The Emergence of Artificial Creole by the EM Algorithm

Makoto Nakamura and Satoshi Tojo

Graduate School of Information Sciences,
Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Tatsunokuchi-machi, Nomi-gun, Ishikawa, 923-1211, Japan
{mnakamur, tojo}@jaist.ac.jp

Abstract. Studies on evolutionary dynamics of grammar acquisition on the computer have been widely reported in recent years, where an agent learns grammars of other agents through the exchange of sentences between them. Particularly, Nowak et al. [11] generalized an evolutionary theory of language with the universal grammar mathematically. In this paper, we propose a model of language evolution for the emergence of creole based on their theory, and try to discover the criteria conditions for creolization. In our experimentation, we utilize the inside-outside (EM) algorithm to find the grammar of a new generation. As a result, we contend that creolization is strongly affected by the popularity of community of the original language, rather than the similarity of original grammars.

1 Introduction

Studies on evolutionary dynamics of grammar acquisition on the computer have been often reported in recent years, where an autonomous and active agent learns grammars of other agents through the exchange of sentences between them. However, those experimental results hardly seem to reflect the phenomena of language evolution in the real world, because the models and the systems in those reports were too abstract. In this paper, we especially pay attention to the model of language evolution that retains intrinsic linguistic features, and try to discover the criteria conditions for the emergence of creole.

For us to realize language dynamics on the computer, a rational agent should behave as follows:

- an agent composes messages to other agents with her own grammar,
- listens to and recognizes messages from other agents with her own grammar,
- evaluates and learns grammars of the messages of other agents, and
- leaves offsprings proportional to the payoff, that is the recognition rate of other agents' grammars.

Thus far, Hashimoto et al. [6, 7] modeled an evolution of symbolic grammar systems by agents who use a very simple grammar and a simple learning mechanism. However, these attractive reports seem to be no more stories of cognitive agents than the primitive protocol matching of lower animals is, and were not about a behaviour of human beings but about just artificial organisms. In order to simulate human language dynamics, we need to make the language system more sophisticated. Because we cannot implement the mechanism of human learning process, nor that of diachronical evolution of languages directly on the computer, the adequateness of a model is difficult to evaluate.

Some works hypothesize UG (universal grammar) [4], that is an innate grammar available to human babies when they begin to learn a language [8]. It is believed that UG is the product of some special neural circuitry within the human brain, which is called 'language organ' by Chomsky, and 'language instinct' by Pinker [13]. The advantage of UG is to restrict the search space of possible candidate grammars. Briscoe [3] reported models of human language acquisition on UG, where each agent had a hierarchical lattice of categories, and a given set of parameters specified a category grammar. Instantiating UG by agents in their model, they could express the evolutionary dynamics of natural language which consisted of eight basic language families in terms of the unmarked, canonical order of verbs (V), subject(S) and objects(O).

Nowak et al. take a middle position between those two extreme models: Hashimoto's simplified model and Briscoe's sophisticated model[11]. They generalized an evolutionary theory of language with UG mathematically. It was assumed that the search space provided by principles in UG was finite and all the possible grammars could be enumerated. From the assumption, defining the similarity matrix and the payoff between grammars, they represented the transition of population of grammars as a differential equation. Consequently, they succeeded in representing an equilibrium of language evolution.

Based on this Nowak's framework, we will discuss the conditions that allow creolization to occur. As for the learning mechanisms of each agent, Nowak et al. assumed a stochastic framework. Similarly, we adopt a statistical learning algorithm called the Expectation Maximization (EM) algorithm.

Section 2 describes how creole emergence should be represented mathematically in the model. Section 3 describes our experiments in which an agent utilizes the inside-outside(EM) algorithm for a grammar estimation. In Section 4, we discuss our contribution.

2 The Model of Creole Emergence

In this section, we describe how creolization emerges based on *population dynamics* for which Nowak et al. [11] proposed a framework. Their purpose is to develop a mathematical theory for the evolutionary and population dynamics of grammar acquisition [8]. Particularly, they do not pay attention to an ability of each agent but the whole behaviour of the population, and in this point, this work is different from other works in which each agent obtains a target grammar by the mutual interaction [3, 6].

They employed the similarity matrix and the payoff between grammars to represent the differential equation for the population dynamics as mentioned

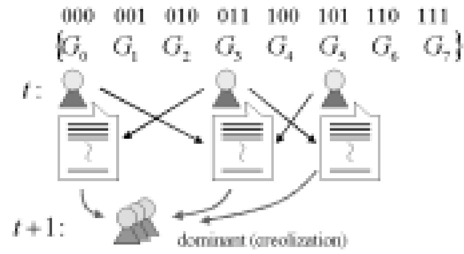

Fig. 1. Population dynamics of creolization.

above. The similarity matrix consists of elements of the probability s_{ij} that an agent who uses a grammar G_i may utter a sentence that is compatible with G_j. The Q matrix is defined as the member of the equation of the payoff and the differential equation of the population dynamics, where q_{ij} is the probability that a child born to an agent using G_i will develop G_j. Thus, the probability that a child will develop G_i if the parent uses G_i is given by q_{ii}. The q_{ii} represents the *accuracy* of grammar acquisition.[1] Roughly speaking, Nowak's results were: (1) For low accuracy of grammar acquisition (low values of q_{ii}), all grammars, G_i, occur with roughly equal abundance. There is no predominating grammar in the population. (2) For high accuracy of grammar acquisition, the population will converge to a stable equilibrium with one dominant grammar. In the latter, the dominant grammar indicates an existing grammar, namely $x_d(0) > 0$, where x_d is the relative abundance of agents who use the dominant grammar G_d at the last. According to Bickerton, creole, namely a new language, emerges under peculiar environments [2]. In most cases when multiple language contact, one of them would eventually dominate the others. Nowak's work has not considered the emergence of a new language, and thus there was no concept of creole.

If all languages are enumerated, creole must be also included in them. Let us consider a creole grammar, G_c. At first, any agent does not have the grammar. That is $x_c(0) = 0$. Creolization means that the grammar emerges and eventually dominates in the population. We propose a model of creole emergence in terms of population dynamics. This model can be considered as a natural extension of Nowak's work. Creole is defined in the model as a grammar G_c such that:

$$x_c(t) = 0, x_c(t+1) \geq \theta_c \tag{1}$$

where θ_c is the threshold to admit the grammar to be dominant (See Fig. 1.)

According to Briscoe's [3], the principles restrict all conceivable grammars to relevant ones and they are anchored to a natural language grammar by the parameters for word order. The parameters are three independent binary variables, hence there can be eight grammars in the search space. All grammar rules in the principle are represented in Chomsky's normal form (CNF). Each parameter is assigned to a specific rule of the grammar to change the order of the non-terminal symbols in the right hand side of the rule. For example, a rule "$S \rightarrow VP\ NP$" in the grammar has the parameter 0, then the rule "$S \rightarrow NP\ VP$" would have 1.

[1] The matrix $Q = \{q_{ij}\}$ depends on the matrix $S = \{s_{ij}\}$ because the latter one defines how close different grammars are to each other [8]. The accuracy of language acquisition also depends on the learning mechanism that is specified by UG.

Each of the eight grammars is named by the value of a set of parameters. Namely, if the values of the parameters (b_2, b_1, b_0) are $(1, 0, 0)$ for example, then the name of the grammar is $G_{(b_2 b_1 b_0)_2}$, that is G_4.

Table 1. The rules of CNF.

b_2	S → NP VP	b_2	S1 → VP1 NP1
	S → VP		S1 → VP1
	VP → Vi		VP1 → Vi
b_1	VP → Vt NP	b_1	VP1 → Vt NP1
b_0	VP → VP1 PP	b_1	NP1 → Det N
	NP → NP1		NP1 → N
b_0	NP → NP1 PP	b_1	PP → Prp NP1
b_0	NP → NP1 S1		

In this model, there are 15 (no recursive) rules of context free grammar as principles and each value of the parameters (b_2, b_1, b_0) determines the order of the right hand side of the rules. In Table 1, the parameters in the first column affect the rules of the second column.

Each agent utilizes the inside-outside algorithm [10] to estimate the probability of each rule. The algorithm is a kind of EM algorithm, which is a way of estimating the values of the hidden parameters of a model by a stochastic method, and is used for grammar acquisition from plain corpus in natural language processing [9, 12]. Through the communication with other agents at time t, an agent memorizes all sentences which are heard from other agents as well as those which the agent herself uttered. Then each agent estimates the application probability of each rule of the grammar and decides the parameter value (See Fig. 1.) To learn rules of the other grammars with the algorithm, agents need to own all the possible rules *a priori* with low probabilities at the initial stage. After the estimation, comparing the probabilities of two contradicting rules assigned, each agent adopts the one with higher probability at $t + 1$.

3 The Experiments

In this section, we detail our experiments. We calculate $S = \{s_{ij}\}$ and $Q = \{q_{ij}\}$ first, and observe the criteria conditions for creolization. The procedure in which agents obtain sentences and learn grammars at t is as follows:

1. An agent generates a sentence from her own grammar, and speaks it in turn to another agent. The listener agent memorizes the sentence in her memory. This is executed once for all the agents.
2. Repeat Process 1 until an agent memorizes 1,000 sentences in total.
3. Each agent estimates the probability of each rule for the sentences in her memory by the inside-outside algorithm, decides the values of parameters (b_2, b_1, b_0) from the probabilities and then decides the grammar.

3.1 The Calculation of the S Matrix

First, we calculate the similarity matrix $S = \{s_{ij}\}$, that was obtained as the probability that a speaker who uses G_i uttered a compatible sentence with G_j, which is derived by rules randomly chosen in proportion to the application probability. Each element of S was calculated from 30,000 sentences, and the result is as in Table 2. Each diagonal element, s_{ii} in S is slightly smaller than 1, because agents happen to speak sentences with rules of low probability and recognize the

Table 2. The S matrix between grammars in principles

	G_0	G_1	G_2	G_3	G_4	G_5	G_6	G_7
G_0	**.968**	.421	.217	.217	.469	.294	.169	.168
G_1	.396	**.969**	.186	.207	.224	.468	.160	.165
G_2	.209	.198	**.968**	.387	.175	.179	.484	.285
G_3	.216	.204	.412	**.968**	.183	.181	.227	.483
G_4	.485	.223	.181	.179	**.968**	.411	.207	.216
G_5	.286	.481	.181	.181	.391	**.970**	.191	.209
G_6	.167	.170	.470	.225	.210	.191	**.969**	.393
G_7	.171	.167	.293	.465	.216	.221	.418	**.970**

Table 3. The Q matrix under the simulation run.

	G_0	G_1	G_2	G_3	G_4	G_5	G_6	G_7
G_0	**.516**	.082	.114	.041	.126	.040	.060	.021
G_1	.131	**.368**	.068	.115	.084	.113	.054	.068
G_2	.136	.074	**.413**	.148	.063	.037	.084	.044
G_3	.033	.095	.081	**.479**	.024	.076	.055	.158
G_4	.157	.055	.074	.023	**.480**	.082	.096	.032
G_5	.043	.082	.037	.064	.149	**.415**	.074	.136
G_6	.069	.054	.114	.085	.115	.068	**.364**	.133
G_7	.022	.059	.041	.126	.043	.114	.079	**.517**

sentence only with the innate own grammar. In this principle, we can claim that $F(G_i, G_j) \neq 0$ among all grammars because the contradicting rules for word order always exist. Moreover, each grammar can derive a sentence 'Vi' by S $\rightarrow$ VP, VP $\rightarrow$ Vi with the probability of 0.167, so that each element in the matrix is roughly over this value. Grammars are roughly symmetrical with regard to the word order.

3.2 Creole Emergence

We distribute eight agents in eight grammars. Possible combinations of population in eight grammars are $15!/(8!7!) = 6,435$. We calculated the transitions of agents between grammars in all possible combinations.

For $\theta_c = 1.0$ in Equation 1, creole emerged in 18 ways, and for $\theta_c = 0.5$, there were 80 ways. We can analyse the relationship between creolization and the number of grammars that the agents use in t. The rate of creolization for the number of grammars was enumerated in Table 4. When $\theta_c = 1.0$, creole emerged only in case agents were classified into three or four languages. In the other cases, creole also emerged when $\theta_c = 0.5$, especially in case agents were classified into three or four languages. When agents are distributed in three languages, creole emerged in 19 out of 1,176 combinations, that is 1.62%; in four languages, 48 out of 2,450, that is 1.96%. These two cases occupied 84% of all the creolization. In case five or more languages, the possibility of a new language itself was difficult, because most languages had been already spoken by at least one agent in t.

This result means that creole is easier to emerge when agents are exposed in a situation of three or more languages. The contact of two languages tends to converge one of the two. This result coincides with actual creole [14].

3.3 Correlation between S and Q Matrices

From the result of population dynamics in the previous section, we calculated $Q = \{q_{ij}\}$ as follows :

$$q_{ij} = \frac{\text{the population of } G_j \text{ users at } t+1 \text{ in those who used } G_i \text{ at } t}{\text{the population of } G_i \text{ users at } t}. \qquad (2)$$

The result of $Q = \{q_{ij}\}$ is shown in Table 3, where $\sum_j q_{ij} = 1$. The diagonal elements, q_{ii} in the Q matrix came to the highest in each column or row. It means that agents of the dominant grammar at t tend to reuse the grammar at $t+1$. The rate that a half of, or more, agents use a same (dominant) grammar at t is over 40 percent among all the combinations. Besides the fact that q_{ii}'s have rather higher values, there seems no notable difference in q_{ij}'s ($i \neq j$). Therefore, this Q matrix does not have any meaning except that agents like to converge to the dominant grammar at t. We could not find further relationship between S and Q matrices.

Table 4. The number and the rate of creolization for the number of grammars.

G	D	Creolization	
		$\theta_c = 0.5$	$\theta_c = 1.0$
2	196	1 (0.51%)	0 (0%)
3	1176	19 (1.62%)	8 (0.68%)
4	2450	48 (1.96%)	10 (0.41%)
5	1960	11 (0.56%)	0 (0%)
6	588	1 (0.17%)	0 (0%)
Total	6435	80 (1.24%)	18 (0.28%)

G : The number of grammars.
D : The number of possible distributions.

3.4 Conditions of Creolization

In each experiment, every agent listens to sentences certain fixed times, and as a result, each agent memorizes a number of sentences. Each of these sentences is generated by another agent with her own grammar, so that sentences in a memory of an agent are divided into groups by the grammars that generated them. Thus, the number of sentences in each group comes to be proportional to the population of the grammar user. Note that this proportion is approximately identical in memories of agents. Because agents learn the grammar of next generation from this memory, and the grammar ratio of each memory is identical, the grammar tends to converge to a common one. Even in case the grammar does not converge, the population of the most dominant grammar at t consequently increases, or at least remains same at $t+1$.

We can observe the emergence of creole under the condition of Equation 1 in Fig. 2. All the agents are classified in three or four languages; for example, G_0, G_5 and G_6 in (a), each population of which are quite similar. In this case, a new language that no agent used at t, emerges at $t+1$.

Here, let us consider what feature the new language owns from the viewpoint of parameters (b_2, b_1, b_0). As for the parameter b_2 in Fig. 2(a), grammars of two agents are set to 0, and those of the other six are set to 1. Therefore, the language of the next generation tends to use those rules with $b_2 = 1$. With the same reason, b_1 tends to be 0, and b_0 to be 0, where '#' denotes a wildcard. Thus, we reason that the grammar of the new generation would have the parameters $(b_2, b_1, b_0) = (1, 0, 0)$, that is, G_4. The sample (b) is the case the grammars of all agents did not converge to a common grammar, because the parameter b_0 could not be fixed.

Therefore, creolization is strongly affected by the distribution of the population among different grammars, rather than the similarity between original grammars. The result can be generalized to include the case creole does not

(a)

	G_0	G_1	G_2	G_3	G_4	G_5	G_6	G_7
t	**2**	0	0	0	0	**3**	**3**	0
$t+1$	0	0	0	0	**8**	0	0	0

$b_2\ b_1\ b_0$	$b_2\ b_1\ b_0$	Value
0 # # : 2	1 # # : **6**	1
# 0 # : **5**	# 1 # : 3	0
# # 0 : **5**	# # 1 : 3	0

$\Rightarrow G_4$

(b)

	G_0	G_1	G_2	G_3	G_4	G_5	G_6	G_7
t	0	**3**	**1**	0	0	0	**1**	**3**
$t+1$	0	0	0	**5**	0	0	0	**3**

$b_2\ b_1\ b_0$	$b_2\ b_1\ b_0$	Value
0 # # : 4	1 # # : 4	-
# 0 # : 3	# 1 # : **5**	1
# # 0 : 2	# # 1 : **6**	1

$\Rightarrow \begin{cases} G_3 \\ G_7 \end{cases}$

Fig. 2. Samples of the result.

appear. If there is a dominant grammar group in population, the grammar survives or attracts more agents in the next generation; however, those who were attracted did not necessarily use similar grammars. Thus, we conclude that what decides the next prevalent language is the majority of each parameter.

4 Discussion

In Nowak et al. [11], the matrix S, that indicates similarities between two grammars, and Q, that represents the accuracy of grammar acquisition, played important roles in population dynamics. To simulate population dynamics on the computer, the S and Q matrices need to be defined in advance. The S matrix could be simply calculated from the given grammars. However, the definition of the Q matrix seemed to be problematic in that they considered $Q = \{q_{ij}\}$ diachronic constants. As we have mentioned in the previous section, the rate how many of the population of G_i users change their language to G_j in the next generation strongly depends upon the balance between x_i and x_j, as well as the population of other grammars. Thus, q_{ij} should be a function of abundance of population: $q_{ij}(x_0(t), x_1(t), \ldots, x_{n-1}(t))$.

Table 5 is a sample of the Q matrix which depends on one specific distribution of the population, that is, $x_1 = 0.375$ (3 of 8), $x_4 = 0.250$ (2 of 8), and $x = 0.375$ (3 of 8) at a certain time t. Each value of the matrix is the result of ten times calculation in the same way in Section 3, the distribution of the population being kept same. In this case, if the majority decision of parameters is considered, the expected dominant grammar in $t+1$ should be $G_{(1,0,1)_2}$, that is G_5. Actually, the experimental result shows that the values of q_{i5} $(i = 1, 4, 7)$ are rather higher in Table 5. Although the expected grammar may not always appear in the next generation, we can contend that the Q matrix should depend on the distribution of x_i's at t. This observation affects the definition of the differential equation of population dynamics.

In this paper, we have discussed the conditions for the emergence of creole based on a mathematical theory of the evolutionary and population dynamics.

Our contributions are summarized as follows.

- First, we adopted EM algorithm, that is one of the standard methods to find grammar rules for large corpora, and showed the actual experimental result on a large scale computer simulation.
- Secondly, we observed the qualitative conditions for creolization, that a new language which had not been used by anyone in the previous generation emerges. Our experimental results coincide with linguistic reports in the following two points; one is that creole tends to appear when there was no dominant language in terms of population, and the other is that it emerges easier when three or more languages contact rather than two languages do [14].
- Thirdly, we calculated the Q matrix on a specific distribution of population over various grammars, and predicted that the values of q_{ij} should be dependent on the abundance of population: $x_0(t), x_1(t), \cdots, x_{n-1}(t)$.

Table 5. A sample of the Q matrix which depends on a distribution of the population.

	G_0	G_1	G_2	G_3	G_4	G_5	G_6	G_7
G_1	0	0.067	0	0	0.330	0.600	0	0
G_4	0	0.050	0	0	0.550	0.400	0	0
G_7	0	0.167	0	0	0	0.800	0	0.033

Our future research target is to reconstruct the differential equation of population change, which incorporates the generation-dependent Q matrix.

References

1. Arends, J., Muysken, P., Smith, N. (eds.): Pidgins and Creoles, J. Benjamins Publishing Co. (1995)
2. Bickerton, D.: Language and Species, The University of Chicago Press (1990)
3. 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)
4. Chomsky, N.: Lectures on Government and Binding. Dordrecht: Foris (1981)
5. DeGraff, M. (ed.): Language Creation and Language Change, The MIT Press (1999)
6. Hashimoto, T., Ikegami, T.: Evolution of symbolic grammar systems, In: Moran, F., Moreno, A., Merelo, J.J., Chacon, P. (eds.): Advances in Artificial Life. Springer, Berlin (1995) pp.812-823
7. Hashimoto, T., Ikegami, T.: Emergence of net-grammar in communicating agents, BioSystems, **38** (1996) pp.1-14
8. Komarova, N.L., Niyogi, P., Nowak, M.A.: The Evolutionary Dynamics of Grammar Acquisition, J.Theor.Biol. **209** (2001) 43-59
9. Lari, K., Young, S.J.: The estimation of stochastic context-free grammars using the Inside-Outside algorithm, Computer Speech and Language, Vol.4 (1990) pp.35-56
10. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing, The MIT Press (1999)
11. Nowak, M.A., Komarova, N.L.: Towards an evolutionary theory of language, Trends in Cognitive Sciences **5**(7) (2001) pp.288-295.
12. Pereira, F. Schabes, Y.: Inside-Outside reestimation from partially bracketed corpora, Proceedings of ACL (1992)
13. Pinker, S.: The Language Instinct, W. Morrow & Co. (1994)
14. Todd, L.: Pidgins and Creoles, Routledge & Kegan Paul (1974)