Title	Efficiency of the Symmetry Bias in Grammar Aquisition
Author(s)	Matoba, Ryuichi; Nakamura, Makoto; Tojo, Satoshi
Citation	Information and Computation, 209(3): 536-547
Issue Date	2010
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/9823
Rights	NOTICE: This is the author's version of a work accepted for publication by Elsevier. Ryuichi Matoba, Makoto Nakamura, Satoshi Tojo, Information and Computation, 209(3), 2010, 536-547, http://dx.doi.org/10.1016/j.ic.2010.11.018
Description	Special Issue: 3rd International Conference on Language and Automata Theory and Applications (LATA 2009)

Efficiency of the Symmetry Bias in Grammar Acquisition

Ryuichi Matoba[a], Makoto Nakamura[b], Satoshi Tojo[b]

[a] *Department of Electronics and Computer Engineering, Toyama National College of Technology, 1–2, Neriya, Ebie, Imizu, Toyama, 933–0293, Japan*
[b] *School of Information Science, Japan Advanced Institute of Science and Technology, 1–1, Asahidai, Nomi, Ishikawa, 923–1292, Japan*

Abstract

It is well known that the symmetry bias greatly accelerates vocabulary learning. In particular, the bias helps infants to connect objects with their names easily. However, grammar learning is another important aspect of language acquisition. In this study, we propose that the symmetry bias also helps to acquire grammar rules faster. We employ the Iterated Learning Model, and revise it to include the symmetry bias. The result of the simulations shows that infants could abduce the meanings from unrecognized utterances using the symmetry bias, and acquire compositional grammar from a reduced amount of learning data.

Keywords: symmetry bias, abduction, iterated learning model, compositionality, grammar acquisition

---

1. Introduction

Infants have an amazing capacity to learn their vocabularies. In fact, infants can acquire new words very rapidly, e.g., 7 to 15 words a day over the age of 18 months, and can also learn a word's meaning after just a single exposure, through fast mapping [1, 2, 3, 4, 5]. The difficulty in lexical acquisition includes mapping between a word and a meaning from an infinite

Email addresses: `rmatoba@nc-toyama.ac.jp` (Ryuichi Matoba), `mnakamur@jaist.ac.jp` (Makoto Nakamura), `tojo@jaist.ac.jp` (Satoshi Tojo)

range of possible meanings, as is well known as ‘gavagai problem’ [6]. Some cognitive biases are working for lexical acquisition. Infants overcome the difficult mapping through fast mapping and slow mapping. The process of fast mapping forms an approximate representation of a word’s meaning after just a single exposure on the basis of heuristics [7], or cognitive biases such as whole object bias, mutual exclusivity bias, and symmetry bias. Carey [1] explained that the approximate word meanings were first established by fast mapping, and then were embodied as actual words in infants’ memory through slow mapping. Therefore, infants under 17 months old only learn lexical items slowly, but this knowledge is fragile and those lexical items are prone to being forgotten. On the other hand, infants over 18 months old come to acquire new words firmly, and lexical misapplication subsides [8].

Among these various biases, the *symmetry bias* is said to be saliently effective for lexical learning [9, 10, 11, 12, 13, 14]. The bias says, if infants are taught that an object P has a lexical label Q, then they apply the label Q to the object P. For example, an infant who has learned the name of an object, which is called `apple`, can not only answer the name of the object indicated correctly, but can also pick out an apple from a basket filled with other fruits. The latter action is considered as a human-specific skill, and many experiments have shown that other animals cannot map an object to its label [15, 16, 17]. Although the relation between symmetry bias and lexical acquisition has been reported so far, its relation to grammatical construction has not been mentioned yet. In this paper, we verify the efficacy of the symmetry bias not only in lexical acquisition, but also in grammar acquisition.

Our study is based on the Iterated Learning Model (ILM, hereafter) by Kirby [18]; in each generation, an infant can acquire grammar in his mind given sample sentences from his mother, and a grown infant becomes the next mother to speak to a new-born 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 over multiple generations, not on that in one generation. In ILM’s setting, a sentence is considered to be uttered in an actual situation; thus, an infant can guess what a sentence implies. Therefore, a mother’s utterance is always paired with its meaning. Here, the utterance is regarded as an instance of E–language, while the infant guesses its grammar structure, that is, I–language [19, 20, 21, 22]. Our objective in this paper is to embed the symmetry bias into this combination of utterance and meaning, to show its efficacy.

This paper is organized as follows: in Section 2 we explain Kirby's ILM [18], and in Section 3 we revise it to include the symmetry bias. Section 4 presents the details of our experimental model, and gives specific experiment designs. We analyze our experimental results in Section 5, and conclude and discuss our results in Section 6.

2. Iterated Learning Model

ILM is a framework for investigating the cultural evolution of linguistic structure. In the experiment by Kirby [18], a parent is a speaker agent and her infant is a learner agent. The key issue of this model is the learning bottleneck, that is, the infant acquires sufficiently versatile grammar in spite of the limited amount of sentence examples from the parent; in other words, this bottleneck is the very cause of grammar generalization. The infant tries to guess sentence structure, as utterances are always paired with their meanings, which are intrinsically compositional. This process is iterated generation by generation, and finally in a certain generation an agent would acquire a compact, limited set of grammar rules.

2.1. Utterance Rule of Kirby's Model

According to Kirby's model, we present a signal–meaning pair as follows.

$$S/hit(john, ball) \rightarrow \text{hjsbs} \tag{1}$$

where the meaning, that is the speaker's intention, is represented by a Predicate–Argument Structure (PAS, hereafter) $hit(john, ball)$ and the signal is the utterance `hjsbs`; the symbol 'S' stands for the category Sentence. The following rules can also generate the same utterance.

$$S/hit(x, ball) \quad \rightarrow \quad \text{h}\, N/x\, \text{sbs} \tag{2}$$

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

The variable x in (2) can be substituted for an arbitrary element of category N. A rule without variables, i.e., the whole signal indicates the whole meaning of a sentence as in Formula (1) is called a *holistic rule* (Fig. 1), while a rule with variables as in Formula (2) is called a *compositional rule* (Fig. 2). Also, a rule which corresponds to a word, as in Formula (3) is called a *lexical rule.* Here, we review the definitions of these rules.

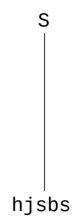

Figure 1: Holistic rule

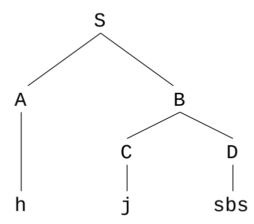

Figure 2: Compositional rule

Definition 1 (Compositional rule, Holistic rule, and Lexical rule).

Compositional rule: a grammar rule including non-terminal symbols for categories.
Holistic rule: a grammar rule consisting of terminal constants.
Lexical rule: a rule consisting of a monadic terminal constant.

In addition, we define the *variable level*.

Definition 2 (Variable level). *Variable level is the number of variables included in a grammar rule.*

For example, the level of Formula (1) is 0, while that of Formula (2) is 1.

2.2. Rule Subsumption

The learner agent has the ability to change his knowledge with learning. The learning algorithm consists of the following three operations; *chunk*, *merge*, and *replace* [18].

2.2.1. Chunk

This operation takes pairs of rules and looks for the most–specific generalization. For example,

$$\left\{ \begin{array}{l} S/read(john, book) \rightarrow \texttt{ivnre} \\ S/read(mary, book) \rightarrow \texttt{ivnho} \end{array} \right. \Rightarrow \left\{ \begin{array}{l} S/read(x, book) \rightarrow \texttt{ivn}N/x \\ N/john \rightarrow \texttt{re} \\ N/mary \rightarrow \texttt{ho} \end{array} \right.$$

2.2.2. Merge

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

$$\left\{\begin{array}{l} S/\mathit{read}(x, \mathit{book}) \rightarrow \texttt{ivn}A/x \\ A/\mathit{john} \rightarrow \texttt{re} \\ A/\mathit{mary} \rightarrow \texttt{ho} \\ S/\mathit{eat}(x, \mathit{apple}) \rightarrow \texttt{apr}B/x \\ B/\mathit{john} \rightarrow \texttt{re} \\ B/\mathit{pete} \rightarrow \texttt{wqi} \end{array}\right. \Rightarrow \left\{\begin{array}{l} S/\mathit{read}(x, \mathit{book}) \rightarrow \texttt{ivn}A/x \\ A/\mathit{john} \rightarrow \texttt{re} \\ A/\mathit{mary} \rightarrow \texttt{ho} \\ S/\mathit{eat}(x, \mathit{apple}) \rightarrow \texttt{apr}A/x \\ A/\mathit{pete} \rightarrow \texttt{wqi} \end{array}\right.$$

2.2.3. Replace

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

$$\left\{\begin{array}{l} S/\mathit{read}(\mathit{pete}, \mathit{book}) \rightarrow \texttt{ivnwqi} \\ B/\mathit{pete} \rightarrow \texttt{wqi} \end{array}\right. \Rightarrow \left\{\begin{array}{l} S/\mathit{read}(x, \mathit{book}) \rightarrow \texttt{ivn}B/x \\ B/\mathit{pete} \rightarrow \texttt{wqi} \end{array}\right.$$

3. Symmetry Bias Model

In this section, we introduce our modified Kirby model, including the symmetry bias as a kind of abductive reasoning, which is the reversely-directed implication. Especially, we reinterpret the notion of the symmetry bias to be built into the meaning–signal model.

3.1. Reverse Reasoning

If a learner is given an utterance always paired with its meaning (PAS), the learner can understand the speaker's intention correctly at all times. However in actual situations, this seldom occurs. In our model, we loosen this assumption and regard that some utterances lack meanings, corresponding to the following situations.

- A learner cannot understand the intention of a speaker's utterance.
- The learner fails to communicate with the speaker because of a lack of other modalities, like finger pointing.

Even in such cases, the learner attempts to complement the speaker's intention by using his own previously-acquired knowledge. We define this notion as the *symmetry bias*. For example, if the learner cannot get the meaning of the left-hand side of '→' in

$$S/p(a,b) \rightarrow \texttt{fjaljla},$$

the learner guesses its meaning backward, namely:

$$??? \leftarrow \texttt{fjaljla}.$$

This backward directional guess is regarded as the effect of the symmetry bias, and we build this process into our model. In ordinary circumstances, this reverse reasoning can lead to a mistake, though there is also the possibility to accelerate language acquisition.

3.2. ILM with the Symmetry Bias

Here we include the symmetry bias into ILM. The symmetry bias supposedly works well when the meaning and the signal are mapped one–to–one. It has been suggested that the Contrast Principle [23] favors one–to–one mappings between meanings and words in language acquisition. Otherwise, linguistic communicability would deteriorate and an infant could hardly acquire language, as was mentioned with a computer simulation by Smith [24].

Now we summarize our settings.

1. In the actual world, a learner cannot always understand a speaker's intention from an utterance.
2. However, the learner accepts this utterance as a reasonable linguistic representation, and he has an ability to understand it using his own knowledge.
3. An utterance is composed by the speaker's intention, and the utterance reflects the speaker's intention.
4. The process inferring the meaning from an utterance has the reverse direction of the utterance generation.
5. We regard this inference to be caused by the symmetry bias in language acquisition.
6. Because most meaning/utterance pairs correspond one-to-one, the symmetry bias works effectively.

For verifying the above intuitions, we have the following conjectures.

Conjecture 1 If a learner receives only an utterance, he looks for similar utterances that he has once received. If the utterance is new, he has to generate an appropriate meaning. In terms of computer simulation, this may increase computational time due to the addition of a learning process, compared with Kirby's model.

Conjecture 2 Despite the disadvantage of Conjecture 1, if the learner's partial grammatical/lexical knowledge is sufficient, he may be able to guess meanings of unrecognized utterances. Thus, the learning process saves memory space.

In the next section, we show the adequacy of these conjectures by computer simulation.

4. Experiments in Symmetry Bias Model

In this section, we show the procedure and the result of our experiment. The purpose of the experiment is to demonstrate acquisition of compositional grammar, even in a case that a learner agent may not always understand the meaning of an utterance. In order to examine the efficacy of our model, we compare our symmetry bias model to the original model.

First of all, we reiterate the experiment by Kirby [18] as a pilot one, to grasp the features of the original model; how many generations are needed to organize a compositional grammar, when does an agent acquire a grammar that can represent the whole meaning space, how many grammar rules does the agent acquire, and so on.

After the pilot experiment, we examine the following three strategies when a learner cannot understand the meaning of an utterance.

(I) The learner ignores such utterances, and does not use them in his learning.

(II) The learner complements meanings of such utterances, randomly assigning his previously learned meanings.

(III) The learner applies the symmetry bias to such utterances to complement their meanings, and uses them in his learning.

The purpose of experiment (I) is to observe differences in acquired grammar, dependent on the amount of data for learning, by comparison with the pilot experiment. Also, we compare experiment (I) to experiment (III) to observe the effect of complementary process. Next, we compare experiment (II) to experiment (III) to observe the superiority of the symmetry bias to use of simple meaning complementation.

4.1. Experimentation Environment

In our model, we have employed the following five predicates and five object words, which are the same in Kirby's experiment [18].

predicates:	admire, detest, hate, like, love
objects:	gavin, heather, john, mary, pete

Preserving Kirby's settings, we prohibit two identical arguments in a predicate like $love(pete, pete)$. This implies that there are 100 distinct meanings (5 predicates $\times$ 5 possible first arguments $\times$ 4 possible second arguments). Algorithm 1 is the procedure of the simulation. Algorithm 2, appearing in Algorithm 1, will be explained in Section 4.4.

Since the number of utterances is limited to 50, the learner cannot learn the whole meaning space, the size of which is 100; thus, the learner comes across Kirby's bottleneck. To obtain the whole meaning space, the learner has to generalize his own knowledge by self-learning.

To evaluate the accomplishment of the learning, we investigate *expressivity* in the following definition, as well as the number of grammar rules.

Definition 3 (Expressivity). *Expressivity is the ratio of the utterable meanings to the whole meaning space.*

We regard that grammar has converged under the following condition.

Definition 4 (Grammar convergence). *We regard that the grammar has converged when the average expressivity of 100 trials attains to 95%.*

The generation when the grammar converges is called the *convergence generation.*

We have carried out each experiment until the 200th generation. However, the tendency of expressivity and the number of rules fixes by the 100th generation, and thus we discuss the results by the 100th generation hereafter.

Prior to our experiments, we have simulated Kirby's experiment. In the early stages, the language has low expressivity and a large number of grammar rules; however, through generations, the language overall acquires higher expressivity and the number of grammar rules decreases. First, the increase

```
generation := 0;
repeat
    a speaker := a parent;
    a learner := an infant;
    for 1 to 50 do
        the speaker chooses a meaning (PAS) from the whole meaning
        space;
        the speaker generates an utterance by her own grammar rules;
        if the speaker could not generate one then
            attach random string to the meaning (PAS);
            put the rule into her grammar set;
        end
        if the learner receives an utterance without its meaning then
            switch Experiment do
                case Experiment (I)
                    the learner does not accept the utterance;
                end
                case Experiment (II)
                    the learner chooses a meaning arbitrarily, and
                    accepts the signal–chosen meaning pair holistically;
                end
                case Experiment (III)
                    the learner follows Algorithm 2 to guess the meaning
                    , and accepts the signal–meaning pair holistically;
                end
            end
        else
            the learner accepts the signal–meaning pair holistically;
        end
        the learner executes learning process, i.e., chunk, merge,
        replace;
    end
    discard the speaker, together with her grammar;
    a parent := the learner; /*the infant becomes a new parent*/
    an infant is created;
    generation ++;
until 200-th generation ;
```

Algorithm 1: The simulation algorithm

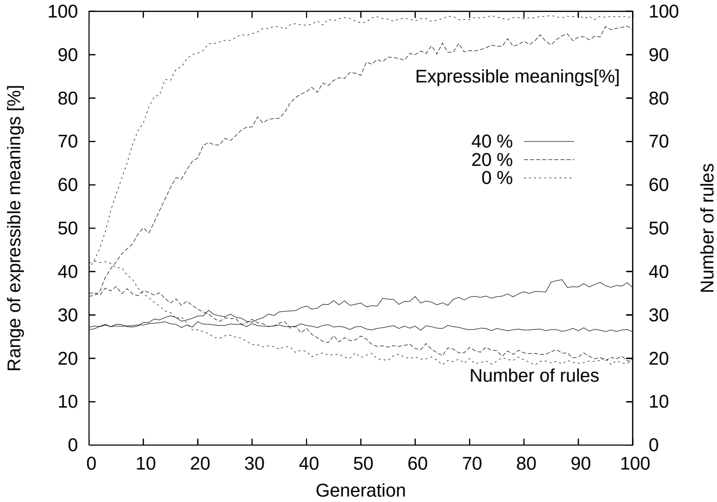

Figure 3: The trends of the number of rules and expressivity per generation: experiment (I)

of holistic rules leads to the growth of expressivity. Then, the grammar becomes compositional in the later stages. We can observe that the more the grammar becomes compositional the smaller the number of rules while expressivity persists or improves. In fact, we have found a grammar with 100 % expressivity consisting of 11 rules, among which are ten lexical rules and one variable level–3 compositional rule (see Definition 2).

4.2. Experiment (I): learner ignores unrecognized utterances

In this experiment, the learner ignores unrecognized utterances and misses that amount of data for learning, i.e., the setting is the same as the pilot experiment, except that the number of meaningful utterances is reduced.

Fig. 3 shows the average tendency of the number of rules and expressivity per each generation, after 100 trials. Multiple lines denote the difference of the ratio of unrecognized utterances; we set the rate 0 %, 20 %, and 40 % of the 50 received utterances. Actually, we have experimented with unrecognized rates beyond 40 %, but grammar never converged, staying near to the initial expressivity. Therefore, we will not mention those futile results hereafter.

At 0 % unrecognized rate, the grammar converges, i.e., the expressivity goes beyond 95 % (Definition 4) at the 32nd generation with 23.14 rules, and at the 100th generation the expressivity is 98.79 % with 18.76 rules on an average. As the unrecognized rate increases, the convergence generation is retarded; at 20 % the grammar converges at the 98th generation and at 40 % the expressivity remains 37.44 % even at the 100th generation. On the contrary, the higher the recognized rate is, the earlier the convergence generation becomes.

4.3. Experiment (II): learner arbitrarily interprets unrecognized utterances

In this experiment, when the learner cannot recognize the meaning of an utterance, he chooses a meaning (PAS) from the meaning space arbitrarily and combines it to the utterance.

A certain rate of unrecognizable utterances are received by the learner agent after recognizable ones; e.g., 10 meaningless sentences are heard after 40 meaningful utterances in case of 20 % rate. That is, the learner first learns firm knowledge of the utterances with meanings, and after then guesses meanings for missing ones.

Similarly to experiment (I), Fig. 4 shows the average tendency of the number of rules and expressivity per each generation after 100 trials, and each line corresponds to the ratio of unrecognized utterances among the whole input as 0 %, 20 %, and 40 %.

Note that there is no difference between experiments (I) and (II) at 0 % rate. In experiment (II), at 20 %, the grammar converges at the 29th generation, and at the 100th generation the expressivity is 98.94 % with 34.49 rules. At 40 %, the grammar converges at the 72nd generation, and at the 100th, 96.24 % expressivity with 40.47 rules.

We can observe that the grammar converges faster at 0 % (experiment (I)) while the expressivity is rather higher in the early generations at 20 %. As learners complement the missed meanings and augment the grammar knowledge in experiment (II), the expressivity tentatively increases; but in the later generation these arbitrary knowledge seems to hinder the grammar convergence.

4.4. Experiment (III): learner applies the symmetry bias to interpret unrecognized utterances

In this strategy, the learner agent follows Algorithm 2. First, he searches grammar rules which exactly match the utterance, and if he can find one he can interpret it. Or else, i.e., when the utterance is unrecognized, he tries to compose the utterance from those rules of variable level one or more in his knowledge; if it is successful, he attaches the composed meaning to the

if *there is a string among variable level 0 (holistic) rules which exactly match the utterance* **then**
    apply the meaning of this rule;
else
    if *the learner can derive the same string as the utterance using rules whose variable levels are greater than or equal to 1* **then**
        apply the meaning derived from these rules;
    else
        Decompose the string to substrings;
        Search lexical rules which match the longest substrings of the utterance;
        if *there exist such lexical rules* **then**
            The learner chooses utterances which include any one of such words;
            From the chosen utterances, the learner picks out as candidates those which are uttered most frequently;
            if *the candidate is only one* **then**
                apply the meaning of this candidate to the unrecognized utterance;
            else
                Among the other substrings, search lexical rules which correspond to one of the longest substrings;
                Choose the most frequent candidates;
                if *such candidate is only one* **then**
                    apply the meaning of this candidate to the unrecognized utterance;
                else
                    Choose one from the candidates;
                end
            end
        else
            Choose a meaning from the meaning space randomly, and apply it to the recognized utterance.
        end
    end
end

Algorithm 2: The bias algorithm

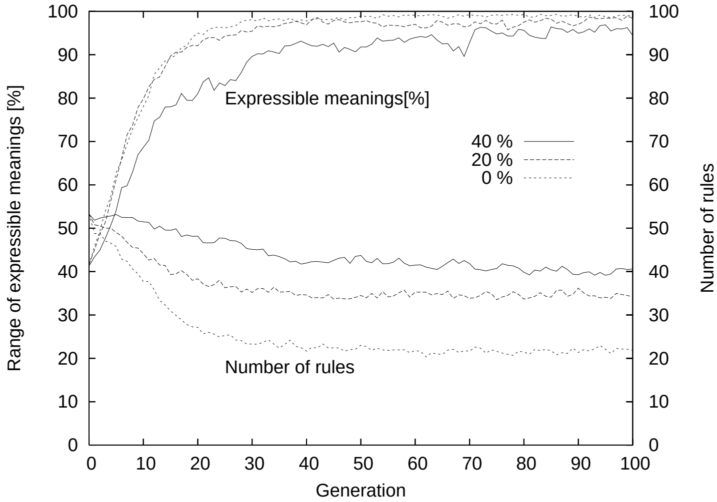

Figure 4: The trends of the number of rules and expressivity per generation: experiment (II)

utterance. If it still fails, he decomposes the utterance to substrings and looks for a longest matching rule; when multiple rules hit at the same length, he applies the one which has been most frequently evoked. When all the above trials fail, he randomly assigns a meaning (PAS).

Similarly to the previous experiments, the ratios of unrecognized utterances are set to 0 %, 20 %, and 40 % as in Fig. 5. Also, unrecognizable sentences are given at the tail part of the 50 utterances.

In experiment (III), at 20 %, the grammar converges at the 17th generation with 22.59 rules, and at the 100th generation the expressivity is 98.74 % with 21.42 rules. At 40 %, the grammar converges at the 27th generation and at the 100th generation the expressivity is 95.68 % with 23.32 rules.

There appeared insignificant difference between 0 % and 20 %, both in expressivity and the number of rules after convergence generation. However, in early stages, 20 % earns larger expressivity; the learner supposedly acquires compositional grammar beyond the limited mother grammar.

By the way, we considered the case that unrecognizable sentences are disseminated randomly among recognizable ones. At 20 %, the grammar converges at the 28th generation with 30.02 rules, and at the 100th generation the expressivity is 97.92 % with 30.91 rules. At 40 %, the grammar did not

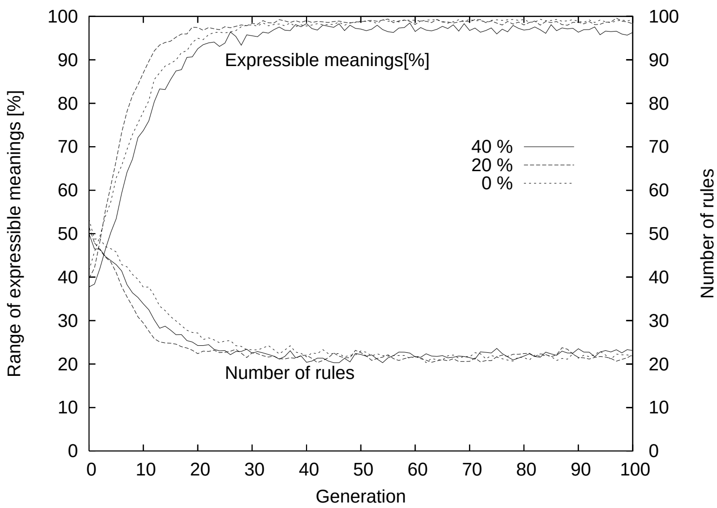

Figure 5: The trends of the number of rules and expressivity per generation: experiment (III)

converge even at the 100th when 79.82 % expressivity. That is, this randomization deteriorates grammar convergence, as was expected. The tendency is shown in Fig. 6.

5. Efficacy of the Symmetry Bias

5.1. Comparison of Experimental Results

In this section, we compare the result of experiment (I) to that of experiment (III), and the result of experiment (II) to that of experiment (III).

Firstly, we compare experiment (I) to experiment (III). The learner in experiment (I) can acquire rather smaller amount of learning data as unrecognized utterances are only discarded. On the contrary, as the learner agent in experiment (III) tries to guess missing meanings by the symmetry bias, the learning data are complemented. Therefore, we can expect that the expressivity of experiment (III) converges earlier than experiment (I). However, those guessed rules are not necessarily correct; namely, the learner agent may acquire rules which were originally not included in the speaker's grammar. In the later generation, these inadequate rules, which are rather holistic than compositional, tend to remain; this is why Table 1 shows that the number of rules in experiment (I) is larger than that in experiment (III).

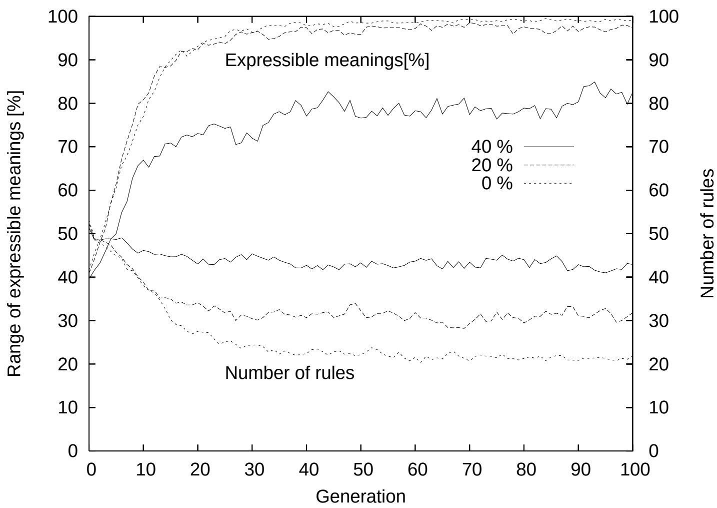

Figure 6: Random unrecognized utterances in experiment (III)

Next, we compare experiment (II) to experiment (III). Experiment (III) surpasses (I) and (II) both in expressivity and convergence generation as is shown in Fig. 7 at 40 % ratio of unrecognized utterances where the difference is prominent as in Table 1. Thus, we can find that the symmetry bias works saliently.

Table 1: Average number of rules at the 100th generation

	experiment (I)	experiment (II)	experiment (III)
0 %	18.76		
20 %	37.90	34.49	21.42
40 %	54.35	40.47	23.32

5.2. Example of Efficacy of the Symmetry Bias

Here, we show how the symmetry bias works to construct compositional grammars, using a concrete example from experiment (III) where the ratio of unrecognized utterances is 20 %.

Now, let us look into the learner's grammatical knowledge in the 3rd generation, to trace how the learner infers meanings of unrecognized utterances.

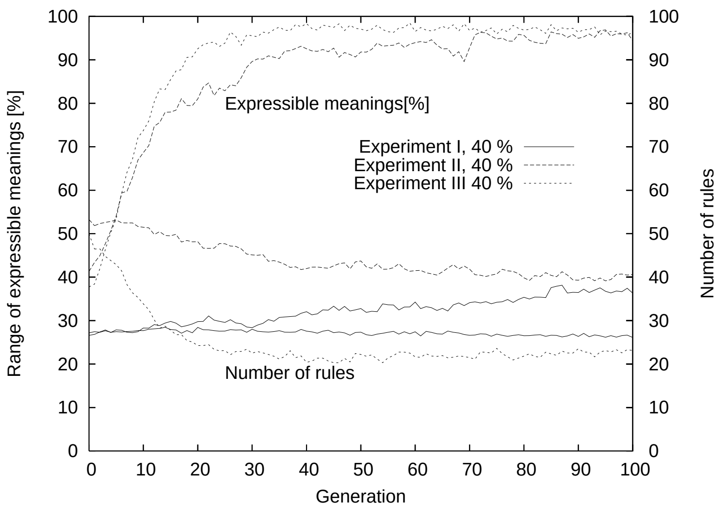

Figure 7: The trends of experiments (I), (II), versus (III)

Table 2 shows the grammar of the learner in the 3rd generation after the 47th utterance.

The speaker utters $admire(mary, heather) \rightarrow$ `ljd` as the 48th utterance, and the learner receives this utterance without meaning, i.e., only `ljd`. At this stage shown in Table 2, the learner does not have a holistic rule for `ljd`, nor substrings corresponding to `lj`, `jd`, while he does have rules for `j`, such as r_{32} and r_{43}. According to Algorithm 2, the learner applies a lexical rule in this case, so r_{43} is employed:

$$C/admire \rightarrow \texttt{j}.$$

Therefore, the learner infers as follows:

$$admire(???, ???) \leftarrow \texttt{ljd}.$$

Looking up words which co-occur most often with the word *admire* from the speaker's utterances up to this time, the learner finds $(john, pete)$ has occurred twice, while $(john, gavin)$, $(pete, john)$, $(gavin, pete)$, $(john, mary)$, $(mary, gavin)$ have each occurred once. Thus, the learner chooses $(john, pete)$, and adds the following rule to his knowledge:

$$admire(john, pete) \rightarrow \texttt{ljd}.$$

Table 2: Learner's acquired knowledge after 47th input(Generation 3)

Generation 3				Speaker's 48th utterance: $admire(mary, heather) \rightarrow$ `ljd`			
r_1 :	$S/admire(gavin, john)$	$\rightarrow$	`mm`	r_{23} :	$S/hate(pete, gavin)$	$\rightarrow$	`w`
r_2 :	$S/admire(gavin, pete)$	$\rightarrow$	`t`	r_{24} :	$S/like(gavin, heather)$	$\rightarrow$	`nc`
r_3 :	$S/admire(john, gavin)$	$\rightarrow$	`d`	r_{25} :	$S/like(gavin, john)$	$\rightarrow$	`bf`
r_4 :	$S/admire(john, mary)$	$\rightarrow$	`v`	r_{26} :	$S/like(heather, y)$	$\rightarrow$	$\mathtt{q}B/y$
r_5 :	$S/admire(john, pete)$	$\rightarrow$	`n`	r_{27} :	$S/like(john, mary)$	$\rightarrow$	`nts`
r_6 :	$S/admire(mary, gavin)$	$\rightarrow$	`mkh`	r_{28} :	$S/like(mary, gavin)$	$\rightarrow$	`dwa`
r_7 :	$S/admire(pete, john)$	$\rightarrow$	`fgf`	r_{29} :	$S/like(pete, gavin)$	$\rightarrow$	`cgb`
r_8 :	$S/detest(heather, gavin)$	$\rightarrow$	`fl`	r_{30} :	$S/like(pete, heather)$	$\rightarrow$	`gcm`
r_9 :	$S/detest(heather, john)$	$\rightarrow$	`o`	r_{31} :	$S/like(pete, john)$	$\rightarrow$	`ze`
r_{10} :	$S/detest(heather, pete)$	$\rightarrow$	`diiw`	r_{32} :	$S/love(mary, heather)$	$\rightarrow$	`j`
r_{11} :	$S/detest(john, pete)$	$\rightarrow$	`nj`	r_{33} :	$S/love(pete, heather)$	$\rightarrow$	`u`
r_{12} :	$S/detest(mary, john)$	$\rightarrow$	`eru`	r_{34} :	$S/love(x, gavin)$	$\rightarrow$	$A/x\mathtt{d}$
r_{13} :	$S/detest(mary, john)$	$\rightarrow$	`ziz`	r_{35} :	$S/p(john, mary)$	$\rightarrow$	$\mathtt{im}C/p$
r_{14} :	$S/detest(pete, mary)$	$\rightarrow$	`kr`	r_{36} :	$S/p(john, pete)$	$\rightarrow$	$C/p\mathtt{i}$
r_{15} :	$S/hate(gavin, heather)$	$\rightarrow$	`bd`	r_{37} :	$S/p(john, pete)$	$\rightarrow$	$\mathtt{g}C/p\mathtt{c}$
r_{16} :	$S/hate(gavin, mary)$	$\rightarrow$	`cmf`	r_{38} :	$S/p(pete, john)$	$\rightarrow$	C/p
r_{17} :	$S/hate(gavin, mary)$	$\rightarrow$	`tru`	r_{39} :	$A/john$	$\rightarrow$	`zc`
r_{18} :	$S/hate(gavin, pete)$	$\rightarrow$	`ftc`	r_{40} :	$A/mary$	$\rightarrow$	`ngb`
r_{19} :	$S/hate(heather, gavin)$	$\rightarrow$	`iri`	r_{41} :	$B/gavin$	$\rightarrow$	`zh`
r_{20} :	$S/hate(john, gavin)$	$\rightarrow$	`k`	r_{42} :	$B/john$	$\rightarrow$	`tw`
r_{21} :	$S/hate(john, heather)$	$\rightarrow$	`wg`	r_{43} :	$C/admire$	$\rightarrow$	`j`
r_{22} :	$S/hate(john, mary)$	$\rightarrow$	`do`	r_{44} :	$C/detest$	$\rightarrow$	`c`

After finishing the complementary process with the symmetry bias, the learner agent executes learning process. Adding a new rule enables the learner to generalize his grammar by operation *replace*. As a result, the learner acquires a new compositional rule, changing his grammar as follows:

$$\begin{cases} C/admire \rightarrow \mathtt{j} \\ S/p(john, pete) \rightarrow \mathtt{l}\ C/p\ \mathtt{d}. \end{cases}$$

In general, the higher the variable level of added rules is, the higher expressivity the language has. Therefore, comparing experiment (I) to experiment (III), the expressivity of the latter converges in a shorter period.

Table 3 shows the result of the guessing of meanings of unrecognized utterances by inference in the 3rd generation. Each item shows the number of each utterance, utterance itself, the meaning and the way the learner guessed the meaning. From Table 3, out of ten unrecognized utterances, meanings of four utterances are guessed by the inference, and the rest of them are guessed randomly because of failure of inference.

Fig. 8 shows the number of rules which are derived from inference with the symmetry bias, and from random application over generations. We set the ratio of unrecognized meanings to 20%. In the early stages, because of failures of inference, the learner has applied meanings to unrecognized

Table 3: Result of learner's inference by symmetry bias (Generation 3)

Num	Speaker's Utterance	Meaning (by Learner's guess)	Application
41	$hate(john, pete) \rightarrow$ `gjc`	$admire(john, pete)$	$inference$
42	$detest(gavin, pete) \rightarrow$ `nts`	$like(john, mary)$	$random$
43	$detest(mary, heather) \rightarrow$ `eru`	$detest(mary, john)$	$random$
44	$detest(mary, gavin) \rightarrow$ `mm`	$admire(gavin, john)$	$random$
45	$hate(heather, john) \rightarrow$ `fl`	$detest(heather, gavin)$	$random$
46	$like(gavin, heather) \rightarrow$ `w`	$hate(pete, gavin)$	$random$
47	$detest(john, heather) \rightarrow$ `do`	$hate(john, mary)$	$inference$
48	$admire(mary, heather) \rightarrow$ `ljd`	$admire(john, pete)$	$inference$
49	$like(john, mary) \rightarrow$ `ru`	$love(gavin, mary)$	$random$
50	$hate(heather, mary) \rightarrow$ `js`	$admire(john, pete)$	$inference$

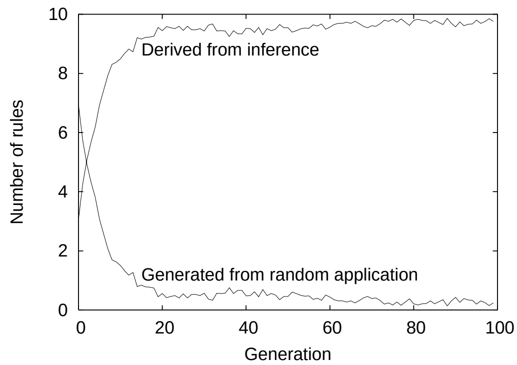

Figure 8: The number of rules which are derived from inference vs. random application

utterances randomly. In later generations, the learner comes to be able to apply meanings by inference. From Fig. 8, we can observe that the more generations go by, the more the number of rules with random application decreases. Also, from Fig. 5, we can see that expressivity and the number of rules converge around the 20th generation. Thus, we can observe that inference failures had steadily decreased in the agents of later generations, the learner having acquired a certain level of grammar knowledge.

If some rules are added to grammar knowledge by the symmetry bias, compositional rules are inevitably generated. However, in the case of random meaning application, it is not always true that learning produces compositional rules.

6. Conclusion

In this paper, we verified the efficacy of the symmetry bias not only in lexical acquisition, but also in grammar acquisition. For this purpose, we have modified Kirby's model [18], and have built the symmetry bias into our grammar conversion model. In the simulation, the learner received utterances without the speaker's intention at a certain rate, and the learner interpreted such unrecognized utterances by three kinds of strategies such as (I) ignoring, (II) attaching meanings randomly, and (III) attaching meanings using the symmetry bias. For each of the strategies, we have observed expressivity, number of rules, and variable level. As a result of the experiments, the learner who has employed strategy (III),

- could acquire high expressivity faster than those who used the strategy (I).
- could construct his grammar with a higher variable level, i.e., the number of rules was smaller than that of the learner who employed strategy (II).

Our model is designed for language evolution, and we have not proved the change of learning speed in one generation. However, we may contend that the learning speed was accelerated somehow by the bias because infants could acquire a certain grammar from 40 utterances while the original Kirby's experiment required 50 utterances, when 20 percent utterances lack for meanings, for example. Namely, infants could learn grammar with the smaller number of input.

According to the trace of inference process using the symmetry bias, as has been shown in Section 5.2, we observed that rules added by the symmetry bias lead to generation of compositional rules inevitably. Therefore, in terms of variable level, the symmetry bias works effectively.

Our future works are planned as follows. First, we should consider including other cognitive biases than the symmetry bias. Second, we should expand the meaning space and grammars of our model, and restudy its effectiveness, since the effect of the bias was not so obvious with the current simple and limited language. Third, we should investigate what happens if we attach incorrect PASs intentionally. If infants are much more versatile in learning language, they also may be able to learn a grammar robustly.

Acknowledgment

The authors would like to thank Shoki Sakamoto and Shingo Hagiwara of JAIST who helped to implement our experimental system.

Also, this work was partly supported by Grant-in-Aid for Young Scientists (B) (KAKENHI) No.20700239 from MEXT Japan.

References

[1] S. Carey, E. Bartlett, Acquiring a single new word, Child Language Development 15 (1978) 17–29.

[2] A. L. Woodward, E. M. Markman, Early word learning, in: W. Damon, D. Kuhn, R. Siegler (Eds.), Handbook of child psychology, volume 2: cognition, perception, and language, John Wiley and Sons, 1998, pp. 371–420.

[3] V. K. Jaswal, E. M. Markman, Learning proper and common names in inferential versus ostensive contexts, Child Development 72 (2001) 768–786.

[4] K. M. Wilkinson, K. Mazzitelli, The effect of "missing" information on children's retention of fast mapped labels, Child Language 30 (2003) 47–73.

[5] J. S. Horst, L. K. Samuelson, Fast mapping but poor retention by 24-month-old infants, Infancy 13 (2008) 128–157.

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

[7] R. A. Blythe, K. Smith, A. D. M. Smith, Learning times for large lexicons through cross-situational learning, Cognitive Science 34 (2010) 620–642.

[8] G. Schafer, K. Plunkett, Rapid word learning by 15-month-olds under tightly controlled conditions, Child Development 69 (1998) 309–320.

[9] M. Imai, D. Gentner, Children's theory of word meanings: The role of shape similarity in early acquisition, Cognitive Development 9 (1994) 45–75.

[10] M. Imai, D. Gentner, A crosslinguistic study of early word meaning: Universal ontology and linguistic influence, Cognition 62 (1997) 169–200.

[11] B. Landau, L. B. Smith, S. S. Jones, The importance of shape in early lexical learning, Cognitive Development 3 (1988) 299–321.

[12] B. Landau, L. B. Smith, S. S. Jones, Syntactic context and the shape bias in children's and adult's lexical learning, Journal of Memory and Language 31 (1992) 807–825.

[13] E. M. Markman, Constraints children place on word meanings, Cognitive Science: A Multidisciplinary Journal 14 (1990) 57–77.

[14] E. M. Markman, J. L. Wasow, M. B. Hansen, Use of the mutual exclusivity assumption by young word learners, Cognitive Psychology 47 (2003) 241–275.

[15] M. Hattori, Adaptive Heuristics of Covariation Detection : A Model of Causal Induction, in: Proceedings of the 4th International Conference on Cognitive Science and the 7th Australasian Society for Cognitive Science Joint Conference (ICCS/ASCS 2003), volume 1, pp. 163–168.

[16] M. Sidman, R. Rauzin, R. Lazar, S. Cunningham, W. Tailby, P. Carrigan, A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children, Journal of the Experimental Analysis of Behavior 37 (1982) 23–44.

[17] Y. Yamazaki, Logical and illogical behavior in animals, Japanese Psychological Research 46 (2004) 195–206.

[18] S. Kirby, Learning, bottlenecks and the evolution of recursive syntax, in: T. Briscoe (Ed.), Linguistic Evolution through Language Acquisition, Cambridge University Press, 2002.

[19] N. Chomsky, Knowledge of Language, Praeger, New York, 1986.

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

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

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

[23] E. Clark, On the logic of contrast, Journal of Child Language 15 (1988) 317–337.

[24] K. Smith, Natural selection and cultural selection in the evolution of communication, Adaptive Behavior 10 (2002) 25–45.