Process Acceleration in the Iterated Learning Model with String Clipping

Ryuichi Matoba[1], Hiroki Sudo[1], Makoto Nakamura[2*], Shingo Hagiwara[1], Satoshi Tojo[3]

[1] National Institute of Technology, Toyama College, 1-2 Neriya, Ebie, Imizu-City, Toyama, 933-0293, Japan.
[2] Japan Legal Information Institute, Graduate School of Law, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan.
[3] School of Information Science, JAIST, 1-1 Asahidai, Nomi, Ishikawa, 923-1292, Japan.

* Corresponding author. Tel.: +81-52-788-6236; email: mnakamur@law.nagoya-u.ac.jp

Abstract: In evolutionary linguistics, the Iterated Learning Model (ILM) is often used for simulating a first language acquisition. Whereas an infant agent acquires a grammar through communication with his/her parent in ILM, the length of syntax rules tends to increase rapidly over generations due to the addition of symbols of meaningless terminal symbols. In the case learning agents potentially have more than one teacher agent, this problem causes an unnatural learning, which results in a combinatorial explosion. In this paper, we propose a learning method in ILM to solve the problem by string clipping. Our experimental result showed that the length of utterances decreases without potential influence in intergenerational language propagation.

Key words: Language acquisition, simulation, iterated learning model, clipping.

1. Introduction

Thus far, simulation studies have played an important role in proving a variety of linguistic theories in the field of the evolution of language [1], [2]. One great achievement is Simon Kirby's Iterated Learning Model (ILM, hereinafter) [3], which shows a process of grammatical evolution through generations. In each generation, an infant can acquire grammar in his/her mind given sample sentences from his/her mother. When the infant has grown up, he/she becomes the next parents 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. Although the poverty of stimulus explains the necessity of the universal grammar [4], Kirby [3] modeled it as *learning through bottlenecks* and such bottlenecks are rather necessary for the learning.

Kirby's ILM has often been used in simulation models concerning language evolution [5]. One important reason for this is that ILM is robust against input sentences in terms of a syntactic learning. As long as learning from a single parent, its infant agent receives sentences derived from a consistent grammar, i.e., the previous study; it is possible to acquire a concise grammar. Currently, the learning situation in ILM is extended to a multiple families connecting with a network, not only in parent-child agents, and thus we could observe not only diachronic but also synchronic changes in grammar.

On the other hand, we are faced with a problem that has been reported by Smith and Hurford [6], that is, in the case 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 symbols of meaningless terminal symbols. This problem causes an unnatural learning, which results in a fatal combinatorial explosion. Therefore, our purpose in this paper is to develop a new method in ILM that avoids the combinatorial explosion. Our approach is to remove those long strings with futile symbols. This abbreviation is often observed in actual utterance.

This paper is organized as follows. In Section 2, we introduce Kirby's ILM. In Section 3, we propose a new learning method used in ILM. In Section 4, we examine our proposed method, and conclude in Section 5.

2. Briefing Kirby's Iterated Learning Model

Kirby [3] introduced the notions of *compositionality* and *recursion* as fundamental features of grammar, and showed that they made it possible for a human to acquire compositional language. Also, he adopted the idea of two different domains of language [7]-[10], namely, I-language and E-language; I-language is the internal language corresponding to 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 listener's mind, through learning process. This process is iterated generation by generation, and converges to a compact, limited number of grammar rules.

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, that is the speaker's intension, 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.

$$S/love(x, mary) \rightarrow \text{h } N/x \text{ sbs} \tag{2}$$

$$N/john \rightarrow j \tag{3}$$

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 [3]; chunk, merge, and replace.

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

$$\begin{cases} S/love(john, pete) \rightarrow \text{ivnre} \\ S/love(mary, pete) \rightarrow \text{ivnho} \end{cases} \Rightarrow \begin{cases} S/love(x, pete) \rightarrow \text{ivn}N/x \\ N/john \rightarrow \text{re} \\ N/mary \rightarrow \text{ho} \end{cases}$$

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

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

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

In Kirby's experiment [3], five predicates and five object words shown in Fig. 1 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.

Since the number of utterances is limited to 50, 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 her own grammar rules. When the parent agent cannot utter because of lack of her grammar rules, 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 to utter, she utters a randomly composed sentence.

> **Verb:** admire, detest, hate, like, love
> **Noun:** john, mary, pete, heather
>
> e.g.) *love(mary, john)*
> (Identical arguments are prohibited.)

Fig. 1. Words used in the experiment.

3. Process for String Clipping

A serious problem takes place if an infant agent has a number of teacher agents consisting of his/her parent and neighbors. Even if 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. Therefore, the chunk process generates compositional rules with a short or long string from utterances. Since there is little probability of making a chunk from short strings, only long ones are likely to survive towards next generations. As a result, learning agents tend to have compositional rules with extremely long strings over generations.

For this reason, we introduce a clipping process in our model. After learning process of the infant agent, he/she curtails his/her grammar rules as long as it does not contain ambiguity. As a result, when the infant agent becomes the new parent agent in the next generation, the grammar set does not contain extremely long rules any more.

Fig. 2 illustrates the clipping process in our model. The infant agent tries to utter strings of *like(john, mary)* as shortly as possible. Firstly, he/she chooses a grammar rule from his/her grammar set for generating utterance of *like(john, mary)*, and deletes symbols one by one, i.e., "cba", "ed", "abef". In case of "cba", this string does not exist in the grammar rules of the infant agent, then the infant agent executes back clipping,

and the string becomes "cba" to "cb". The string "cb" does not exist in the grammar rules of the infant agent, so the infant agent executes back clipping, and the string becomes "cb" to 'c'. Since 'c' exists in the grammar rules of infant agent, 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".

Actually, such phenomenon occurs in the real world, as cutting the beginning and/or the end of a word off. The deletion of a part of a word constructs a new and shorter word;

e.g.) Ham**burger**→**burger**,   In**flu**enza → flu,   **Exam**ination → exam

A position of clipping is dependent on a phonologic reason [11]. Since ILM does not deal with phonological information, we need to find another way to how to shorten strings. Nonetheless, backclipping cutting off the final part of the word is the most common way of abbreviation in English [12];

e.g.) **ad**vertisement, **doc**tor, **lab**oratory, **prof**essor, **demo**nstration, **cap**tain, etc.

Thus our clipping process introduces backclipping, i.e., it deletes uttered strings symbol by symbol from the tail of strings.
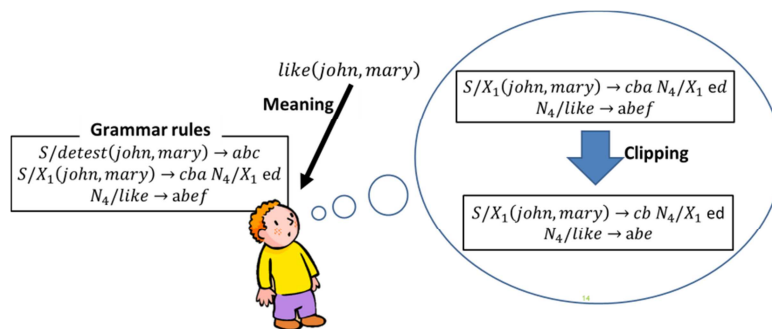

Fig. 2. Image of the clipping process.

4. Evaluating Clipping Process

In this section, we show the procedure and the result of our experiment. The purpose of the experiment is to verify the efficacy of our revision.

4.1. Evaluation Procedure

In order to examine the efficacy of our clipping process, we compare two models; one is the original Kirby's ILM and the other is ILM with our clipping process. We measure *expressivity*, and *distance in language*, as well as the number of grammar rules in each generation. The expressivity of a grammar is defined as the ratio of the number of utterable meanings derived from the grammar rules to the whole meaning space. The expressivity and the number of grammar rules represent only quality of grammar in a single agent viz., two agents who own completely different grammar rules from each other may have the same scores. But, the distance in languages represents grammatical difference between the parent and infant agents [13], i.e., they can communicate with each other with their own grammar rules.

We define the distance in languages by the edit distance, known as the Levenshtein distance; we count the number of insertion/elimination operations to change one word into the other. For example, the distance between 'abc' and 'bcd' becomes 2 (erase 'a' and insert 'd').

All the compositional grammar rules are expanded into a set of holistic rules, which do 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 up a grammar rule (g_c) which 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 parent's grammar rules (G_p), in terms of the Levenstein 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.

    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.

$$Dist_{G_c to G_p} = \frac{1}{|G_c|}\left(\sum_{i=0}^{|G_c|} \frac{d(u_{ci}, u_p^{p_{ci}})}{|u_{ci}| + |u_p^{p_{ci}}|}\right)$$

4.2. Experiments and Results

In this section, we compare two iterated learning models, i.e., with/without the clipping process. To test the efficacy of clipping, we prohibit a parent agent from inventing strings while speaking. We have carried out each experiment until 300th generation, and each result is the average of 100 trials.

Fig. 3 to Fig. 5 show the comparison of the results of ILM with/without the clipping process. Fig. 3 shows the tendency of the number of rules and expressivity per generation. In the early stages, parents' grammar is non-compositional, that is, the language has low expressivity, and the size of grammar rules is large due to holistic rules. However, operations are applied among these holistic rules over generations, and as a result, the grammar turns out to be compositional after generations, that is, the language acquires higher expressivity and the number of grammar rules decreases.

Fig. 4 shows the distance between parents and their infants in language. The closer the distance in Fig. 4 is, the more faithfully the parental language is transferred. In the early stages, their languages are quite different from each other. Since most of grammar rules of a parent agent are not compositional, his/her infant agent cannot find compositionality in them. In contrast, those utterances generated by compositional rules are more easily transferable to infant agents.
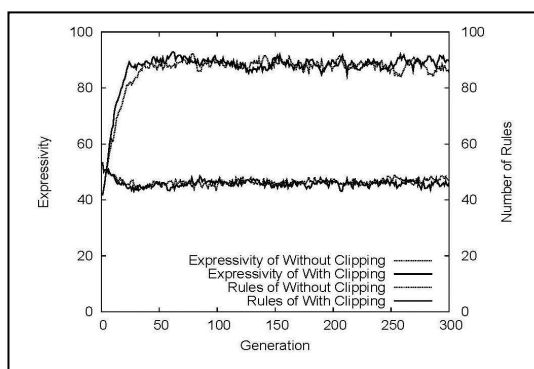

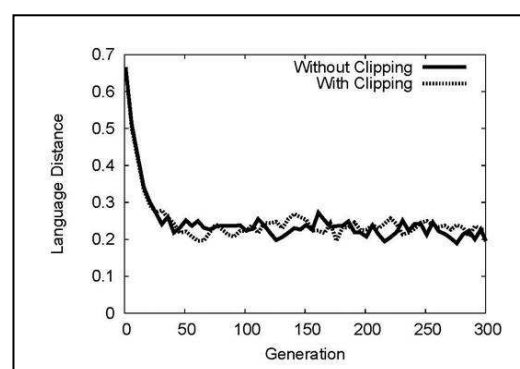Fig. 3. Expressivity and number of grammar rules.

Fig. 4. Distance in language.

Between two processes of ILM, i.e., with/without clipping, we cannot find a big difference with regard to expressivity, the number of rules, and the distance in language. Namely, our proposed clipping process does not affect the simulation. However, we found a great change in the average length of an utterance in Fig. 5. Note that each side of the figure is different from the other in digits in the vertical axis. The utterance of agents without the clipping process is going to explode over generations (the left-hand side), while that of

with the clipping process converges into 5 or 6 letters (the right-hand side).

The series of experiments showed that our proposed method has succeeded to prevent redundant utterances, keeping the expressivity and the number of rules and distance in language from the original ILM.
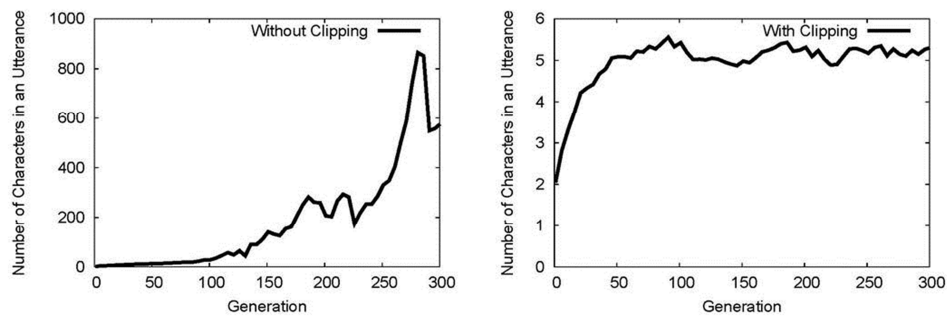

Fig. 5. Number of characters in an utterance (Left: without clipping, right: with clipping).

5. Conclusion

In this paper, we proposed a new method for learning syntax rules in ILM. The advantage of our method is to avoid increasing the length of utterance, which results in an efficient process, even if an infant agent learns from more than one parental agent. Experimental results showed that our proposed method works better than the original, keeping its expressivity, the number of rules and distance in language from the parent, which shows that we have succeeded in the improvement of the entire process.

In the near future, we plan to model a linguistic community with learning agents connected with a social network. The network model makes it possible to observe not only diachronic but also synchronic changes in grammar. For example, we could observe an emergence of dialect in a local area. Furthermore, although our method is based on an actual abbreviation process, we employed it just for faster learning in the process. In other words, our framework may contribute to modeling the abbreviation process, considering that of the real world.

References

[1] Cangelosi, A., & Parisi, D. (2001). *Simulating the Evolution of Language*. Springer.
[2] Lyon, C., Nehaniv, C., & Cangelosi, A. (2007). *Emergence of Communication and Language.* Springer.
[3] Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models,* 173-203. Cambridge, MA: Cambridge University Press.
[4] Chomsky, N. (1980). *Rules and Representations*, Oxford: Basil Blackwell.
[5] Delz, M., Layer, B., Schulz, S., & Wahle, J. (2012). Overgeneralization of verbs – the change of the German verb system. *The Evolution of Language*, 96-103.
[6] Smith, K., &Hurford, J. (2003). Language evolution in populations: extending the iterated learning model. *Advances in Artificial Life, Lecture Notes in Computer Science, 2801,* 507-516.
[7] Bickerton, D. (1990). *Language and Species*. University of Chicago Press.
[8] Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Praeger.

[9] Hurford, J. (1987). *Language and Number: the Emergence of a Cognitive System*. Blackwell.

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

[11] Jamet, D. (2009). A morphological approach of clipping in English. Can the study of clipping be formalized. *Lexis-E-Journal in English Lexicology*, 15-31.

[12] Andrejs, V. (1999). Clipping in English and Latvian. *Poznan Studies in Contemporary Linguistics,* 35, 153-163.

[13] Matoba, R., Sudo, Hagiwara, S., & Tojo., S. (2013). Evaluation of the Symmetry Bias in Grammar Acquisition. *Proceedings of the International Symposium on Artificial Life and Robotics*. Daejeon, Korea.

Ryuichi Matoba received his Master and Ph.D. degrees in information science from Japan Advanced Institute of Science and Technology in 2006 and 2010, respectively.

He is currently an associate professor of National Institute of Technology, Toyama College, Department of Electronics and Computer Engineering.

Hiroki Sudo received his associate degree from National Institute of Technology, Toyama College, Department of Information Engineering in 2013. He is currently a student of Advanced Course of Control Information System Engineering Program at NIT Toyama.

Makoto Nakamura received his master and Ph.D. degrees in information science from Japan Advanced Institute of Science and Technology in 1997 and 2004, respectively.

He is currently a designated assistant professor of Japan Legal Information Institute, Graduate School of Law, Nagoya University.

Shingo Hagiwara received his master and Ph.D. degrees in information science from Japan Advanced Institute of Science and Technology in 2004 and 2009, respectively.

He is currently an associate professor of Department of International Business, National Institute of Technology, Toyama College.

Satoshi Tojo received his master and Ph.D. degrees in engineering from the University of Tokyo in 1983 and 1995, respectively.

He is currently a professor of School of Information Science, Japan Advanced Institute of Science and Technology.