The Eighteenth International Symposium on Artificial Life and Robotics 2013 (AROB 18th '13),
Daejeon Convention Center, Daejeon, Korea, January 30-February 1, 2013

Evaluation of Efficiency of the Symmetry Bias in Grammar Acquisition

R. Matoba[*1], H. Sudo[*1], S. Hagiwara[*1], and S.Tojo[*2]

**1 Toyama National College of Technology*
1-2, Neriya, Ebie, Imizu, Toyama, 933-0293, Japan
(tel:+81-766-86-5283; Fax:+81-766-86-5110)
mail:{rmatoba ,i08323, hagiwara}@nc-toyama.ac.jp

**2 Japan Advanced Institute of Science and Technology*
1-1, Asahidai, Nomi, Ishikawa, 923-1292, Japan
(Tel : +81-0761-51-1372; Fax :+ 81-0761-51-1149)
mail: tojo@jaist.ac.jp

Abstract: The aim of this study is to evaluate the efficacy of cognitive biases in grammar acquisition. So far, we have constructed Iterated Learning Model (ILM) in generation by generation, where a pair of a parent agent and an infant agent resides in a generation, and the infant becomes the parent of the next generation. Then, we have added the effect of such cognitive biases as symmetry bias and mutual exclusivity bias. Thus far, in evaluating results of acquired grammar, the expressivity and the number of grammar rules have been focused on. When we consider the efficacy of cognitive biases, however, we should observe how the infant agent could learn the parent agent's intentions faithfully. Therefore, in this paper, we suggest assessing the difference of linguistic knowledge between a parent agent and an infant agenet, using Levenshtein distance. As a current result, we could observe that the distance becomes further in generations.

Keywords: Cognitive Bias, Levenshtein distance, Grammar Acquisition, Iterated Learning Model.

1. Introduction

It is well known that infants over 18 months old can acquire new words very rapidly, e.g., 7 to 15 words a day [1]. To enable this phenomenal learning, cognitive biases which are constraint to limit possibilities of words-meaning mapping in a situation have suggested [2, 3]. This mapping is considered to be generally difficult as is well known as `gavagai problem' [4], though infants achieve this operation.

So far, we have already reported efficacy of cognitive biases in grammar acquisition using computer simulation [5, 6]. Our model is based on Simon Kirby's Iterated Learning Model [7] where a pair of a parent agent as and an infant agent resides in a generation. In this model, the parent agent is a speaker and the infant agent is a listener. A number of utterances would form a compositional grammar rules in the infant agent's mind, being substrings are chunked. This process is iterated generation by generation, and finally, a certain generation would acquire a compact, limited number of grammar rules. We include cognitive biases into this process. We implement agents with the bias in a virtual world, and make them learn a grammar by computer simulation.

In general, the more grammar is compositional, the higher expressivity and the less number of rules. Therefore, we have evaluated the progress of evolution by these two criteria, so far. However, when we pay attention to the efficacy of cognitive biases, we should observe how the infant agent could acquire the parent agent's intentions correctly. Therefore, we need to assess the difference of linguistic knowledge, consisting of pairs of meanings and utterances, between the parent agent and the infant agent. More precisely, we employed a notion of Levenshtein distance to evaluate two different linguistic knowledge. Applying our suggested distance, we have gotten plausible results of Kirby's ILM to measure the efficacy of cognitive biases.

2. Distance between Two Different Linguistic Knowledge

2.1. Agent's Linguistic Knowledge

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

S/love(john, mary) → lovejohnmary

Where a speaker's intention is a PAS l*ove(john, mary)* and its utterance becomes `lovejohnmary'; the symbol `S' stands for the category Sentence. After the listener receives a pair of a PAS and an utterance, she/he tries to guess his/her parent's grammar rules, as utterances are always paired with their meanings, which are intrinsically compositional. This guessing process consists of the following two operations; `chunk' is to find a common substring, and to substitute it for a new category.

S/read(john, book) → johnreadsbook
S/read(mary, book) → maryreadsbook

↓

S/read(x, book) → N/x readsbook

while `merge' is to unify the identical category names as:

N/mary → mary
N/john → john
B/john → john

↓

N/john→ john

As a result, an infant comes to acquire a set of context-free grammar rules.

2.2. Distance between Two Knowedge

For evaluating the distance of two linguistic knowledge, i.e., the distance between the parent knowledge and the infant one, we employed the edit distance, aka Levenshtein distance, which is a metric for measuring the difference between two sequences of symbols; 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 developed to a set of holistic rules, where one PAS is connected directly to a superficial string excluding intermediate categories, beforehand. Now the comparison between a parent agent and an infant agent takes the following procedure.

1. Pick up a grammar rule (g_p) which is constructed by a pair of a PAS(p_p) and an utterance (u_p) from the parent's knowledge (K_p). Choose a grammar rule (g_c) which PAS (p_c) is most similar to p_p from a child's knowledge (K_c), in terms of Levenstein distance. If there are multiple candidates, all of them are kept for the next process.
2. Focus on an utterance (u_c) of g_c and u_p, and measure a distance (d_1) between u_p and u_c using Levenshtein distance. If there are some, choose smallest one.
3. Normalize d1 from 0 to 1.
4. Carry out 1 to 4 for all grammar rules of K_p. Calculate the sum of all the distances and regard the average of them as the distance of two linguistic knowledge. Thus, in this case, the distance between K_c and K_p is calculated as below.

$$Dist_{K_p to K_c} = \frac{1}{i}\left(\sum^{i} \frac{d_i}{|u_{ci}| + |u_{pi}|}\right)$$

The image of this measuring procedure is show in Fig1.

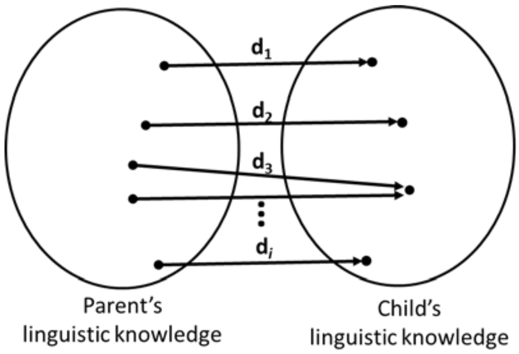

Fig1. Image of measuring procedure

Here, we give a concrete example. Assume that a parent agent has the linguistic knowledge including two rules as below.

rule1: *S/love(pete, mary)* → acd
rule2: *S/hate(mary, john)* → bd

Also, a child agent has

rule3: *love(john, mary)* → abc
rule4: *hate(mary, pete)* → bcd
rule5: *admire(mary, john)* → fcg

Now, pick up rule1 of parent's grammar rule, and search rules which have most similar PAS from the child's knowledge. In this case, rule3 is selected (procedure 1). Considering `acd' and `abc', the elimination of `d' and the succeeding insertion of `b' equate them, i.e., Levenshtein distance of them is 2 (procedure 2). Since the maximum distance is 6, the result of normalization becomes 2/6 (procedure 3). In the case of rule2, rule4 and rule5 are selected in procedure1. In terms of Levenshtein distance, rule4 is selected, and its distance is 1. After the normalization, we obtain 1/5. Therefore, the distance between the knowledge of the child agent and that of the parent agent becomes

$$\left\{\frac{1}{2} \times \left(\frac{1}{3} + \frac{1}{5}\right)\right\} \approx 0.27.$$

3. Experiment and Result

In this section, we show the result of the application of our suggested distance as mentioned in Section 2.2 to Kirby's ILM to testify its plausibility.

3.1 Briefing Experiment of Kirby's ILM

Kirby's ILM employs the following five two-place predicates and five object words:

Predicates: admire, detest, hate, love, like

Objects: gavin, heather, john, mary, pete

where two identical arguments in a predicate like love(john, john) is prohibited. This implies that there are 100 distinct meanings (5 predicates ×five first arguments × 4 second arguments).
Since the number of utterances is limited to 50 in his experiment, the child argent cannot learn the whole meaning space of 100; thus, the child agent comes across the learning bottleneck problem. To obtain the whole possible meanings, the child agent has to generalize his/her own linguistic knowledge by some learning process.

We have carried out this experiment until 100th generation.

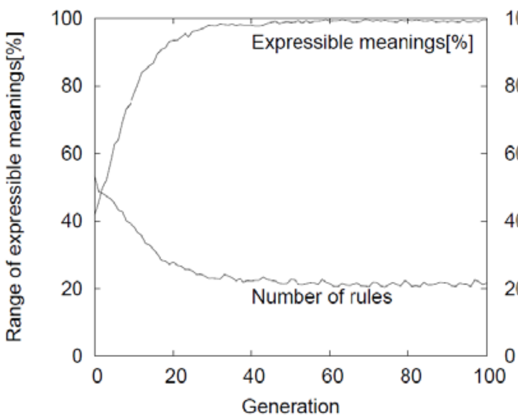

Fig2. Result of Kirby's ILM

The result of this experiment was that the language has low expressivity and a large number of rules in the early stages; however, through generations, the language acquires higher expressivity and the number of rules decreases. We can observe that the more the agent's linguistic knowledge becomes compositional the smaller the number of rules while expressivity persists or improves.

3.2 Application of Distance to Kirby's ILM

In the original experiment, the accomplishment of the learning is evaluated by the number of rules in linguistic knowledge and by the expressivity which is the ratio of the utterable meanings to the whole meaning space. Instead of these two criteria, we employ our own, i.e., the distance between two linguistic knowledge.

Here, we need to distinguish the distance from a parent to his/her offsprings, and that from an infant to his/her ancestors, each of which is shown in Fig. 3 and Fig. 4, respectively.

In Fig. 3, each upward line shows a distance of linguistic knowledge owned by a parent agent to his/her offsprings; the topmost line, for example, shows the distance from the parent agent in the first generation to his/her offsprings in the second, third, ..., and 100th generations. Because the distance becomes larger as the generations go further, the line becomes upward.

In this situation, for example, distance between knowledge of the 2nd generation to the 3rd generation is about 0.39. Also, distance between the 2nd generation to converged generation is about 0.85.

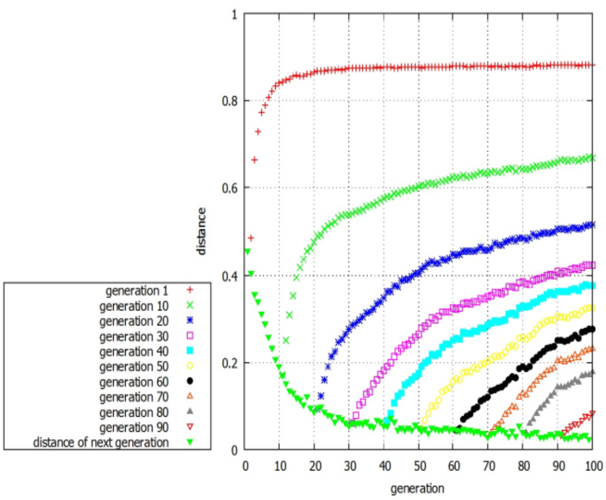

Fig3. distance from ancestor to offspring

In the later stages, linguistic knowledge of the parent agent becomes enough compositional to express the whole meaning world with few grammar rules, then the distance becomes smaller. For example, the distance between the 82nd generation to the 83rd generation is about 0.04. Also, the distance between the 82nd generation to the converged generation is about 0.14. The bottom line in Fig. 3 shows the distance between a generation and its

The Eighteenth International Symposium on Artificial Life and Robotics 2013 (AROB 18th '13),
Daejeon Convention Center, Daejeon, Korea, January 30-February 1, 2013

immediate follwing generation. We can observe that gradient of early generations is steep though that of the latter generations is flat.

In the early stages, the parent agent does not own compositional linguistic knowledge, i.e., expressivity of the knowledge is low, so the parent agent cannot make utterances with his/her own grammar rules, viz., the parent agent has no choice to make utterances randomly. On the other hand, in the latter stages, the parent acquires enough compositional linguistic knowledge, so the child agent can receive inputs in a regular pattern from the parent agent. Thus, the child agent can acquire similar linguistic knowledge to the parent agent, i.e., the distance of their knowledge is small.

The same tendency can be observed in Fig. 4. The topmost line in Fig. 4 shows the distance from the infant in the 100th generation to his/her ancestors. The further his/her ancestor is the larger the distance becomes, so that each line decreases from left to right. In the similar way, the bottom line in Fig.4 shows the distance between two consecutive generations; this line shows the exactly same tendency in Fig. 3.

5. Conclusion and Future Work

The cognitive biases are known to work to reduce the number of possible interpretations for each word, for first language acquisition. The authors thus far have verified that the function of the cognitive biases work not only in the lexical acquisition but also in the grammar acquisition. For this mission, we have revised Kirby's ILM, and built the cognitive biases into our model.

To evaluate the efficacy of the cognitive biases, however, we reconsidered the two traditional criteria; the expressivity and the number of rules. Although these two criteria are advantageous in evaluating the

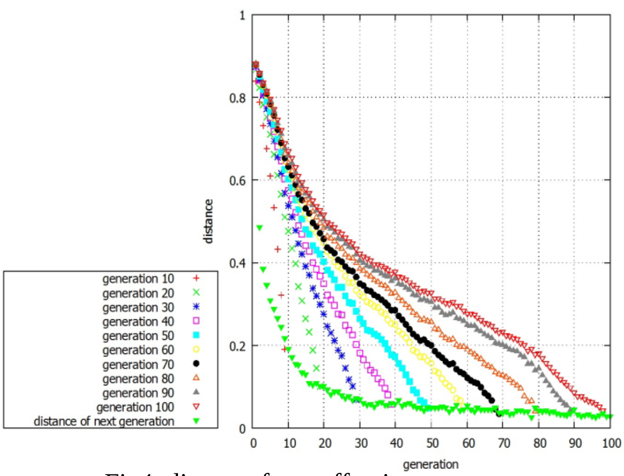

Fig4. distance from offspring to ancestor

acquisition speed of compositional language, they cannot evaluate the correctness of meaning; infants in each generation may still misunderstand the mapping between meanings and utterances in such rapid leanings. Therefore, we suggested a criterion to measure the distance between two linguistic knowledge using Levenshtein distance. As a result, the distances between two linguistic knowledge in the early stages are larger because their grammar are not compositional. On the other hand, after grammar becomes enough compositional, the distances become smaller. Our future works include to employ this measuring procedure to our grammar acquisition model and to verify the function of the cognitive biases.

Refferences

[1] G. Schafer and K. Plunkett. Rapid word learning by 15-month-olds under tightly controlled conditions. Child Development, Vol. 69, pp. 309{320, 1998.

[2] 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, pp. 45~75, 1994.

[3] E. M. Markman. Categorization and naming in children: Problems of induction, MIT Press, Cambridge, 1989.

[4] W.V.O.Quine, Word and Object, MIT Press, Cambridge, MA, 1960.

[5] R.Matoba, M.Nakamura, and S.Tojo, Efficiency of the Symmerty Bias in Grammar Acquisition, Information and Computation, Vol209, Issue 3, Elsevier, pp536-547, 2010.

[6] R.Matoba, M.Nakamura, S.Hagiwara, and S.Tojo, Efficiency of the Cognitive Bias in Grammara Acquisition, CogSci, 2012.

[7] S.Kirby, Learning Bottlenecks and the Evolution of Recursive Syntax, in: T.Briscoe(Ed.), Linguistic Evolution through Language Acquisition, CambridgeUniversity Press, 2002.