SIMULATION OF CREOLIZATION BY EVOLUTIONARY DYNAMICS

MAKOTO NAKAMURA[1]

TAKASHI HASHIMOTO[2]

SATOSHI TOJO[1]

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

The purpose of this abstract is to investigate the characteristics of creole (DeGraff, 1999) using a mathematical formalization of population dynamics. Linguistic studies show that the emergence of creole is affected by contact with other languages, the distribution of population of each language, and similarities among the languages. Constructing a simulation model including these elements, we derive conditions for creolization from theoretical and numerical analyses.

Creoles are full-fledged new languages which children of the pidgin speakers acquire as their native languages. Interesting is the fact that children growing up hearing syntactically simplified languages such as pidgins develop a mature form as creoles (DeGraff, 1999). Pidgins and creoles may concern the mechanism for language acquisition of infants. Particularly, some properties of creoles imply the existence of innate universal grammar.

Simulation studies of language evolution can be represented by population dynamics, examples of which include an agent-based model of language acquisition proposed by Briscoe (2002) and a mathematical framework by Nowak, Komarova, and Niyogi (2001), who developed a mathematical theory of the evolutionary dynamics of language called the *language dynamics equation*, in which the change of language is represented as the transition of population among a finite number of languages. We modified the language dynamics based on social interaction, and then dealt with the emergence of creole (Nakamura, Hashimoto, & Tojo, 2007).

In response to the language dynamics equation, we assumed that any language could be classified into one of a certain number of grammars. Thus, the population of language speakers is distributed to a finite number (n) of grammars $\{G_1 \dots G_n\}$. Let x_i be the proportion of speakers of G_i within the total population. Then, the language dynamics is modeled by an equation governing the transition of language speakers among languages. Our model is different from the language dynamics equation by Nowak et al. (2001) in that we neglect the fitness term in terms of the biological evolution, and focus on the cultural transmission by introducing the degree of language contact, that is:

$$\frac{dx_j(t)}{dt} = \sum_{i=1}^{n} \overline{q}_{ij}(t)x_i(t) - x_j(t) \ \ , \tag{1}$$

where $\overline{Q}(t)$ $(= \{\overline{q}_{ij}(t)\})$ is the transition matrix among languages. Each element, $\overline{q}_{ij}$, is defined as the probability that a child of G_i speaker obtains G_j by the exposure to his/her parental language and to other languages. $\overline{Q}(t)$ depends on the distribution of language population at t, similarity among languages and a learning algorithm.

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

From our experiments, we observed creolization and found a correlation between the number of input sentences and the similarity among languages. Creoles emerged within a certain range of similarity. In our model, languages are defined as similarity between languages, which denotes the probability that a G_i speaker utters a sentence consistent with G_j. If we consider some situation of language contact, the target language is either very similar to speakers' own language or dissimilar at all. Replacing the similarity values with $1 - \epsilon$ for very similar languages and with ϵ for dissimilar languages, the model is very simplified and may be solved analytically. However, if we consider a creole, which is somewhat similar to other contact languages, we cannot replace the values with these simple ones. As a result, our creole model is very difficult to solve analytically. We discuss how to cope with this problem.

References

Briscoe, E. J. (2002). Grammatical acquisition and linguistic selection. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models.* Cambridge University Press.

DeGraff, M. (Ed.). (1999). *Language creation and language change.* Cambridge, MA: The MIT Press.

Nakamura, M., Hashimoto, T., & Tojo, S. (2007). Simulation of common language acquisition by evolutionary dynamics. In *Proc. of IJCAI 2007 Workshop on Evolutionary Models of Collaboration* (pp. 21–26). Hyderabad.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science*, *291*, 114-118.