

Application of Loose Symmetry Bias to Multiple Meaning Environment

Ryuichi Matoba
and Hiroki Sudo

Department of Electronics
and Computer Engineering,
National Institute of Technology,
Toyama College
Email: {rmatoba, i08323}
@nc-toyama.ac.jp

Makoto Nakamura

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

Satoshi Tojo

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

Abstract—It is well known that the cognitive biases much accelerate the vocabulary learning. In addition, other works suggest that cognitive biases help to acquire grammar rules faster. The efficacy of the cognitive biases enables infants to connect an utterance to its meaning; even a single uttered situation contains many possible meanings. In this study, we focus on the symmetry bias which is one of the cognitive biases. The aim of this study is to evaluate the efficacy of the symmetry bias in the multiple meaning environment. In the experiments, two symmetry bias patterns are utilized to evaluate the developed Meaning Selection Iterated Learning Model. The patterns are strict/loose symmetry bias with distance in languages and expressivity.

Keywords—Symmetry Bias; Iterated Learning Model; Language Acquisition.

I. INTRODUCTION

When learning a foreign language, learners might translate foreign words into their first language verbatim using a dictionary, and looking up grammar textbooks, or getting taught the foreign language by somebody who has already acquired it. In other words, learners use their first language to grasp meanings of the foreign language. In the case of the first language acquisition, learners do not know any languages to translate to figure out meanings of input utterances. From this aspect, the first language acquisition is more difficult than the second language, since an infant has no way to understand the meaning of each utterance. Thus, the infant has to identify the parent's intention from the situation. Under an environment which contains possibilities of many missteps to infer the parent's intention, it is hard to imagine that the infant smoothly acquires the first language.

In spite of this situation, infants can acquire new words very rapidly and also learn a word's meaning after just a single exposure [1], through fast mapping [2]. Under the above circumstance, various kinds of cognitive biases such as the shape bias [3] [4], the mutual exclusivity bias [5] [6], the whole object bias [7] and so on, work for infants to limit the possible word meanings [8] [9]. The definition of the cognitive bias is in the following quotation [10]:

Cognitive bias Systematic error in judgment and decision-making common to all human beings which can be due to cognitive limitations, motivational factors, and/or adaptations to natural environments.

In this paper, we especially focus on the symmetry bias, and investigate its efficacy using computer simulation. The

symmetry bias makes a strong relation between an object and its label by characterizing mapping among them as symmetric, i.e., it allows infants who are taught that a red sphere has a lexical label “apple”, to make the reverse implication on their own, namely that the label “apple” refers to the red sphere object [11]. This tendency is said to be one of the peculiar human skills, and many experiments have endorsed that other animals cannot do this reverse implication [12].

This study aims to evaluate the efficacy of the symmetry bias using computer simulation. So far, we simulated the efficacy of the symmetry bias [13] using Iterated Learning Model (ILM) [14], and constructed joint attention frame in learning environment of infant agents [15]. Moreover, we formulated a method for measuring language distance to indicate the efficacy of the symmetry bias [16], and constructed the Meaning Selection Iterated Learning Model (MSILM) where a pair of a parent agent and an infant agent resides in a generation, and the infant agent becomes the parent agent of the next generation. In MSILM, the parent agent and the infant agent are given multiple meanings under the situation, where two agents share a common attention. The symmetry bias in our model works is based on the similarity of utterances.

In the last few years, our study suggests that the symmetry bias which connects a grammar and a meaning with complete symmetry does not accelerate effective grammar acquisition. In this paper, using MSILM, we evaluated the efficacy of two patterns of the symmetry bias which are strict/loose symmetry bias, which will be explained later.

This paper is organized as follows. In Section II, we introduce ILM and MSILM. In Section III, we examine our proposed method, and conclude in Section IV.

II. ILM WITH MEANING SELECTION

A. Briefing Kirby's ILM

Our study is based on ILM by Simon Kirby [14], who 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 [17]–[19] which are I-language and E-language. The former is the internal language corresponding to the speaker's intention or meaning, while the other is the external language, that is, utterances. In Kirby's ILM, a speaker is a parent agent and a listener is an infant 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). The agent's grammar is a set of a pair of a meaning and a string of symbols, as shown in formula (1).

$$S/\text{love}(\text{john}, \text{mary}) \rightarrow \text{lovejohmary} \quad (1)$$

where the meaning, that is the speaker's intension, is represented by a PAS $\text{love}(\text{john}, \text{mary})$ and the string of symbols is the utterance "lovejohmary"; the symbol 'S' stands for the category Sentence. The following rules can also generate the same utterance.

$$S/\text{love}(x, \text{mary}) \rightarrow \text{loveN/xmary} \quad (2)$$

$$N/\text{john} \rightarrow j \quad (3)$$

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

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 finally, a certain generation would acquire a compact, limited number of grammar rules. The learner agent has the ability to generalize his/her grammar with learning. The learning algorithm consists of the following three operations; *chunk*, *merge*, and *replace* [14].

1) *Chunk*: This operation takes pairs of rules and looks for the most-specific generalization. For example,

$$\begin{cases} S/\text{read}(\text{john}, \text{book}) \rightarrow \text{ivnre} \\ S/\text{read}(\text{mary}, \text{book}) \rightarrow \text{ivnho} \end{cases} \quad (4)$$

↓

$$\begin{cases} S/\text{read}(x, \text{book}) \rightarrow \text{ivnN/x} \\ N/\text{john} \rightarrow \text{re} \\ N/\text{mary} \rightarrow \text{ho} \end{cases} \quad (5)$$

A rule without variables, i.e., the whole signal indicates the whole meaning of a sentence is called a *holistic rule*, while a rule with variables is called a *compositional rule*. In the case of the above example, two holistic rules become one compositional rule and two holistic rules by chunk operation.

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

$$\begin{cases} S/\text{read}(x, \text{book}) \rightarrow \text{ivnA/x} \\ A/\text{john} \rightarrow \text{re} \\ A/\text{mary} \rightarrow \text{ho} \\ S/\text{eat}(x, \text{apple}) \rightarrow \text{aprB/x} \\ B/\text{john} \rightarrow \text{re} \\ B/\text{pete} \rightarrow \text{wqi} \end{cases} \quad (6)$$

↓

$$\begin{cases} S/\text{read}(x, \text{book}) \rightarrow \text{ivnA/x} \\ A/\text{john} \rightarrow \text{re} \\ A/\text{mary} \rightarrow \text{ho} \\ S/\text{eat}(x, \text{apple}) \rightarrow \text{aprA/x} \\ A/\text{pete} \rightarrow \text{wqi} \end{cases} \quad (7)$$

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

$$\begin{cases} S/\text{read}(\text{pete}, \text{book}) \rightarrow \text{ivnwqi} \\ B/\text{pete} \rightarrow \text{wqi} \end{cases} \quad (8)$$

↓

$$\begin{cases} S/\text{read}(x, \text{book}) \rightarrow \text{ivnB/x} \\ B/\text{pete} \rightarrow \text{wqi} \end{cases} \quad (9)$$

In Kirby's experiment [14], five predicates and five object words are employed. Also, two identical arguments in a predicate like "hate(mary, mary)" 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 in his experiment, 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.

B. Briefing MSILM

Our model, MSILM (see Figure 1), introduces the notion of joint attention frame, as mentioned the previous section, into the ILM. In MSILM, multiple meanings are presented to both the parent and the infant agent, and the parent agent mentions one of them. The infant agent listens to the utterance from the parent agent, and infers its meaning from the presented meanings using an inference strategy, i.e., the symmetry bias. This model represents a situation in which the infant agent does not always acquire a unique meaning of a parent's utterance.

In our model, we changed two points of Kirby's model, which are (i) taking away the transmittance of meanings between the parent and the infant, and (ii) introducing a set of multiple meanings which contains more than one meaning. This would cause a significant difference from the result of ILM, namely the infant agent has a possibility to connect a parent's utterance to a meaning which is not that of the parent's intention of the utterance, and this leads the infant to acquire a far different grammar from the parent.

So far, following the evaluation method of Kirby, we have only used expressivity which is defined as the ratio of the number of utterable meanings derived from the grammar rules to the whole meaning space, and the number of rules of grammar to evaluate agent's grammar efficacy. However, our motivation of this study is to evaluate the efficacy of the symmetry bias by measuring the differences of grammar between the infant and the parent in a quantitative way. Therefore, we have introduced the distance in languages as well as expressivity as an evaluating method for the infant's acquired grammar. For evaluating the distance of two grammars, 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

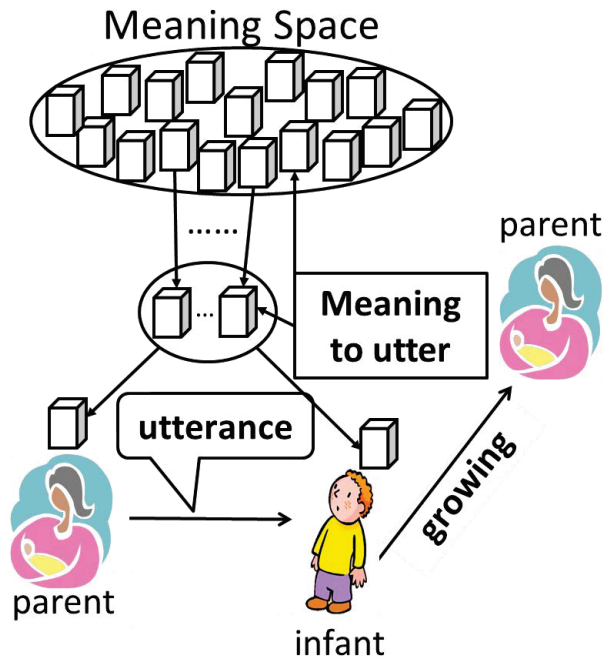


Fig. 1. Illustration of Meaning Selection Iterated Learning Model.

another. 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 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.
- 4) 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 Formula(10).

$$Dist_{G_c to G_p} = \frac{1}{|G_c|} \left(\sum_{i=0}^{|G_c|} \frac{d(u_{ci}, u_p^{p_{ci}})}{|u_{ci}| + |u_p^{p_{ci}}|} \right) \quad (10)$$

The image of this measuring procedure is shown in Figure 2.

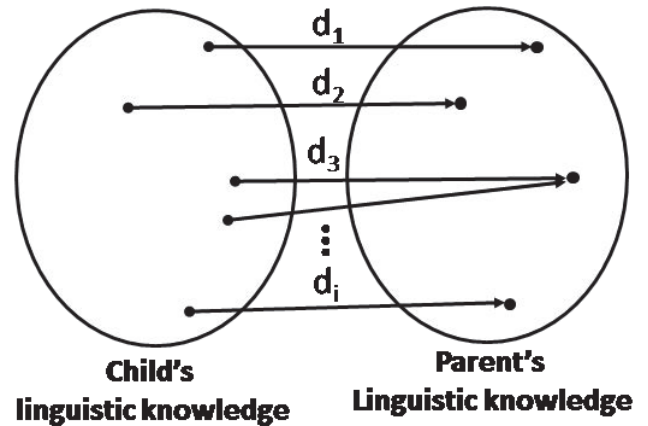


Fig. 2. Image of measuring procedure.

III. EXPERIMENT AND RESULT IN MSILM

In the simulation, preserving Kirby's settings, we employed five predicates and five object words, also prohibited two identical arguments in a predicate. This implies the size of meaning space of the experimental model is 100. A point of difference is the number of meanings which are presented to the agents. In the experiment, two meanings are presented to both the parent and the infant agent, and we examined the following three strategies when the infant agent infers the meaning of a parent's utterance from two meanings.

- 1) **Random:** The infant agent chooses a meaning from presented meanings randomly as a meaning of a parent's utterance.
- 2) **Strict Symmetry Bias:** If the infant agent can generate the same utterance as the utterance from the parent agent using own grammar, and its meaning is found in the presented meanings, he/she connects the utterance and its meaning. Otherwise, the infant agent employs the random strategy.
- 3) **Loose Symmetry Bias:** If strict strategy fails, the infant agent compares all utterances which he/she can generate to the parent's utterance using Levenshtein distance, and chooses the most similar one. Next, he/she compares a meaning of the selected utterance to presented meanings, and chooses the most similar meaning from the presented meanings.

Figures 3 and 4 show the average tendency of expressivity and the distance in languages per each generation, after 100 trials. Each line denotes the result of the strategies of random, strict symmetry bias and loose symmetry bias, respectively.

From Figure 3, we can observe that expressivity of the agent who takes the loose strategy records the highest value of the three strategies, also, the strict strategy is the lowest despite the infant agent applies the symmetry bias. In the case of applying strict symmetry bias, the infant agent receives information that he/she already knows from the parent agent, i.e., he/she does not get new information. Therefore, expressivity of the infant agent who employs strict strategy records the lowest.

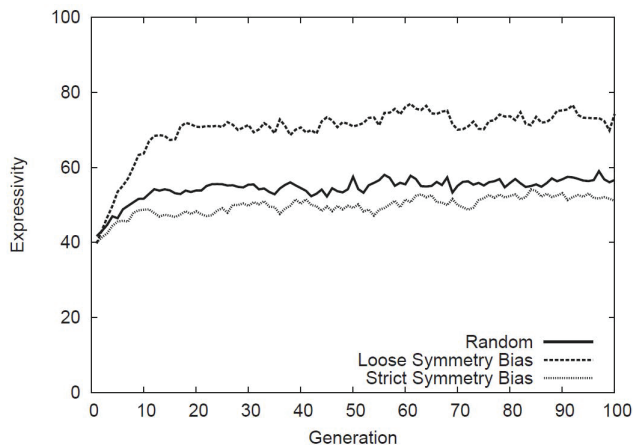


Fig. 3. The movement of the expressivity per generation.

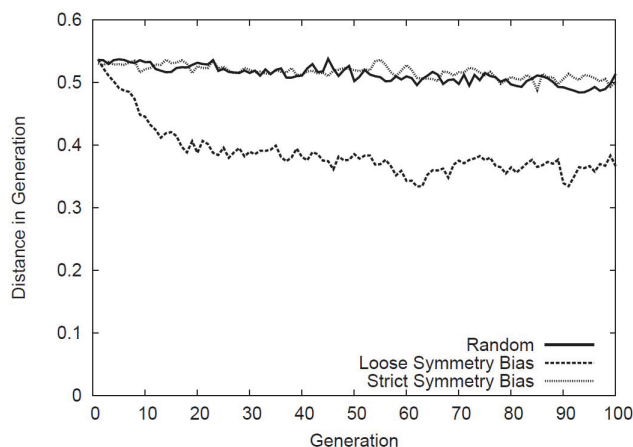


Fig. 4. The movement of the expressivity per generation.

From Figure 4, we can observe that the distance of loose strategy is the smallest, i.e., grammar of the infant agent and the parent agent is the most similar of the three strategies. For the above reasons, loose symmetry bias is the most effective strategy to acquire the parent’s grammar of the three strategies.

IV. CONCLUSION

In this study, we verified the efficacy of the symmetry bias not only in the lexical acquisition, but also in the grammar acquisition. For this purpose, we have revised Kirby’s model [14] to MSILM, and have built two kinds of symmetry bias, which are strict symmetry bias and loose symmetry bias. In the simulation, both of the parent and the infant agent are presented with multiple meanings, and the parent agent chooses one of them to utter. The infant agent receives an utterance and infers its meaning with three kinds of strategies which are random, strict symmetry bias, and loose symmetry bias.

For each of the strategies, we have observed expressivity, and the distance in languages. As a result of the experiments, the infant agent who has employed loose strategy,

- could acquire the highest expressivity,

- could construct the most similar grammar to his/her parents.

Our future works are summarized as follows. So far, we have only implemented the symmetry bias to a computer simulation model, and not compared it to a phenomenon of actual world yet. We should describe the efficacy of the symmetry bias of our model based in a real world experience.

ACKNOWLEDGMENT

This work was partly supported by Grant-in-Aid for Young Scientists(B)(KAKENHI)No. 23700310, and Grant-in-Aid for Scientific Research(C)(KAKENHI)No.25330434 from MEXT Japan.

REFERENCES

- [1] S. Carey and E. Bartlett, “Acquiring a single new word,” *Child Language Development*, vol. 15, 1978, pp. 17–29.
- [2] J. S. Horst and L. K. Samuelson, “Fast mapping but poor retention by 24-month-old infants,” *Infancy*, vol. 13, 2008, pp. 128–157.
- [3] B. Landau, L. B. Smith, and S. S. Jones, “The importance of shape in early lexical learning,” *Cognitive Development*, vol. 3, no. 3, 1988, pp. 299–321.
- [4] —, “Syntactic context and the shape bias in children’s and adult’s lexical learning,” *Journal of Memory and Language*, vol. 31, no. 6, 1992, pp. 807–825.
- [5] E. M. Markman, “Constraints children place on word meanings,” *Cognitive Science: A Multidisciplinary Journal*, vol. 14, no. 1, 1990, pp. 57–77.
- [6] E. M. Markman, J. L. Wasow, and M. B. Hansen, “Use of the mutual exclusivity assumption by young word learners,” *Cognitive Psychology*, vol. 47, no. 3, 2003, pp. 241–275.
- [7] E. M. Markman, *Categorization and naming in children: Problems of induction*. Cambridge: MIT Press, 1989.
- [8] M. Imai and D. Gentner, “Children’s theory of word meanings: The role of shape similarity in early acquisition,” *Cognitive Development*, vol. 9, no. 1, 1994, pp. 45–75.
- [9] —, “A crosslinguistic study of early word meaning: Universal ontology and linguistic influence,” *Cognition*, vol. 62, no. 2, 1997, pp. 169–200.
- [10] A. Wilke and R. Mata, *Cognitive Bias*. Academic Press, 2012, vol. 1.
- [11] M. Sidman, R. Raizin, R. Lazar, S. Cunningham, W. Tailby, and P. Carrigan, “A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children,” *Journal of the Experimental Analysis of Behavior*, vol. 37, 1982, pp. 23–44.
- [12] Y. Yamazaki, “Logical and illogical behavior in animals,” *Japanese Psychological Research*, vol. 46, no. 3, 2004, pp. 195–206.
- [13] R. Matoba, M. Nakamura, and S. Tojo, “Efficiency of the symmetry bias in grammar acquisition,” *Information and Computation*, vol. 209, 2011, pp. 536–547.
- [14] S. Kirby, *Learning, Bottlenecks and the Evolution of Recursive Syntax*. Cambridge University Press, 2002.
- [15] R. Matoba, S. Shoki, and H. Takashi, “Cultural Evolution of Compositional Language under Multiple Cognition of Meanings,” in *Proceedings of the 15th International Symposium on Artificial Life and Robotics(AROB 2010)*, 2010.
- [16] R. Matoba, H. Sudo, S. Hagiwara, and S. Tojo, “Evaluation of Efficiency of the Symmetry Bias in Grammar Acquisition,” in *Proceedings of the 18th International Symposium on Artificial Life and Robotics(AROB 2013)*, 2013, pp. 444–447.
- [17] J. Hurford, Ed., *Language and Number: the Emergence of a Cognitive System*. Blackwell, 1987.
- [18] D. Bickerton, Ed., *Language and Species*. University of Chicago Press, 1990.
- [19] S. Kirby, *Function, Selection, and Innateness: The Emergence of Language Universals*. Oxford University Press, 1999.