Simulation of Emergence of Local Common Languages Using Iterated Learning Model on Social Networks

Makoto Nakamura
Japan Legal Information Institute,
Graduate School of Law,
Nagoya University
Email: mnakamur@nagoya-u.jp

Ryuichi Matoba
Department of Electronics
and Computer Engineering,
National Institute of Technology,
Toyama College,
Email: rmatoba@nc-toyama.ac.jp

Satoshi Tojo
School of Information Science,
JAIST,
Email: tojo@jaist.ac.jp

Abstract—**Thus far, there have been a variety of methodologies proposed on simulating the evolution of languages, each of which belongs to a different level of abstraction. Our goal is to provide a framework that represents the diachronic change in language by the contact among language communities. We propose an agent-based model with Kirby's iterated learning model and complex networks, putting learning agents on each node in the social network. Our proposed model is examined in three aspects: (i) the effectiveness of string-clipping, (ii) the relation between generations for learning and the number of local common languages, and (iii) the relation between network types and local common languages. A series of experiments shows that we have succeeded in modeling the actual situation of language change.**

Keywords–Language Evolution; Language Acquisition; Iterated Learning Model; Social Network.

I. Introduction

Thus far, simulation studies have played an important role in the field of language evolution [1] [2]. In particular, a very important function of simulation is to prove whether a prediction actually and consistently derives from a theory [3]. Thus far, there has been a variety of methodologies proposed on simulating the evolution of languages, each of which belongs to a different level of abstraction. Simulation studies for population dynamics alone include an agent-based model of language acquisition by Briscoe [4], which was developed toward a formal model of a language acquisition device. On the other hand, Nowak [5] proposed a mathematical theory of the language dynamics equation that focuses on its evolutionary aspect. The language dynamics equation is highly abstract, while an agent-based model is considered to be concrete, or less abstract.

Our goal is to provide a framework that represents the diachronic change in language by the contact among language communities. Examples are shown in the emergence of creoles in conventional linguistics and in the rapid change in Internet linguistics [6]. In fact, researchers on creoles have long known that contact speeds up language change [7] [8], in which the emergence of *pidgins* and *creoles* is one of the most interesting phenomena. Pidgins are simplified tentative languages spoken in multilingual communities. They come into being where people need to communicate but do not have a language in common. Creoles are full-fledged new languages that children of the pidgin speakers acquire as their native languages [9]. A common view is that a pidgin or creole is a language that takes its vocabulary from one language and its grammar from another. One language, usually European landowners' in colonial situations, was the original target of language learners [10], which is reflected by a social relation between a European elite and an indigenous underclass.

The framework of language contact would be useful not only for simulating typical language changes but also for novel phenomena taking place in the cyber world. In recent decades, the evolution of the Internet technology has made it possible for users to transcend physical distance to participate in discussions with anonymous people concerning their favorite topics. They do not only exchange small bits of information, but rather seem to establish a durable channel for communication among people sharing common tastes on a bulletin board system, Twitter and so on. They often employ colloquial, rather than formal, language, and their utterances are characterized by abbreviated words and reduced grammatical complexity [11]. This phenomenon is often seen in language contact, but the time and size of the language change on the Internet are extremely fast and large [6]. Using this framework, it would be possible to deal with this rapid language change as a phenomenon of language evolution.

Simulation of the change in languages has been studied in consideration of social networks. Thus far, Nakamura et al. [12] have proposed a mathematical framework for the emergence of creoles based on the language dynamics equation. Toward a more concrete analysis, they introduced a spatial structure to the mathematical framework [13] [14], in which learning agents come into contact with neighbors according to the learning algorithm. Furthermore, the spatial structure was expanded into complex networks [15]. Their studies are based on a hypothesis about the emergence of creoles; that is, language contact is likely to stimulate creolization. However, their learning mechanisms are too simple to observe language changes from a linguistic aspect because languages are defined as similarity measures in a numeric matrix.

In this paper, we set forth two hypotheses: one is that language contact leads to the emergence of local common languages (LCLs); the other is that language divergence depends on the type of network. We test the hypotheses

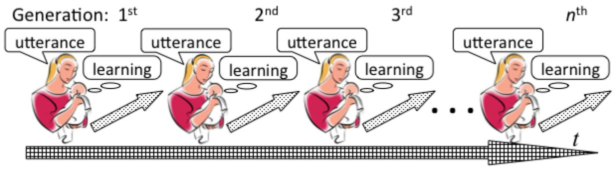


Figure 1. Iterated learning model

with a simulation model parameterizing language contact with network types. We propose an agent-based model to deal with grammatical changes in the language community. Therefore, our purpose in this paper is to show a relationship between communication among learning agents and grammatical changes. We employ Kirby's iterated learning model (ILM) [16], which shows a process of grammatical evolution through generations. This approach has often been used in simulation models concerning language evolution [17]. One important reason for this comes from its robustness against input sentences in terms of syntactic learning. As long as it is learning from a single parent, an infant agent receives sentences derived from a consistent grammar; it is possible to acquire a concise grammar. Currently, the learning situation in ILM is extended to multiple families connecting with a network. We can observe the language change, not only in a diachronic situation, i.e., in a parent-child relation, but also in a synchronic one.

Thus far, we have demonstrated a pilot version [18], where we found a problem reported by Smith and Hurford [19]; that is, in the case where learning agents potentially have more than one teacher agent, the length of syntax rules tends to increase rapidly over generations due to the addition of meaningless terminal symbols. This problem causes unnatural learning, which results in a fatal combinatorial explosion. To solve this problem, we developed a module for acceleration in ILM by string-clipping, which differs from the one for efficient learning with cognitive bias [20] [21] in that we focus on the multi-input environment.

This paper is organized as follows. In Section II, we introduce Kirby's ILM. In Section III, we propose an agent-based model for language contact. We examine our proposed method in Section IV, and conclude in Section V.

II. Modified Iterated Learning Model

In this section, we mention how to deal with ILM on social networks. First, we briefly explain Kirby's ILM. Next, we introduce the modification for social networks by Matoba et al. [24] in order to avoid the combinatorial explosion, which enables the expansion of ILM.

A. Kirby's Iterated Learning Model

Kirby [16] introduced the notions of compositionality and recursion as fundamental features of grammar, and showed that they make it possible for a human to acquire compositional language. Figure 1 illustrates ILM. In each generation, an infant can acquire grammar in his/her mind given sample sentences from his/her mother. After growing up, the infant becomes the next parent to speak to a newborn baby with his/her grammar. As a result, infants can develop more compositional grammar through the generations. Note that the model focuses on the grammar change in multiple generations, not on that in one generation. Also, Kirby adopted the idea of two different domains of language [25] [26] [27] [28], namely, I-language and E-language; I-language is the internal language corresponding to a speaker's intention or meaning, while E-language is the external language, that is, utterances. In his model, a parent is a speaker agent and his/her infant is a listener agent. The speaker agent gives the listener agent a pair of a string of symbols as an utterance (E-language), and a predicate-argument structure (PAS) as its meaning (I-language). A number of utterances would form compositional grammar rules in a listener's mind, through the learning process. This process is iterated generation by generation, and converges to a compact, limited number of grammar rules.

Verb: admire, detest, hate, like, love
Noun: john, mary, pete, heather, gavin

e.g.) *love(mary, john)*
(Identical arguments are prohibited.)

Figure 2. Words used in experiment

According to Kirby's ILM, the parent agent gives the infant agent a pair of a string of symbols as an utterance, and PAS as its meaning. The agent's linguistic knowledge is a set of a pair of a meaning and a string of symbols, as follows.

$$S/love(john, mary) \rightarrow \text{hjsbs}, \tag{1}$$

where the meaning, i.e., the speaker's intention, is represented by a PAS *love(john, mary)* and the string of symbols is the utterance "hjsbs"; the symbol *'S'* stands for the category Sentence. The following rules can also generate the same utterance.

$$\begin{array}{rcl} S/love(x, mary) & \rightarrow & \text{h } N/x \text{ sbs} \\ N/john & \rightarrow & \text{j}, \end{array} \tag{2}$$

where the variable x can be substituted for an arbitrary element of category *N*.

The infant agent has the ability to generalize his/her knowledge with learning. This generalizing process consists of the following three operations [16]: *chunk*, *merge*, and *replace*.

Chunk This operation takes pairs of rules and looks for the most-specific generalization.

$$\begin{array}{l} \left\{ \begin{array}{rcl} S/love(john, pete) & \rightarrow & \text{ivnre} \\ S/love(mary, pete) & \rightarrow & \text{ivnho} \end{array} \right. \\ \Rightarrow \left\{ \begin{array}{rcl} S/love(x, pete) & \rightarrow & \text{ivn } N/x \\ N/john & \rightarrow & \text{re} \\ N/mary & \rightarrow & \text{ho.} \end{array} \right. \end{array} \tag{3}$$

Merge If two rules have the same meanings and strings, replace their nonterminal symbols with one common symbol.

$$
\begin{cases}
S/love(x, pete) & \rightarrow & \text{ivn } A/x \\
A/john & \rightarrow & \text{re} \\
A/mary & \rightarrow & \text{ho} \\
S/like(x, gavin) & \rightarrow & \text{apr } B/x \\
B/john & \rightarrow & \text{re} \\
B/heather & \rightarrow & \text{wqi}
\end{cases}
\Rightarrow
\begin{cases}
S/love(x, pete) & \rightarrow & \text{ivn } A/x \\
A/john & \rightarrow & \text{re} \\
A/mary & \rightarrow & \text{ho} \\
S/like(x, gavin) & \rightarrow & \text{apr } A/x \\
A/heather & \rightarrow & \text{wqi.}
\end{cases}
\quad (4)
$$

Replace If a rule can be embedded in another rule, replace the terminal substrings with a compositional rule.

$$
\begin{cases}
S/love(heather, pete) & \rightarrow & \text{ivnwqi} \\
B/heather & \rightarrow & \text{wqi}
\end{cases}
\Rightarrow
\begin{cases}
S/love(x, pete) & \rightarrow & \text{ivn } B/x \\
B/heather & \rightarrow & \text{wqi.}
\end{cases}
\quad (5)
$$

In Kirby's experiment [16], five predicates and five object words shown in Figure 2 are employed. Also, two identical arguments in a predicate like *love(john, john)* are prohibited. Thus, there are 100 distinct meanings (5 predicates $\times$ 5 possible first arguments $\times$ 4 possible second arguments) in a meaning space.

The key issue in ILM is to create a *poverty of stimulus*, which explains the necessity of universal grammar [29]. Kirby [16] modeled it as learning through bottlenecks, which are rather necessary for the learning. As long as an infant agent is given all sentences in the meaning space during learning, he/she does not need to make a compositional grammar; he/she would just memorize all the meaning-sentence pairs. Therefore, agents are given a part of sentences in the whole meaning space. The total number of utterances the infant agent receives during learning is parameterized. Since the number of utterances is limited, the infant agent cannot learn the whole meaning space, the size of which is 100; thus, to obtain the whole meaning space, the infant agent has to generalize his/her own knowledge by self-learning, i.e., *chunk, merge,* and *replace*. The parent agent receives a meaning selected from the meaning space, and utters it using his/her own grammar rules. When the parent agent cannot make an utterance because of a lack of grammar rules, he/she invents a new rule. This process is called *invention*. Even if the invention does not work to complement the parent agent's grammar rules, he/she utters a randomly composed sentence.

B. Process for String-Clipping

When an infant agent has a number of teacher agents consisting of his/her parent and neighbors, as the teacher agents have their own compositional grammar rules, they are inconsistent with each other. Although the infant agent tries to find a common chunk among utterances, it would be a short string. Since there is little probability of making a chunk from short strings, only long ones are likely to survive into subsequent generations. As a result, learning agents tend to have compositional rules with extremely long strings over generations [19].

Matoba et al. [24] proposed a clipping process in their model, which solves the above problem. This process is called

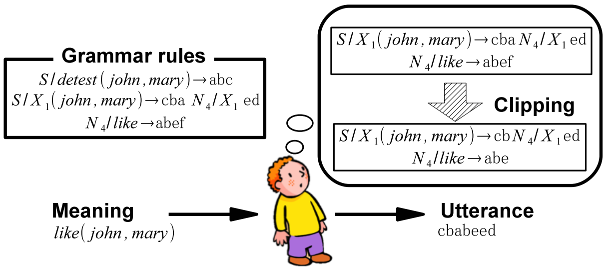


Figure 3. Image of clipping process

backclipping. After the learning process of the infant agent, he/she curtails symbols in his/her grammar rules from the tail of a string, unless it contains ambiguity. As a result, when the infant agent becomes the new parent agent in the next generation, the grammar set no longer contains extremely long rules.

Figure 3 illustrates the clipping process in our model. The infant agent tries to utter strings of *like(john, mary)* as shortly as possible. First, he/she chooses a grammar rule from his/her grammar set to generate an utterance of *like(john, mary)*, and deletes symbols one by one, i.e., "cba", "ed", "abef". In the case of "cba", this string does not exist in the grammar rules of the infant agent, so the infant agent executes backclipping, and the string is shortened from "cba" to "cb". The string "cb" does not exist in the grammar rules of the infant agent, so the infant agent executes backclipping, and the string is shortened from "cb" to "c". Since "c" exists in one of the infant's grammar rules, "$S/detest(john, mary) \rightarrow$ abc", the infant agent does not abridge it anymore, and adopts "cb" as the clipped word of "cba". The same process is also applied to the other words. As a result, the sentence becomes shortened from "cbaabefed" to "cbabeed".

Of course, such phenomena occur in the real world. The deletion of part of a word constructs a new and shorter word.

e.g.) Ham**burger** $\rightarrow$ **burger**, In**flu**enza $\rightarrow$ **flu**, **Exam**ination $\rightarrow$ **exam**

The position of clipping is dependent on a phonological reason [30]. Since ILM does not deal with phonological information, we need to find an alternative way to shorten strings. Nonetheless, backclipping by cutting off the final part of a word is the most common method of abbreviation in English [31].

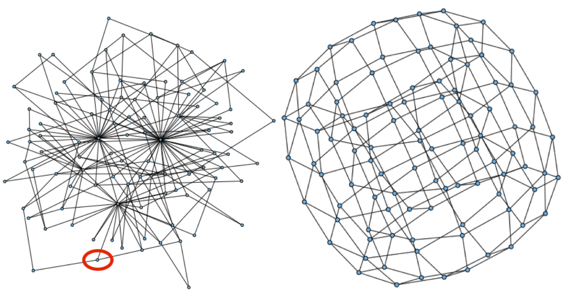
e.g.) **ad**vertisement, **doc**tor, **lab**oratory, **prof**essor, **demon**stration, **cap**tain

III. Agent-Based Model for Language Contact

In this section, we explain the agent-based model toward the emergence of local common languages (LCLs). Section III-A explains social networks for language communities. In Section III-B, we discuss the number of agents on each node in the network. We introduce a language exposure in Section III-C. Section III-D explains measuring language distance and clustering LCLs.

TABLE I. NETWORK CHARACTERISTICS

	Network type ($N_{\text{agt}} = 100$)	Average Degree	Average shortest path
	Complete graph	99.00	1.00
	Star	1.98	1.98
(a)	Scale-free	3.96	2.74
	Small-world	4.00	4.26
(b)	2D lattice	4.00	5.05
	Ring	2.00	25.25



(a) Scale-free (b) 2D-lattice

Figure 4. Examples of networks ($N_{\text{agt}} = 100$)

A. *Social Networks for Language Communities*

Social networks play an important role in language change, regardless of whether they are connected by an actual or virtual relationship. Some simulation studies deal with complex networks [15] [22] [23]. There are several types of networks, each of which characterizes many real-world communities.

Table I shows network characteristics, in which each value is calculated based on 100 nodes [23]. The average degree denotes the average number of edges connected to a node. The average shortest path length stands for the average smallest number of edges, via which any two nodes in the network can be connected to each other. In this paper, we examine scale-free and 2D-lattice networks.

Figure 4 shows examples of networks, in which the former network is regarded as a complex network and the latter is for comparison. Agents are assigned to each node in the networks.

B. *Iterated Learning Model on a Social Network*

Here, we discuss how to deal with languages in the framework of ILM on a social network. The population consists of non-overlapping generations; that is, infants at each generation are born at the same time, become parents at the same time, and die at the same time. The network is fixed through generations. The number of agents on each node depends on the model.

A group of agents on a node proceeds with learning in the following way.

1) In the group, there is a teacher agent, who makes utterances to all the learning agents.
2) Every learning agent acquires the same grammar.
3) At the end of a generation, one of the learning agents becomes the teacher agent in the next generation.
4) Repeat 1) to 3) for a certain number of generations.

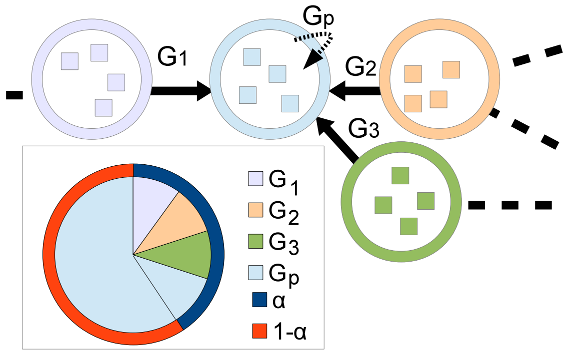


Figure 5. Language input from neighbors connected in a network depending on exposure rate α

As a result, all of the agents could acquire the same grammar. In this paper, we simplify our network model so that only one learning agent resides on each node, though we suppose the network represents a much larger community.

C. *Exposure Rate α*

Nakamura et al. [12] introduced an exposure rate α, which determines how often language learners are exposed to a variety of language speakers other than their parents. They modified the learning algorithm of Nowak et al. [5], taking the exposure rate into account in order to model the emergence of a creole community. They have shown that a certain range of α is necessary for a creole to emerge. This parameter was further employed for the following network studies [13] [14] [15].

In some communities, a child learns language not only from his/her parents but also from other adults, whose language may be different from the parental one. In such a situation, the child is exposed to other languages, and thus may learn the most communicative language. To assess how often the child is exposed to other languages, let us divide the language input into two categories; one is from his/her parents, and the other is from other language speakers. The ratio of the latter to the total amount of language input is called an *exposure rate* α. This α is subdivided into smaller rates corresponding to those other languages, where each rate is in proportion to the population of the language speakers.

Figure 5 illustrates a situation of language contact where an agent receives language input from neighbors connected in a network depending on the exposure rate α. The encircled agent in Figure 4a corresponds to the group of agents in the center. Here, we focus on the agent on each node due to the simplification in Section III-B. Let G_i be the language of Agent i. Suppose a child has parents who speak G_p; he/she receives input sentences from G_p in the proportion of $\alpha x_p + (1 - \alpha)$, and from non-parental languages $G_i (i \neq p)$ in the proportion of αx_i, where x_i denotes a population rate of G_i speakers among the neighbors.

An infant receives a meaning-signal pair from his parent and neighbors according to the exposure rate α. The number of utterances an infant receives is fixed, and he/she receives them in proportion to the language distribution for neighbors, like the pie chart shown in Figure 5.

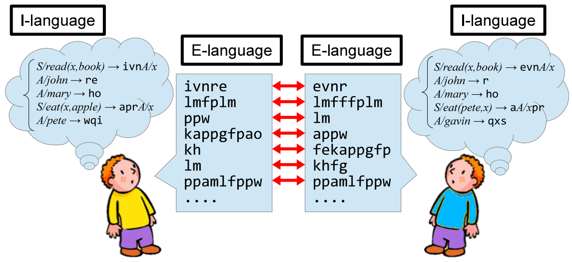


Figure 6. Calculation of distance between languages

D. Distance between Local Common Languages

To compare between languages, we define the distance in languages by the edit distance, known as the Levenshtein distance [32]; we count the number of insertion/elimination operations to change one word into another. For example, the distance between "abc" and "bcd" becomes 2 (erase 'a' and insert 'd'). Once the learning process is finished, each agent has his/her own grammar rules. In other words, each agent can enumerate all the sentences he/she can utter as E-language derived from I-language. Figure 6 depicts an image of enumeration. Note that all of the compositional grammar rules are expanded into a set of *holistic rules,* which does not include any variable, i.e., a rule consists of a sequence of terminal symbols.

Now, the comparison between a parent agent and an infant agent takes the following procedure.

1) Pick a grammar rule (g_c) that is constructed by a pair of a PAS (p_c) and an utterance (u_c) from the child's grammar rules (g_c). Choose a grammar rule ($g_p^{p_c}$) in which PAS ($p_p^{p_c}$) is the most similar to p_c from the parent's grammar rules (G_p), in terms of the Levenshtein distance. If there are multiple candidates, all of them are kept for the next process.
2) Focus on an utterance ($u_p^{p_c}$) of $g_p^{p_c}$ and u_c, and measure a distance ($d(u_c, u_p^{p_c})$) between $u_p^{p_c}$ and u_c using the Levenshtein distance. If there are multiple candidates, choose the smallest one.
3) Normalize d from 0 to 1. Carry out 1) to 3) for all grammar rules of G_c. Calculate the sum of all the distances and regard the average of them as the distance of two sets of linguistic knowledge. Thus, in this case, the distance between G_c and G_p is calculated as below:

$$\mathrm{Dist}(G_c, G_p) = \frac{1}{|G_c|}\left(\sum_{i=0}^{|G_c|} \frac{d(u_{c_i}, u_p^{p_{c_i}})}{|u_{c_i}| + |u_p^{p_{c_i}}|}\right). \quad (6)$$

Since agents independently invent languages, their acquired languages are different from each other. To classify agents into local communities by language similarity, we introduce a clustering method, recognizing a cluster as a local language community. The relationship among languages is represented by a dendrogram shown in Figure 7. The vertical axis denotes the height of the tree, which represents the language distance between two merged clusters. We employed the complete linkage method throughout the experiments. Therefore, the number of languages depends on the cutting point of the tree.

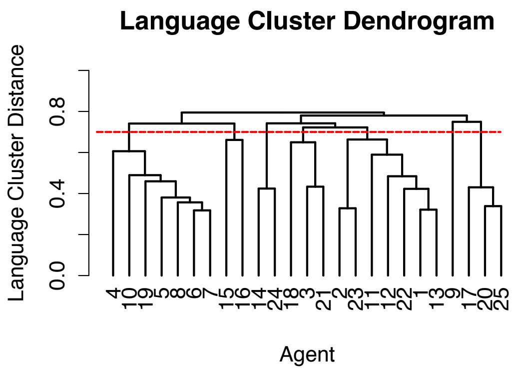


Figure 7. Clustering languages ($N_{\text{agt}} = 25$)

In this case, the community is regarded as having seven LCLs at the height of $\theta = 0.7$.

IV. Experiments

A. Experimental Settings

Our purpose in the experiments is to examine how the configuration of networks affects the language change. We evaluate the method of string-clipping for high-speed processing, comparing between ILMs with and without string-clipping. In addition, we show that the more generations the simulation takes for learning, the more compositional the grammars. Likewise, the greater the number of nodes, the more network characteristics appear.

We expect LCLs to emerge depending on the types of networks and other conditions. Therefore, we examine two types of networks: Scale-free and 2D-lattice networks. Scale-free networks are drawn with BA [33] models, where the number of edges to add in each step is 2, and the power is set to 1. The generation of multiple edges is allowed. In this paper, experiments on small-world networks drawn with WS [34] models were omitted due to a lack of outstanding results.

Table II shows the list of experiments. Exp-1 is to confirm the effectiveness of string-clipping, comparing two models in the same environment without string-clipping. Since the model without string-clipping in the network becomes memory- and time-consuming through generations, the number of generations is set to 30, which is substantially lower than the former experiment [24]. To eliminate the network effect, we examine models only in the 2D-lattice network in Exp-1. Exp-2 is to observe behaviors of the model with different generations, which are set to 30, 100 and 200. Exp-3 is to observe the relation between the configuration of networks and the language change in the large scale of networks. Therefore, the number of agents is set to $N_{\text{agt}} = 100$, while the preceding experiments employ $N_{\text{agt}} = 25$.

There are some parameters for parental and infant agents in ILM. Each infant agent receives 50 sentences, while the meaning space is 100 (5 predicates × 5 possible first arguments × 4 possible second arguments). Therefore, agents need to acquire a compositional grammar for high expressivity. When

TABLE II. LIST OF EXPERIMENTS

Experiment	N_{agt}	Generation	Network	String-Clipping
Exp-1	25	30	2D Lattice	No
				Yes
Exp-2	25	30 / 100 / 200	2D Lattice	Yes
			Scale Free	
Exp-3	100	100	2D Lattice	Yes
			Scale Free	

TABLE III. LABEL SUFFIXES

Suffix	Definition
-lcl	Number of local common languages
-exp	Degree of expressivity (%)
-cmp	Number of rules where non-terminal symbol 'S' is put on left-hand side
-lex	Number of lexical rules
-prev	Language distance from previous generation

a parent agent fails to derive a sentence from his/her own grammatical rules, he/she invents a holistic rule with a random string, the maximum length of which is set to 10. In the process of generating a sentence with his/her grammatical rules, parental agents randomly choose one of the candidate rules when facing a syntactic ambiguity. The agents are set to keep choosing the same rule through the generation, although on the other option agents could randomly choose a rule again.

B. Evaluation Methods

We measure (a) the number of *local common languages* (LCLs), (b) the number of *grammar rules* and (c) *expressivity* of the grammar. (a) is calculated by setting the threshold to distinguish languages in the dendrogram to $\theta = 0.7$. (b) denotes the average number of grammar rules created in an agent at the final generation. (c) is defined as the ratio of the number of utterable meanings derived from the grammar rules to the whole meaning space.

Since the exposure rate α activates communication with neighbors, we also expect local dialects to emerge through communication. We parameterize $0 \leq \alpha \leq 1$, where the larger the value α is, the more frequently the neighbors speak to the infant agent. Note that, even at $\alpha = 1$, infants receive sentences from their parents as frequently as from one of the neighbors.

The notation of labels in Figures 8 to 13 are defined as follows:

$$\left\{ \begin{array}{c} \text{sc} \\ \text{lat} \end{array} \right\} /a \left\{ \begin{array}{c} 25 \\ 100 \end{array} \right\} /g \left\{ \begin{array}{c} 30 \\ 100 \\ 200 \end{array} \right\} / \left\{ \begin{array}{c} \text{w} \\ \text{wo} \end{array} \right\} \left\{ \begin{array}{c} \text{-lcl} \\ \text{-exp} \\ \text{-cmp} \\ \text{-lex} \end{array} \right\}, \tag{7}$$

where the first part distinguishes a network type (**sc**ale-free / 2D-**lat**tice), the second and third numerals denote the numbers of agents and generations, respectively, and 'w' and 'wo' stand for the learning methods 'with' and 'without' string-clipping. The suffixes, which are shown in Table III, denote a type of experimental result; that is, '-lcl' and '-exp' correspond to the number of LCLs and the degree of expressivity, respectively. The other suffixes '-cmp' and '-lex' show the number of grammatical rules; the former is the number of rules where the non-terminal symbol 'S' is put on the left-hand side, which means a compositional or holistic rule, while the latter denotes the number of lexical rules.

All the data are plotted as an average of 50 trials with 95% confidence intervals. Since scale-free networks are randomly drawn for each trial, no phenomenon peculiar to a specific network appears in the results.

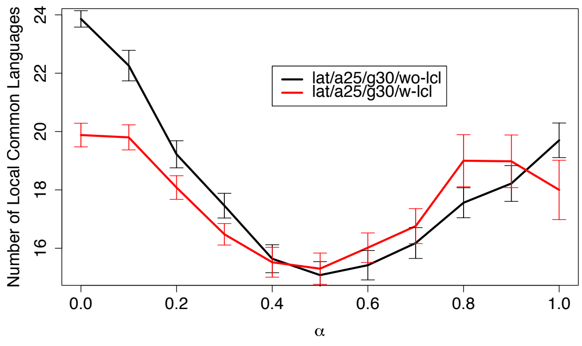
C. Experiment 1

Figure 8 shows the experimental result of Exp-1. We first explain the characteristics of whole behaviors, focusing on 'lat/a25/g30/wo', the black lines, in which 25 agents in the 2D-lattice network learn languages without string-clipping in 30 generations. Figure 8a shows the number of LCLs. At $\alpha = 0$, every agent is isolated without connection, speaking an independent language different from each other. Therefore, the number of LCLs is almost the same as that of agents. With increasing communication with neighbors at the range $0.1 \leq \alpha \leq 0.5$, agents are likely to acquire common languages. As a result, the average number of LCLs decreased to 13.2 at $\alpha = 0.5$. However, after the low peak, LCLs increased to 19.7 at $\alpha = 1$. The experimental result may suggest that excessive communication at $\alpha > 0.5$ causes language divergence. The results of LCLs form a 'V'-shape overall.

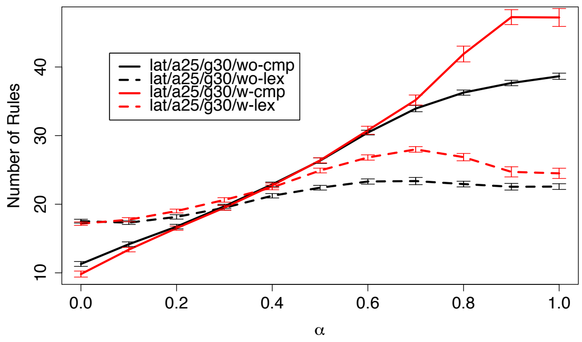
We also investigate acquired grammars. The number of grammar rules and expressivity are shown in Figures 8b and 8c, respectively. Figure 8b shows that grammars become non-compositional according to the increase of α. In general, compositional rules decrease as the whole grammar gets compositional, while lexical rules increase. Although both types of rules increase, the number of compositional rules eventually exceeds the other at $\alpha = 0.3$, which implies that holistic rules occupy agents' knowledge, while an ideal compositional grammar consists of one compositional rule and ten lexical rules. Figure 8c is inevitably reflected by the compositionality. The excessive language exposure negatively affects common languages.

Next, we explain the comparison between learning models with and without string-clipping. At $\alpha = 0$, the number of LCLs of the model with string-clipping is less than that without it. This is because the length of utterances shortens due to string-clipping, which leads to agents' languages getting closer in language distance based on the Levenshtein distance. On the other hand, with increase of α, the number of LCLs gets close to that of the model without string-clipping, and eventually the positions are reversed at $\alpha \geq 0.5$. The string-clipping method may become unstable in the excessive communication environment; at $\alpha = 0.9$, the expressivity suddenly drops to 79.1% in the model with string-clipping. Moreover, at $\alpha = 1$, the number of LCLs decreases from 19.0 to 18.0. At $\alpha = 1$, every infant agent listens to utterances equally from his/her parent and each neighbor. This uniform language sourcing may facilitate common language acquisition with string-clipping. In fact, the results are less than or insignificantly different from that of $\alpha = 0.9$.

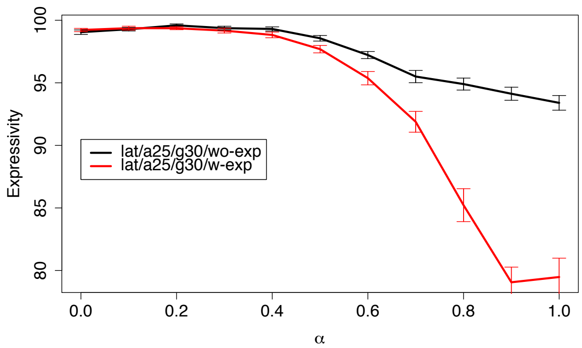
With regard to 95% confidence intervals, the difference between with/without string-clipping is significant. However, the macro-scopic behaviors seem to be similar to each other



(a) Local common languages ($\theta = 0.7$)



(b) Grammar rules



(c) Expressivity (%)

Figure 8. Experimental results of Exp-1 with 95% confidence intervals (n=50)

as both form 'V'-shaped lines, except for the large values of α. We have introduced string-clipping for computational efficiency; otherwise, a combinatorial explosion takes place in their composition of utterances after many generations. This method also supports our human behavior; that is, we do not speak sentences that are too long and beyond our presumed competency. Therefore, considering that word abbreviation actually occurs in the real linguistic utterances, we would like to justify employing this string-clipping even for the following experiments.

D. Complement of Experiment 1

We examine characteristics of LCLs with Figure 9, which shows language distance from neighbors and parents for each α. Our first expectation was simple: the more frequent the communication with neighbors, the larger the common language communities. If so, the experimental result should show that the greater the exposure rate α is, the smaller the number of LCLs. However, as shown in Figure 8a, the relation between the exposure rate α and the number of LCLs forms a 'V'-shape, where the number of LCLs reaches the lowest at $\alpha = 0.5$.

Figure 9a shows the average height of each branch in the dendrogram, which corresponds to clustering languages shown in Figure 7. Note that the height of the dendrogram represents the language distance between two merged clusters in the complete linkage method, meaning that a cluster consisting of 25 agents draws 24 lines of language distance. The red dashed line at 0.7 on the Y-axis denotes the threshold to distinguish languages, which means the number of lines above it corresponds to the number of LCLs. Therefore, every point of height would be below the threshold when everyone comes to speak a common language.

At $\alpha = 0$, since every agent learns a language only from his/her parent, all the languages spoken in the community are independent. Therefore, according to the number of grammatical rules and expressivity shown in Figures 8b and 8c, each language is sophisticated. Nevertheless, from the viewpoint of language similarity, they are far from each other.

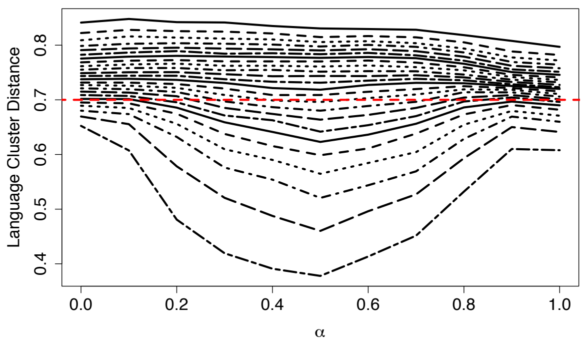
At $0.3 \leq \alpha \leq 0.5$, we can find some agents within a very close distance. This is because the agents learn a language mainly from their parents and a part from neighbors. We consider that these appropriate ranges of contact frequency succeed in producing some local common languages.

At $\alpha \geq 0.9$, although the formation in height is similar to that at $\alpha = 0$, the languages spoken in the community are quite different. Figure 9b shows the average language distance of agents' languages from their parental ones. While agents at $\alpha = 0$ learn a language similar to that of their parents, the languages at large values of α are far from that of the previous generation. We can regard them as diffusion due to learning from multipronged language sources. This is a reason for the 'V'-shape in Figure 8a.

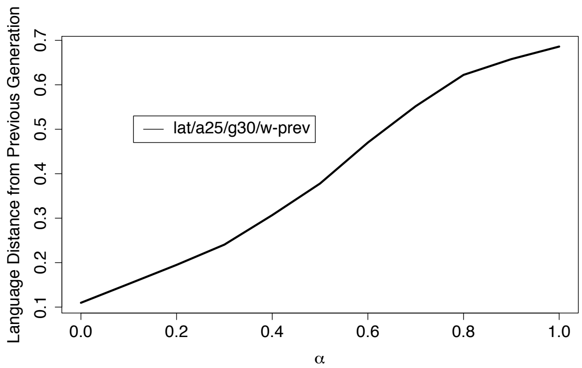
E. Experiment 2

The purpose in this experiment is to show an appropriate generation for grammar acquisition, taking effectiveness and computational time into account. We examined the model with a generation parameter at the range between 30, 100 and 200. Figure 10 shows the experimental results of Exp-2.

As far as the confidence intervals in Figure 10a, at the range of $0 \leq \alpha < 0.4$, these three results form three lines with significant difference. On the contrary, these significant differences among generation settings disappear at the range of $\alpha \geq 0.4$. This observation shows that the more generations there are for learning, the easier the observation of the emergence of LCLs at small values of α. In other words, we



(a) Height of branches in dendrogram (lat/a25/g30/w)



(b) Language distance of agents' languages from previous generation

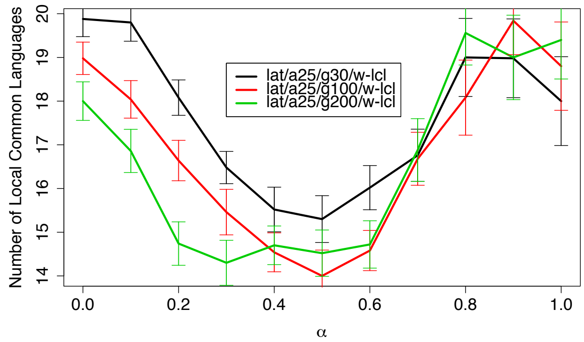
Figure 9. Language distance from neighbors and parents for each α

should take as much time as possible in the 2D-lattice network. However, Figures 10b and 10c show the lines are split into 30-th generation and the others.

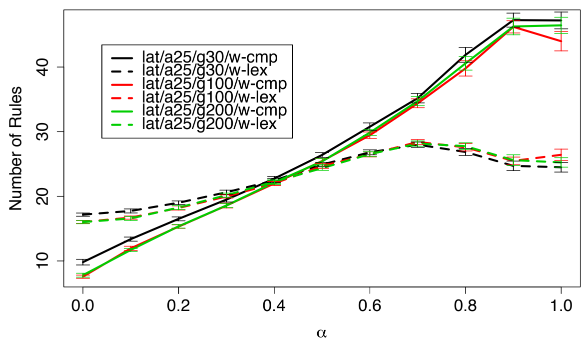
Figure 11 shows another experimental result of Exp-2 with scale-free networks. The number of LCLs in the result of the 30-th generation is significantly greater than the others at the range of $0 \leq \alpha \leq 0.5$. Rather, it is clear that the results of 100-th and 200-th generations are closer to each other. The results of grammar rules and expressivity, which are omitted, showed similar characteristics as Figures 10b and 10c. Despite a little difference between the model of the 30-th generation and the others, they are significantly different from each other at the smaller values of α. Taking computational time into account, the number of generations should be set to 100 in the following experiments.

F. Experiment 3

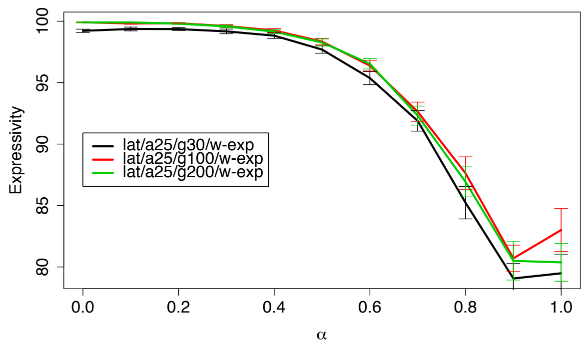
Figure 12 shows the experimental results of Exp-3 with 100 agents in 100 generations. Looking at the result of scale-free networks (black lines), we can see a large 'S'-shape in which the number of LCLs at $\alpha = 0.8$ is greater than that at $\alpha = 0$. As far as Figures 12b and 12c, agents' grammars have almost the same characteristics as that of 25 agents. This may come from the characteristics of scale-free networks.



(a) Local common languages ($\theta = 0.7$)



(b) Grammar rules



(c) Expressivity (%)

Figure 10. Experimental results of Exp-2 with 95% confidence intervals (n=50)

Figure 13 shows language distance from neighbors (Figure 13a) and parents (Figure 13b) for each α. As was shown in Section IV-D, those large values of α cause language diffusion supposedly due to too many sources of language. According to the above observation, the number of lines above the threshold ($\theta = 0.7$) at $\alpha = 0.8$ must be the greatest, since the number of LCLs at $\alpha = 0.8$ is significantly different from those at

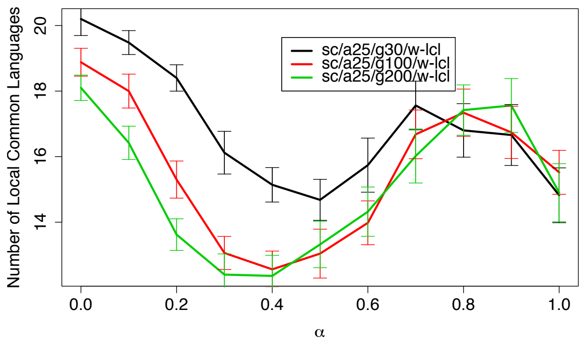


Figure 11. Local common languages ($\theta = 0.7$)

$\alpha = 0.7$ and 0.9.

We focus on the result of the 2D-lattice network (red lines). At large values of α, since every agent in the 2D-lattice network is surrounded by four neighbors and his/her parent, everyone can be an influencer in the whole network. For example, an utterance by any agent can equally affect the learning of two steps away in two generations. Note that a neighbor of two steps away is far from the other in the opposite direction, which is different from Exp-1 in that the neighbors are connected in the 2D-lattice network consisting of 5×5 vertices. As a result, the agents in such a situation may acquire not well-organized grammars, which interferes the emergence of language communities.

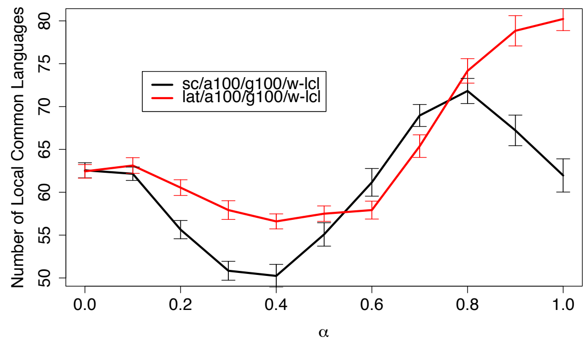
As for Figure 12a, agents in scale-free networks are likely to make local language communities at a certain range of α. Although two different types of networks drew similar curves, the one by scale-free networks reached the least number of LCLs.

Therefore, we conclude confidently that the emergence of local language communities depends on network types and an appropriate communication with neighbors.

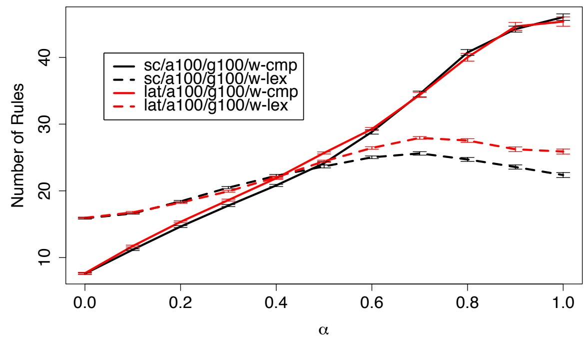
V. Conclusion

In this paper, we proposed an agent-based model for language contact. We employed Kirby's iterated learning model and complex networks. Languages were measured based on the Levenshtein distance of utterances, which enabled us to show the language divergence by the clustering. Language exposure is expected to cause neighbors to communicate with each other. Overall, we succeeded in implementing a linguistic community with learning agents connected by a social network. The network model makes it possible to observe not only diachronic but also synchronic changes in grammar. We achieved implementation of a large-scale, agent-based model where 100 processes of ILM run in parallel, which contributes to the simulation study on language evolution.

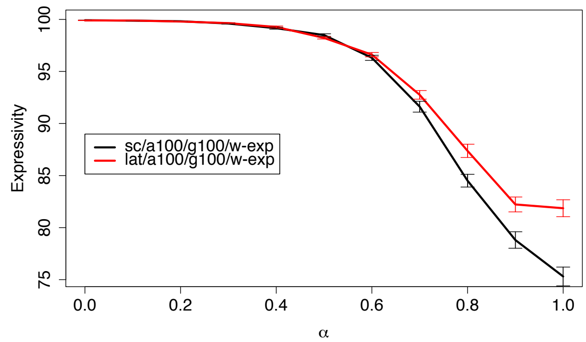
Our proposed model was examined in three aspects: (i) the effectiveness of string-clipping, (ii) the relation between generations for learning and the number of local common languages, and (iii) the relation between network types and the local common languages. Through the experiments, the language exposure was parameterized. We confirmed that the string-clipping method works effectively for grammatical learning, but it seems unstable in the multi-input environment. The complement experiment showed that excessive communication causes language divergence, which suggests that there is an appropriate degree of exposure to other languages during learning toward the emergence of a common language. We observed that local languages are likely to emerge after generations



(a) Local common languages ($\theta = 0.7$)

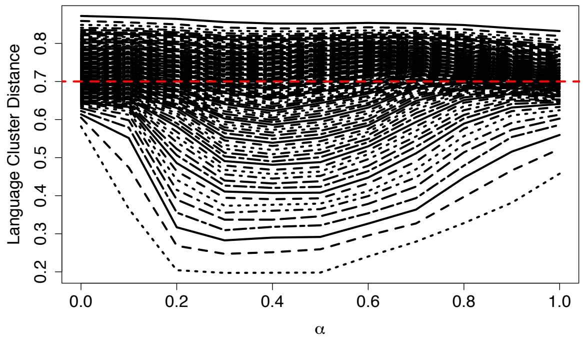


(b) Grammar rules

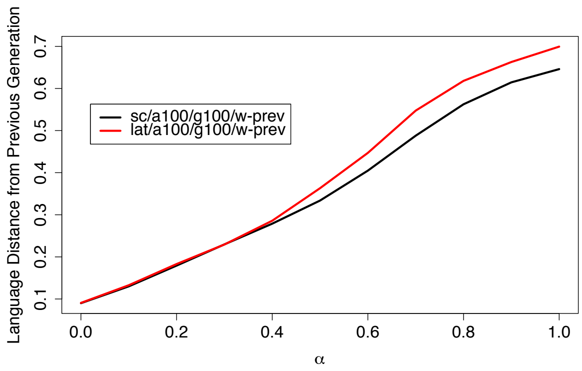


(c) Expressivity (%)

Figure 12. Experimental results of Exp-3 with 95% confidence intervals



(a) Height of branches in dendrogram (sc/a100/g100/w)



(b) Language distance of agents' languages from previous generation

Figure 13. Language distance from neighbors and parents for each α

for learning. In addition, we confirmed that the network type affects the number of local common languages. We conclude that we succeeded in modeling the actual situation of language change through the series of experiments.

In the near future, we plan to run different types of simulations toward the framework of the diachronic change in languages by language contact. Although we employed indirect graphs for simplicity in the experiments, direct graphs with weighted edges could reflect more real-world social structures.

Acknowledgment

We would like to give a special thanks to Dr. Shingo Hagiwara from the National Institute of Technology, Toyama College, and Hiroki Sudo from JAIST for technical support. This work was partly supported by Grant-in-Aid for Young Scientists (B) (KAKENHI) No. 23700310 and No. 15K16013, Grant-in-Aid for Scientific Research (C) (KAKENHI) No. 15K00201 and No. 25330434 from MEXT Japan.

References

[1] M. Nakamura, R. Matoba, and S. Tojo, "Simulation of the Emergence of Language Groups Using the Iterated Learning Model on Social Networks," in Proceedings of the 7th International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE2015), 2015, pp. 175–180.

[2] C. Lyon, C. Nehaniv, and A. Cangelosi, Eds., Emergence of Communication and Language. Springer, 2007.

[3] A. Cangelosi and D. Parisi, Eds., Simulating the Evolution of Language. London: Springer, 2002.

[4] E. J. Briscoe, "Grammatical acquisition and linguistic selection," in Linguistic Evolution through Language Acquisition: Formal and Computational Models, T. Briscoe, Ed. Cambridge University Press, 2002, ch. 9.

[5] M. A. Nowak, N. L. Komarova, and P. Niyogi, "Evolution of universal grammar," Science, vol. 291, 2001, pp. 114–118.

[6] D. Crystal, Internet Linguistics: A Student Guide, 1st ed. New York, NY, 10001: Routledge, 2011.

[7] A. Van Name, "Contributions to Creole Grammar," Transactions of the American Philological Association, vol. 1, 1869, pp. 123–67.

[8] J. Holm, "Contact and change: Pidgins and creoles," in The Handbook of Language Contact, R. Hickey, Ed. Blackwell Publishing Ltd., 2013, ch. 12.

[9] J. Arends, P. Muysken, and N. Smith, Eds., Pidgins and Creoles. Amsterdam: John Benjamins Publishing Co., 1994.

[10] M. Sebba, Contact Languages: Pidgins and Creoles, ser. Modern Linguistics. Palgrave Macmillan, 1997.

[11] D. Crystal, Txtng: The Gr8 Db8. OUP Oxford, 2008. [Online]. Available: http://books.google.co.jp/books?id=HyNVuCxTtW0C

[12] M. Nakamura, T. Hashimoto, and S. Tojo, "Exposure dependent creolization in language dynamics equation," in New Frontiers in Artificial Intelligence, ser. Lecture Notes in Artificial Intelligence, A. Sakurai, K. Hasida, and K. Nitta, Eds., vol. 3609. Springer, 2006, pp. 295–304.

[13] ——, "Self-organization of creole community in spatial language dynamics," in Proceedings of 2nd IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO2008), Venice, 2008, pp. 459–460.

[14] ——, "Prediction of creole emergence in spatial language dynamics," in LATA 2009 (Proceedings of 3rd International Conference on Language and Automata Theory and Applications), ser. Lecture Notes in Artificial Intelligence, A. H. Dediu, A. M. Ionescu, and C. Martin-Vide, Eds., vol. 5457. Tarragona: Springer, 2009, pp. 614–625.

[15] ——, "Self-organization of creole community in a scale-free network," in Proceedings of 3nd IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO2009), San Francisco, 2009, pp. 293–294.

[16] S. Kirby, "Learning, bottlenecks and the evolution of recursive syntax," in Linguistic Evolution through Language Acquisition: Formal and Computational Models, T. Briscoe, Ed. Cambridge University Press, 2002, ch. 6.

[17] M. Delz, B. Layer, S. Schulz, and J. Wahle, "Overgeneralization of Verbs - the Change of the German Verb System," in The Evolution of Language: Proceedings of the 9th International Conference (EVOLANG9), T. C. Scott-Phollips, M. Tamariz, E. A. Cartmill, and J. R. Hurford, Eds. World Scientific Pub Co. Inc., 2012, pp. 96–103.

[18] M. Nakamura, S. Hagiwara, and S. Tojo, "Multilayered formalisms for language contact," in Proceedings of WS on Constructive Approaches to Language Evolution, Kyoto, 2012, pp. 145–147.

[19] K. Smith and J. R. Hurford, "Language Evolution in Populations: Extending the Iterated Learning Model," in Proceedings of ECAL03, 2003, pp. 507–516.

[20] R. Matoba, M. Nakamura, and S. Tojo, "Efficiency of the symmetry bias in grammar acquisition," Information and Computation, vol. 209, no. 3, 2010, pp. 536–547.

[21] R. Matoba, H. Sudo, M. Nakamura, and S. Tojo, "Application of Loose Symmetry Bias to Multiple Meaning Environment," in Proceedings of the 7th International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE2015), 2015, pp. 62–65.

[22] X. Castelló, V. M. Eguíluz, M. S. Miguel, L. Loureiro-Porto, R. Toivonen, J. Saramäki, and K. Kaski, "Modelling language competition: bilingualism and complex social networks," in The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7), A. Smith, K. Smith, and R. Cancho, Eds. World Scientific Pub Co. Inc., 2008, pp. 59–66.

[23] T. Gong, L. Shuai, M. Tamariz, and G. Jäger, “Studying Language Change Using Price Equation and Pólya-urn Dynamics.” PLoS One, vol. 7, no. 3:e33171, 2012.

[24] R. Matoba, H. Sudo, M. Nakamura, S. Hagiwara, and S. Tojo, “Process Acceleration in the Iterated Learning Model with String Clipping,” International Journal of Computer and Communication Engineering, vol. 4, no. 2, 2014, pp. 100–106.

[25] D. Bickerton, Language and Species. University of Chicago Press, 1990.

[26] N. Chomsky, Knowledge of Language: Its Nature, Origin, and Use. New York: Praeger, 1986.

[27] J. R. Hurford, Language and Number: the Emergence of a Cognitive System. Oxford: Basil Blackwell, 1987.

[28] S. Kirby, Function, Selection, and Innateness: The Emergence of Language Universals. Oxford University Press, 1999.

[29] N. Chomsky, Rules and Representations. Oxford: Basil Blackwell, 1980.

[30] D. Jamet, “A Morphological Approach of Clipping in English. Can the Study of Clipping Be Formalized?” Lexis, no. 1, 2009, pp. 15–31.

[31] A. Veisbergs, “Clipping in English and Latvian,” Poznan Studies in Contemporary Linguistics, no. 35, 1999, pp. 153–163.

[32] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000.

[33] A.-L. Barabasi and R. Albert, “Emergence of scaling in random networks,” Science, vol. 286, no. 5439, 1999, pp. 509–512.

[34] D. J. Watts and S. H. Strogatz, “Collective dynamics of ‘small-world’ networks,” Nature, vol. 393, 1998, pp. 440–442.