Efficiency of the Symmetry Bias
in Grammar Acquisition

Ryuichi Matoba, Makoto Nakamura, and Satoshi Tojo

School of Information Science,
Japan Advanced Institute of Science and Technology,
1–1, Asahidai, Nomi, Ishikawa, 923–1292, Japan
{r-matoba,mnakamur,tojo}@jaist.ac.jp

Abstract. It is well known that the symmetry bias much accelerates the 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 would help to acquire grammar rules faster. We employ Iterated Learning Model, and revise it to include the symmetry bias. The result of the experiments shows that infants could abduce the meanings from incomprehensible utterances using the symmetry bias, and acquire the compositional grammar from a reduced amount of learning data.

1 Introduction

It is well known that infants under 17 months old only slowly acquire lexical items, and these lexicons are hardly fixed. Also, infants in these ages tend to misapply words to objects. On the other hand, infants over 18 months old can acquire new words very rapidly, e.g. 7∼15 words a day, and such lexical misapplication subsides [1]. We cannot explain this phenomenal learning, considering that infants cannot take account of all possible meanings from an utterance and they learn meanings of words only from a few limited examples. For this problem, several studies have suggested that infants infer meanings efficiently to limit possibilities in a situation using constraints called cognitive biases [2,3]D A simple expression of lexical acquisition is a mapping between an object and a lexical label. This mapping is considered to be generally difficult as is well known as 'gavagai problem' [4], though infants achieve this operation.

It has been reported that various kinds of cognitive biases work for infants to limit the possible word meanings [2,5,6,7,8,9], and among which, the *symmetry bias* is said to be saliently effective. The bias says, if infants are taught an object P has a lexical label Q, then they apply the label Q to the object P without being taught. 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 [10,11,12].

Our study is based on the Iterated Learning Model (ILM, hereafter) by Kirby [13], who introduced the notion of *compositionality* and *recursion* as fundamental features of grammar, and showed that they made it possible for a human

to acquire compositional language without LAD (Language Acquisition Device), opposing to Chomsky's idea [14]. Also, he adopted the idea of two different domains of language [15,14,16,17], 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 [16]. In his model, he regarded that a parent is a speaker agent and 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, 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 the symmetry bias into this process. We implement agents with the bias in a virtual world, and make them learn a grammar by computer simulation.

This paper is organized as follows: in Section 2 we explain Kirby's ILM [13] and revise it to include the symmetry bias. Section 3 presents the details of our experimental model, and gives specific experiment designs. We analyze our experimental results in Section 4, and conclude and discuss our issues in Section 5.

2   Iterated Learning Model with the Symmetry Bias

2.1   Utterance Rule of Kirby's Model

According to Kirby's model, we show a pair of I–language and E-language as follows.

$$S/eat(john, apple) \rightarrow \text{eatjohnapple}$$

where a speaker's intention is a PAS '$eat(john, apple)$' and its utterance becomes 'eatjohnapple'; the symbol 'S' stands for that they belong to category Sentence. Thus, as far as a listener is given an utterance paired with its meaning (PAS), the listener can understand the speaker's intention precisely at all times. However, compared to the actual situation, it seems a very strong assumption. In our model, we loosen this assumption and regard that some utterances lack meanings, to show the efficacy of the cognitive bias.

2.2   Rule Subsumption

The listener agent has an ability to change his knowledge with learning. The learning algorithm consists of the following three operations such as *chunk*, *merge*, and *replace* [13].

Chunk. This operation takes pairs of rules and looks for the least–general generalization. For example,

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

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

$$\begin{cases} S/read(x, book) \rightarrow \text{ivn}A/x \\ A/john \rightarrow \text{re} \\ A/mary \rightarrow \text{ho} \\ S/eat(x, apple) \rightarrow \text{apr}B/x \\ B/john \rightarrow \text{re} \\ B/pete \rightarrow \text{wqi} \end{cases} \Rightarrow \begin{cases} S/read(x, book) \rightarrow \text{ivn}A/x \\ A/john \rightarrow \text{re} \\ A/mary \rightarrow \text{ho} \\ S/eat(x, apple) \rightarrow \text{apr}A/x \\ A/pete \rightarrow \text{wqi} \end{cases}$$

Replace. If a rule is embeddable in another rule, replace the latter for a compositional rule with variables.

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

2.3   Abductive Reasoning

An utterance lacking its meaning (PAS) corresponds to the following situations.

– A listener cannot understand what the intention of a speaker's utterance is.
– The listener fails to communicate with the speaker because of lacking other modalities, like finger pointing.

Even such cases, the listener 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 listener cannot get the meaning of the left-hand side of '$\rightarrow$' in

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

The listener guesses its meaning backward, namely:

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

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 abductive reasoning can be a mistake though there is a possibility to accelerate the language acquisition.

2.4   Extensional Model by the Symmetry Bias

The relation between an utterance and a meaning is not only one–to–one mapping, but also one–to–many or many–to–one. However, from the viewpoint of our linguistic performance, our language strongly favors one–to–one mapping. Thus, it can be reasonably expected that infants use the symmetric bias to obtain a meaning of the utterance if the infants possess a certain level of grammatical/lexical knowledge. Now, we summarize our intuition which we mentioned so far, as a policy of our study.

1. In the actual world, a listener cannot always understand a speaker's intention from an utterance.
2. However, the listener 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 is caused by the symmetry bias in language acquisition.
6. If most of meaning/utterance pairs are one-to-one, the symmetry bias must work effectively.

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

Conjecture 1. If a listener 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 learning process, compared with Kirby's model.

Conjecture 2. However, if the listener's partial grammatical/lexical knowledge is sufficient, he may be able to complement meanings of incomprehensible utterances. Thus, the learning process saves memory space.

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

3   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 listener 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 of Kirby's model as a pilot one, to grasp the features of the original model; how many generations are needed to organize a compositional grammar, when an agent acquires a grammar that can represent the whole meaning space, how many grammar rules the agent acquires, and so on.

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

(I) The listener ignores such utterances, and does not use them in his learning.
(II) The listener complements meanings of such utterances, randomly assigning his previous meanings.
(III) The listener 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 of acquired grammar, dependent on the data amount for learning, by comparison with the pilot experiment. Also, we compare experiment (I) to experiment (III) to observe the effect of complemental process. Next, we compare experiment (II) to experiment (III) to observe the superiority of the symmetry bias to the mere simple meaning complementation.

3.1   Experimentation Environment

In our model, we have employed the following five predicates and five object words, which are the same as Kirby's experiment [13].

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

The arguments of a predicate must not be identical, i.e., representations like $love(pete, pete)$ are prohibited. 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.

```
generation := 0;
repeat
    a speaker := a parent;
    a listener := 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
        give string–PAS to the listener;
        the listener guesses grammar rules in his own mind;
    end
    discard the speaker, together with her grammar;
    a parent := the listener; /*the infant becomes a new parent*/
    an infant is created;
    generation ++;
until the grammar converges ;
```

Algorithm 1. The simulation algorithm

Because the number of utterances is limited to 50 times, the listener cannot learn the whole meaning space, the size of which is 100. To obtain the whole meaning space, the listener has to generalize his own knowledge by learning.

To evaluate the accomplishment of the learning, we investigate *expressivity* and *compositional level* in the following definitions, as well as the number of grammar rules.

(A)

(B)

(C)

(D)
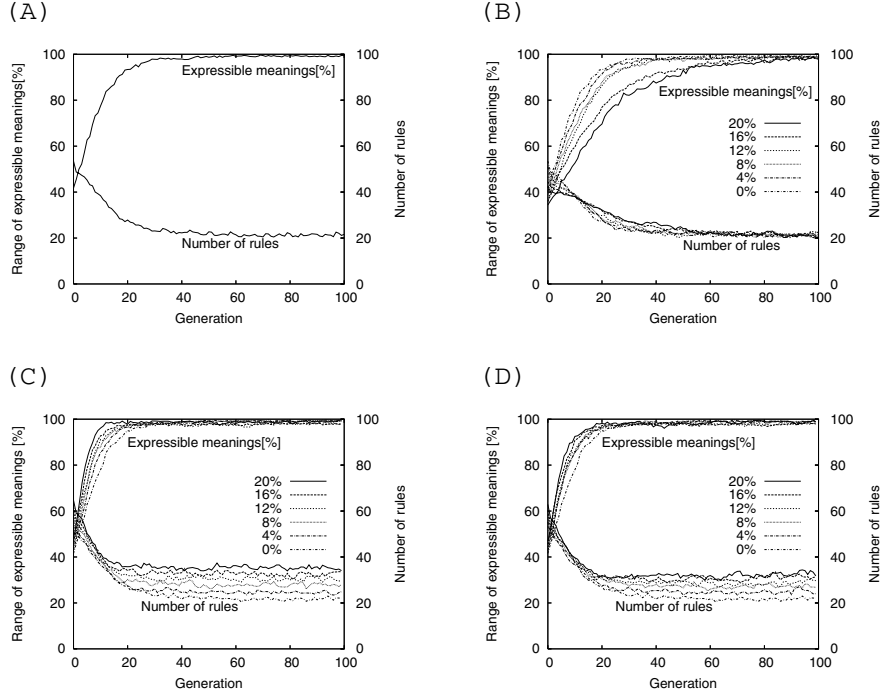
Fig. 1. The movement of the number of rules and expressivity per generation. (A). The result of Kirby(2002). (B). Result of experiment (I). (C). Result of experiment (II). (D). Result of experiment (III).

Definition 1 (Compositional rule, Holistic rule, and Lexical rule). *A grammar rule including non-terminal symbols for categories is called a compositional rule while the others holistic rules. Also, a rule which consists of a monadic terminal constant is called a lexical rule[1].*

Definition 2 (Compositional level and Expressivity). *Compositional level is the number of variables included in a compositional rule. Expressivity is the ratio of the utterable meanings against the whole meaning space.*

The experiment was carried out until the 100th generation. In fact, we have carried out until the 1000th generation; however, both of expressivity and the number of rules converged until the 100th generation, and thus we discuss the result derived by the 100th generation.

3.2   Pilot Experiment: Kirby's Model

Figure 1(A) is the result of the average of 100 trials, and shows the tendency of the number of rules and expressivity per generation. In the early stages, the language has low expressivity and a large size of grammar rules; however, through

---

[1] E.g., $N/john \rightarrow$ j.

the generations, the language universally acquires higher expressivity and the number of grammar rules subsides.

In the early stages, the increase of holistic rules leads the growth of expressivity, and the acceleration of generalization. This process leads the grammar to be compositional, so most of the rules in the later stages become compositional.

In fact, we have found a grammar with 100 % of expressivity consisting of 11 rules, consisting of ten lexical rules and one level–3 compositional rule (see Definition 2).

3.3   Experiment (I): Listener Ignores Incomprehensible Utterances

In this experiment, the listener does not use incomprehensible utterances for learning, i.e., the experimental setting is the same as the pilot experiment, except that the number of available utterances is reduced.

Figure 1(B) is the result of the average of 100 trials. Each line denotes the rate of the incomprehensible utterances to the received utterances (50 times), that is, 0 %, 4 %, 8 %, 12 %, 16 %, and 20 %.

From Figure 1(B), we can observe that the grammars anyhow converge regardless of the incomprehensible rate, and there is not a significant difference in the number of rules. However, there is a difference in convergence generation. When the incomprehensible rate is 0 %, the number of grammar rules converges around the 30th generation, but when the rate is 20 %, it becomes around the 50th generation and the expressivity does not converge until around the 80th generation. This indicates that the listener with more amount of learning data can acquire the higher expressivity and the smaller grammar rules in a shorter period.

3.4   Experiment (II): Listener Arbitrarily Interprets Incomprehensible Utterances

In this experiment, the listener chooses a meaning from the meaning space arbitrarily, and puts it to an incomprehensible utterance.

Figure 1(C) is the result of the average of 100 trials. Each line denotes the rate of the incomprehensible utterances to the received utterances as 0 %, 4 %, 8 %, 12 %, 16 %, 20 %, as Figure 1(B).

As compared to experiment (I) (Figure 1(B)), the result of experiment (II) (Figure 1(C)) shows high expressivity in shorter period. Complementing meanings to incomprehensible utterances, the listener can superficially increase his grammar knowledge. However, these ungrounded rules become noise to generalize the knowledge, and thus the size of the grammar rules becomes larger. In terms of expressivity, there is an insignificant difference between experiment (I) and experiment (II) after convergence.

3.5   Experiment (III): Listener Applies the Symmetry Bias to Interpret Incomprehensible Utterances

If the listener receives incomprehensible utterances, he follows Algorithm 2.

```
if there is a string among compositional level 0 rules which exactly matches the
utterance then
    apply the meaning of this rule;
else
    if the listener can derive the same string as the utterance using rules whose
    compositional 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 longest to any one of the substrings;
        if there exists such lexical rules then
            The listener chooses utterances which include any one of such words;
            From the chosen utterances, the listener picks out utterances as
            candidates which are uttered most frequently;
            if the candidate is only one then
                apply the meaning of this candidate to the incomprehensible
                utterance;
            else
                Among the other substrings, search lexical rules which correspond
                to one of the substrings longest;
                Choose the most frequent candidates;
                if such candidate is only one then
                    apply the meaning of this candidate to the incomprehensible
                    utterance;
                else
                    Choose one from the candidates;
                end
            end
        else
            Choose a meaning from the meaning space randomly, and apply it to
            the incomprehensible utterance.
        end
    end
end
```

Algorithm 2. The bias algorithm

Similar to the previous three experiments, the incomprehensible rates are set to 0%, 4%, 8%, 12%, 16%, and 20%. However, different from previous three experiments, incomprehensible utterances are given at the tail of the 50 utterances, e.g., in case of 20%, the first 40 sentences are paired with PAS while the rest 10 utterances are meaningless. This is because Algorithm 2 must be evoked after a certain accumulation of grammar knowledge. Namely, the listener could consider in such a way that the intention of the speaker was not clear but referring back to the previous knowledge the listener could partially guess what the speaker had said. We contend that this is exactly the inference by the symmetry bias.

Figure 1(D), the result of the average of 100 trials, shows that the expressivity is higher when the listener does not understand all the utterances, since the listener augments the number of grammar rules as experiment (II).

4   Comparison of Experimental Results

In this chapter, we compare the result of experiment (I) to that of experiment (III) and the result of experiment (II) to that of experiment (III), where we use the results of incomprehensible rate is 20 % as prominent examples. Table 1 shows the average number of rules at the 100th generation.

Firstly, we compare experiment (I) to experiment (III). By the effect of inference, the listener can get more learning data excluding noise data for leaning. This is why we can observe the expressivity of experiment (III) converges faster than experiment (I), i.e., inferring the meanings of incomprehensible utterances is more efficient in terms of convergence speed.

Here, we show how the symmetry bias works to construct compositional grammars using a concrete example from experiment (III) where the incomprehensible rate is 20 %, and the speaker's 45th utterance is 'esk'. The listener has the following lexical rule:

$$A/love \rightarrow \text{s}$$

The utterance 'esk' being decomposed to substrings, 's' corresponds to the lexical rule, so the listener infers as follows:

$$love(???,???) \leftarrow \text{esk}$$

Investigating the frequency of the words which co-occur with the word *love* from the speaker's utterances up to this time, the listener finds *gavin* has occurred three times, *john* has occurred twice, and *heather*, *mary*, *pete* has occurred once, respectively. Thus, the listener chooses *gavin*. Next, the listener chooses either one of the followings based on frequency:

$$love(gavin,???)$$

$$love(???,gavin)$$

Table 1. Average number of rules: Generation 100

	experiment (I)	experiment (II)	experiment (III)
0 %	21.1	22.1	22.1
4 %	22.3	24.9	24.0
8 %	20.9	28.0	26.8
12 %	22.6	29.6	29.7
16 %	20.0	33.6	31.3
20 %	20.5	34.1	31.8

Looking back to the log, *mary* has occurred once in the form of $love(mary, gavin)$ and the others have not occurred. Thus, the listener adds the following rule to his knowledge:

$$love(mary, gavin) \rightarrow \text{esk}$$

This enables the listener to learn a lexical rule '$A/love \rightarrow$ s', so his grammar changes as follows:

$$\begin{cases} A/love \rightarrow \text{s} \\ S/p(mary, gavin) \rightarrow \text{ek } A/p \end{cases}$$

Thus, the listener acquires the compositional rule. Since the compositional level of the added rules is higher, the language necessarily has higher expressivity. Therefore, compared experiment (I) to experiment (III), the expressivity of the latter converges in shorter period.

When the complement of meanings by inference is incomplete, the listener acquires rules which were originally not included in the speaker's grammars, so the listener cannot generalize such rules well. Thus, there remains low level compositional rules in the listener's grammar. This is why Table 1 shows that the number of rules of experiment (I) is less than that of experiment (III) as 20.5 versus 31.8.

The listener's strategy of experiment (I) which does not treat incomprehensible utterances to learning, decrease data for learning. On the other hand, the listener's strategy of experiment (III) keeps the same amount of data for learning. The symmetry bias being applied, the grammar converges in shorter period, although the number of rules increases. In terms of convergence speed, it is efficient to apply the symmetry bias to learning.

Next, we compare experiment (II) to experiment (III) under the incomprehensible rate is 20 %. There are not so much of differences in convergence speed and the expressivity between these experiments. On the other hand, there is a difference in the number of rules between them, as 34.1 and 31.8. This difference indicates the efficiency of the symmetry bias in learning.

Figure 2 shows the number of guessed meanings by the symmetry bias under the rate is 20 % by the average of 100 trials. In early stages, because of the failure of the
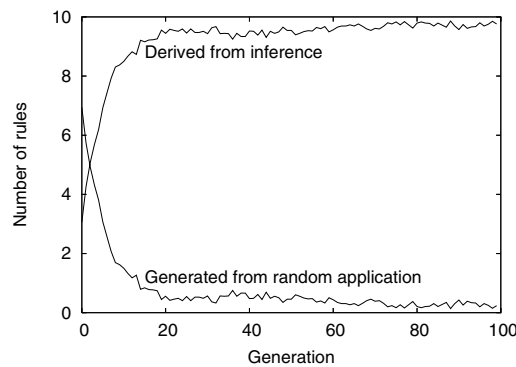
Fig. 2. The number of rules which are derived from inference vs. random application

inference, the listener has applied meanings to the incomprehensible utterances randomly. In later generations, the listener becomes to be able to apply a meaning by the inference. From Figure 2, we can observe that the number of rules with random PAS decreases, while those which own correct PAS increases. Also, from figure 1(D), we can see that the expressivity and the number of rules converges around the 20th generation. Thus, we can observe that the inference failure had steadily decreased after the agent acquired a certain level of grammar knowledge.

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

5   Conclusion

In this paper, 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 [13], and have built the symmetry bias into our model. In the simulation, the listener received utterances without the speaker's intention at a certain rate, and the listener interpreted such incomprehensible utterances with three kinds of strategies such as (I) ignoring, (II) attaching meanings randomly, and (III) attaching meanings by the symmetry bias. For each of the strategies, we observed expressivity, number of rules, and compositional level. As a result of the experiments, the listener who has employed strategy (III),

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

Our future works are summarized as follows. Firstly, we should consider to include other cognitive biases than the symmetry bias. Secondly, we should expand the meaning space and grammars of our model and restudy the effectiveness, since the effect of the bias was not so prominent at the current simple and limited language. Thirdly, 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

References

1. Schafer, G., Plunkett, K.: Rapid word learning by 15-month-olds under tightly controlled conditions. Child Development 69, 309–320 (1998)
2. Imai, M., Gentner, D.: Children's theory of word meanings: The role of shape similarity in early acquisition. Cognitive Development 9(1), 45–75 (1994)

3. Markman, E.M.: Categorization and naming in children: Problems of induction. MIT Press, Cambridge (1989)
4. Quine, W.V.O.: Word and Object. MIT Press, Cambridge (1960)
5. Imai, M., Gentner, D.: A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. Cognition 62(2), 169–200 (1997)
6. Landau, B., Smith, L.B., Jones, S.S.: The importance of shape in early lexical learning. Cognitive Development 3(3), 299–321 (1988)
7. Landau, B., Smith, L.B., Jones, S.S.: Syntactic context and the shape bias in children's and adult's lexical learning. Journal of Memory and Language 31(6), 807–825 (1992)
8. Markman, E.M.: Constraints children place on word meanings. Cognitive Science: A Multidisciplinary Journal 14(1), 57–77 (1990)
9. Markman, E.M., Wasow, J.L., Hansen, M.B.: Use of the mutual exclusivity assumption by young word learners. Cognitive Psychology 47(3), 241–275 (2003)
10. Hattori, M.: 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), vol. 1, pp. 163–168 (2003)
11. Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., Carrigan, P.: A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. Journal of the Experimental Analysis of Behavior 37(1), 23–44 (1982)
12. Yamazaki, Y.: Logical and illogical behavior in animals. Japanese Psychological Research 46(3), 195–206 (2004)
13. Kirby, S.: Learning, bottlenecks and the evolution of recursive syntax. In: Linguistic Evolution through Language Acquisition. Cambridge University Press, Cambridge (2002)
14. Chomsky, N.: Knowledge of Language. Praeger, New York (1986)
15. Bickerton, D.: Language and Species. University of Chicago Press (1990)
16. Hurford, J.: Language and Number: the Emergence of a Cognitive System. Blackwell, Oxford (1987)
17. Kirby, S.: Function, Selection, and Innateness: The Emergence of Language Universals. Oxford University Press, Oxford (1999)