language bioprogram hypothesis. *Behavioral and Brain Sciences*, 7(2):173–222, 1984.

E. J. Briscoe. Grammatical acquisition and linguistic selection. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, chapter 9. Cambridge University Press, 2002.

M. DeGraff, editor. *Language Creation and Language Change*. The MIT Press, Cambridge, MA, 1999.

J. Kegl, A. Senghas, and M. Coppola. Creation through contact: Sign language emergence and sign language change in nicaragua. In M. DeGraff, editor, *Language Creation and Language Change*, chapter 1. The MIT Press, Cambridge, MA, 1999.

N. L. Komarova, P. Niyogi, and M. A. Nowak. The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, 209(1):43–59, 2001.

E. H. Lenneberg. *Biological Foundations of Language*. John Wiley & Sons, Inc., New York, 1967.

M. Nakamura, T. Hashimoto, and S. Tojo. The language dynamics equations of population-based transition – a scenario for creolization. In H. R. Arabnia, editor, *Proc. of the IC-AI'03*, pages 689–695. CSREA Press, 2003.

P. Niyogi. *The Informational Complexity of Learning*. Kluwer, Boston, 1998.

M. A. Nowak, N. L. Komarova, and P. Niyogi. Evolution of universal grammar. *Science*, 291:114–118, 2001.