Proceedings of International Conference
on Artificial Intelligence

Common Language Acquisition by Multi-Agents

Tetsuo Ono Satoshi Tojo Satoshi Sato

School of Information Science
Japan Advanced Institute of Science and Technology, Hokuriku
Tatsunokuchi, Ishikawa, Japan
tono@jaist.ac.jp

December 20, 1996

Abstract

We propose a multi-agent model, where each agent learns a common language in the community, modifying his/her own initial grammar through a number of exchanges of sentences with other agents. As the initial step toward the realization of this model, we presuppose that a child agent who has a primitive grammar set is thrown into a community consisting of adult agents. The child agent modifies his/her grammar and learns a refined one, imitating adult's sentences, while the adult agents also loosen their grammar in order to accept what the child agent says tolerantly.

In this paper, we report the result of experimentation with this model, together with grammar representation and learning methods. We showed that our model could realize adaptability, which is one of the important features of natural language.

1 Introduction

Grammar of natural language is a statistical phenomenon, while that of artificial language is a regulation. To put it concretely, once a new usage in natural language comes to be prevalent, the grammar itself should be changed accordingly. On the other hand, the grammar of formal language defines the language clearly at the beginning stage, so that there is no ambiguity in the generation process.

It is commonly assumed that each individual has to have his/her own grammar when one can generate or recognize a sentence in a community appropriately. Although each grammar representation differs, the language itself must be common in the community. Therefore, those various individual grammars should be changed and reproduced, through interaction among speakers, to meet this social requirement.

The statistical and changeable natures as stated above can be realized by a multi-agent model, where several semi-autonomous agents interact or work together to perform some set of tasks. In other words, the multi-agent model can realize a system that adapts to a dynamic environment by cooperation between agents who have the ability of self-modification. In this paper, we regard the process of constructing the system as a process of forming a common language in the community. An agent is capable of self-modifying his/her grammar, and changing his/her language accordingly. The modification mechanism should be based on the laws of probability, that is, a grammar rule that often parses speeches in the community becomes dominant. We believe that this statistical point of view well explains the nature of our grammar.

Research to construct a common language (or protocol) based on evolutionary and biological interests has been that [Koza 91] attempted to organize a colony of artificial ants by communication through their pheromones evolutionally. [Werner and Dyer 91] designed to self-organize a common protocol for mate finding in a group of artificial organisms. However, it is difficult for them to raise the level from those poor languages to the human language alone by artificial evolution. There is also the research of evolution of symbolic grammar systems as the object to model a higher level language [Hashimoto and Ikegami 95]. Yet it does not intend to explain the natural language phenomena. The chief aim of our research is to model the language acquisition process of human beings by multi-agents.

The final goal of our research is to propose a *foreigner agent* model, where each agent who has a different grammar learns a common language in a community. However, prior to this model, there are many problems to be solved; for example, how to communicate without any common concepts, how to know each other's understanding level, and so on. For this reason, as the first step, we propose a *child* and *adult agent* model, where all agents have the same dictionary as common concepts. In this model, we presuppose that the child agents who have a primitive grammar set are thrown into the adult community. The child agents modify their grammar and learn the refined one, imitating adult's sentences, while the adult agents also loosen their grammar in order to accept what the child agents say tolerantly.

First, we define the common language in this paper. Next, we describe the grammar representation and our learning method for this model in the following sections. Finally, we show the experimental result using the proposed model.

2 Common Language

This section defines a *common language* that we use in this paper. A grammar that an agent has is denoted by G_i, where each agent is identified by index i. A language, which means a set of sentences that the grammar G_i can generate and accept, is denoted by $\mathcal{L}(G_i)$. The language of the community consisting of n agents, which is defined as a union of the languages $\mathcal{L}(G_i)$ $(1 \leq i \leq n)$, is denoted by $\mathcal{L}_G$:

$$\mathcal{L}_G = \mathcal{L}(G_1) \cup \mathcal{L}(G_2) \cup \ldots \cup \mathcal{L}(G_n) = \bigcup_{1 \leq i \leq n} \mathcal{L}(G_i).$$

In the community, the *common language* $\mathcal{L}_C$ is defined as a set of the sentences which have appeared in the community over a given frequency threshold. The relation between $\mathcal{L}_C$ and $\mathcal{L}_G$ is defined as follows:

$$\mathcal{L}_C \subset \mathcal{L}_G$$

Each agent modifies his/her own grammar G_i in order that his/her language $\mathcal{L}(G_i)$ may meet the common language $\mathcal{L}_C$. This is a process of language acquisition by each agent. The common language $\mathcal{L}_C$ is modified accordingly, because each language $\mathcal{L}(G_i)$ that is the element of the $\mathcal{L}_C$ changes. This is also a process of common language acquisition by multi-agents. In the following sections, we propose a multi-agent model for this process.

3 Model of Grammar Acquisition

This section describes a model of grammar acquisition, that is, a formalization of agent and communication between them, a framework of grammar, learning methods, and so on.

3.1 Agents and Communication

Our first model consists of child and adult agents (Figure 1). Each agent has grammar based on GPSG (Generalized Phrase Structure Grammar) [Gazdar et al. 85], and two learning methods.

The child and the adult agents differ from each other on the generativity of grammar, but they have the same learning methods. As a child's initial grammar G_C has not been learned yet, the child speaks ungrammatical sentences for a community involuntarily. On the other hand, as an adult's initial grammar G_A has already been learned sufficiently, the adult speaks grammatically for the community. These agents try to learn a common language $\mathcal{L}_C$ used in the community, modifying their own initial grammar through a number of exchanges of sentences with other agents.

The outline of the processing in this communication is as follows.

1. Each agent speaks a sentence in turn.
2. The sentence is broadcasted to all agents.
3. Every agent tries to parse the sentence.
4. Every agent does learning (assignment of rules' weight).
5. One time step is completed if all agents made an utterance once.

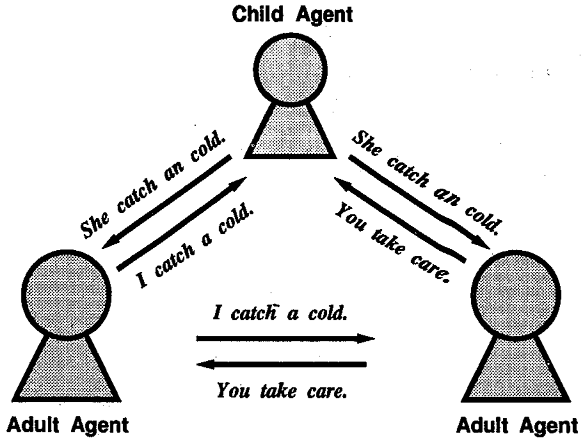

Figure 1: Communication between Agents

6. An agent learns (refinement of rules) if the number of parsed trees is over a given constant.

The agents use simple sentences in English as means of communication. When the agents speak sentences, the adult does so randomly but the child imitatively. That is to say, the adult chooses words and rules randomly, and speaks grammatically, while the child reuses the rules that just used in the adult's speech, changing words randomly.

Two learning methods, that is, the assignment of rules' weight and the refinement of rules, are described in the Section 3.3.

3.2 Framework of Grammar

This section describes the framework of the grammar that each agent has. Each rule of the grammar based on GPSG has a weight that gives the rule an order of priority (Tables 1 and 2). A symbol in these rules is represented by a row of bits (Figure 2). These bits express the following five features and a word identification number (WORD-ID). Here, '+' denotes an applicable feature, '–' denotes an inapplicable feature.

- MAJOR: a primitive feature.
  b_0: a nominal group ([+N]: 1, [–N]: 0).
  b_1: a verbal group ([+V]: 1, [–V]: 0).
- BAR: a bar level.
  b_2, b_3: bar levels (0: 00, 1: 01, 2: 10, 3: 11).
- FORM: a form of word.
  - FORM 1: an ending 1 of word.
    b_4, b_5: adding *-s* (+: 00, –: 01).
  - FORM 2: an ending 2 of word.
    b_4, b_5: adding *-ed* (+: 00, –: 01).
  - FORM 3: a beginning of word.
    b_4, b_5: a vowel (+: 00, –: 01).

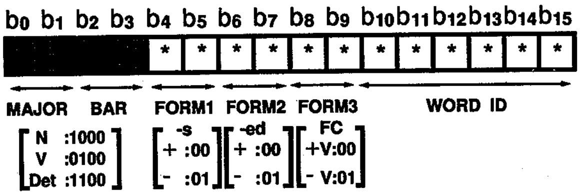

Figure 2: Representation of Symbol

Table 1: Child Agent's Grammar G_C

No.	rule	weight
1	S* → NP* VP*	w_1
2	NP* → Det* N*	w_2
3	NP* → N*	w_3
4	VP* → V* NP*	w_4
5	VP* → V*	w_5

For example, a noun "dogs" is reconstructed as the expression that has the category type characterized by [+N] and [−V], the bar level 0, the ending of word "-s" and not a vowel in the beginning of word. Therefore it is expressed by the row of bits '1000000101' in our model (omitting WORD-ID).

The MAJOR and BAR features (b_0-b_3) are called *known-bits*, whereas FORM1, FORM2 and FORM3 features (b_4-b_9) are called *unknown-bits*. The *known-bit* is a common bit, which can be understood both by the child and by the adult. By contrast, the *unknown-bit* is understood only by the adult and not by the child. The aim of the child is to discover and determine the value of *unknown-bits*.

Tables 1 and 2 represent example grammar in our model. Table 1 is the child's grammar G_C that is characterized by the presence of the wild card '*' which can fit any bit; with this wild card, the child's grammar can be considered as 'general.' Table 2 is the adult's grammar G_A that is characterized by specificity where some features are determined, which is expressed as '[F].' Generally speaking, the weight of a specific rule is larger than that of a general one, so that a specific rule is basically used if multiple rules can be applied in the speaking and parsing processes. Therefore, the adults can speak grammatical sentences, and parse the sentences correctly. In this model, agents parse sentences by the active chart parser [Winograd 83].

3.3 Learning Method

In this model, we adopt two learning methods for grammar acquisition. One is the refinement of each rule, and the other is the weight assignment on each rule. Although these are rather simple, we show that we can realize a robust system with those mechanisms.

Table 2: Adult Agent's Grammar G_A

No.	rule	weight
1	S* → NP* VP*	w_1
2	NP* → Det* N*	w_2
3	NP* → N*	w_3
4	VP* → V* NP*	w_4
5	VP* → V*	w_5
6	S[F] → NP[F] VP[F]	w_6
7	NP[F] → Det[F] N[F]	w_7
8	NP[F] → N[F]	w_8

3.3.1 Refinement of Rules

The refinement of rules is a method for discovering features from a set of the spoken sentences $\mathcal{L}_G$. Procedure of this method is as follows:

1. Input a parsing tree $\mathcal{T} = \{R_i | 1 \leq i \leq n\}$, where R_i is a used rule in the parsing, and n is the number of used rules.

2. Compute a conditional probability $p(F_k|F_j)$ in a rule R_i using CAP (a principle in GPSG):
$$p(F_k|F_j) = \frac{p(F_j \cap F_k)}{p(F_j)},$$
where F_j and F_k are features in R_i, i and j are parameter ($j = 1$ *to* 5, correspondingly $k = j + 1$ *to* 6).

3. Generate a new rule assigning values to features F_j and F_k,
if $p(F_k|F_j)$ is larger than a given threshold θ.

We show this learning process, taking a sentence "dogs come," as an example. This example omits the process of 'NP → N' and 'VP → V,' also the WORD-ID in the representation of the symbol, for simplicity.

1. Sentence: "dogs come."

2. Reference of dictionary:
dogs: 1000000101,
come: 0101010101.

3. Grammar rule: S* → NP* VP*.

4. Bit representation of rule:
0110**** → 1010****** 0101******.**

5. Instantiated rule (Rule after parsing):
0110010101 → 1010000101 0101010101.

6. Conditional probability:
$p(F_1|F_4)$, $p(F_2|F_4)$, $p(F_1|F_2) > \theta$.

7. Rule generation:
011001** → 10100001** 010101****.**

Table 3: Example of Merging Rules

time	rule no.	rule	weight
1	R_0	NP → Det* N*	w_0
2	R_1	NP → Det[a] N[sg]	w_1
	R_0	NP → Det* N*;	w_0
3	R_2	NP → Det[the] N[pl];	w_2
	R_0	NP → Det* N*;	w_0
	R_1	NP → Det[a] N[sg];	w_1
4	R_2	NP → Det[the] N[pl];	w_2
	R_3	NP → Det[a] N[sg];	w_3
	R_0	NP → Det* N*;	w_0
	R_1	NP → Det[a] N[sg];	w_1
5	R_4	NP → Det[a] N[sg];	$w_4 \Leftarrow w_3 + w_1$
	R_2	NP → Det[the] N[pl];	w_2
	R_0	NP → Det* N*;	w_0

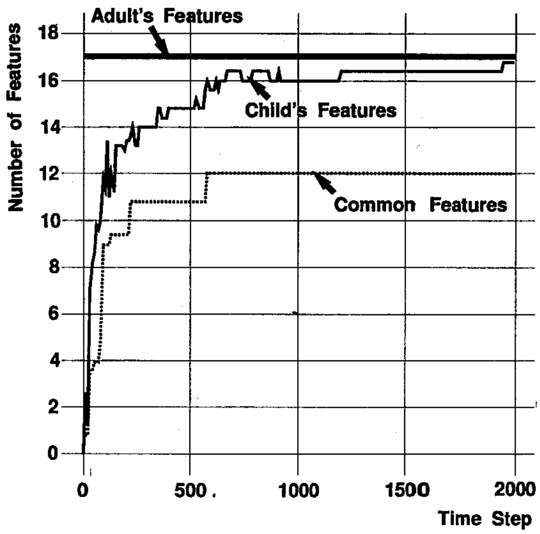

Figure 5: Result of Experiment 1

4 Results of Experiments

We implemented a multi-agent system using the proposed model. We did two experiments on the system, varying the number of child agents.

4.1 Experiment 1

In the experiment 1, we examined that only one child was thrown into the environment where 3 adults exist. We executed 2,000 time steps for each simulation, and we experimented 10 simulations. Here, 1 time step was completed when each agent spoke a sentence. Each agent had 20 words in his/her dictionary. The parameters were fixed as follows: the threshold of the conditional probability $\theta = 0.7$, and the margin of the new rule $r = 100$.

The average result of the 10 simulations is shown in Figure 5, where the number of the child's acquired features is shown by a fine line, and the number of the adult's features is shown by a bold line at 17. The child was acquiring the adult's features quickly up to around 250 time steps, and kept on acquiring around 250-700 time steps slowly. After 700 time steps, he/she preserved the same level, and almost acquired as many features as the adult had done near 2,000 time steps. On the other hand, the number of the adult's features did not change all the time, however the weight of the rules changed a little.

The number of common features among all agents is shown by a broken line in Figure 5. The number of the features increased in the same way as that of the child's acquired ones. The number of the rules which were assigned these features was 6 when the simulations were finished. These rules generated or parsed 71 percent of all the spoken sentences at that time.

As the result of the experiment, the child almost acquired as many features as the adult had done. However, there were a few different features between the child and the adult. The reason this happened is that our model only discovers regularity of the relation between the features on the word forms. Meanwhile all agents held the 6 rules in common when the simulations were completed. We consider that a common language is formed in the community at this time.

4.2 Experiment 2

In the experiment 2, we examined the case that the number of the child agents was the same as that of the adults. Both the numbers of agents of these two kinds were 3. For the rest, this experimental condition was the same as previous one.

The average result of the 10 simulations is shown in Figure 6, where the number of the child's features is shown in the lower part, the number of the adult's features is shown in the upper part. The child was acquiring the features rapidly up to around 300 time steps, in contrast, the adult did not change the number of his/her features at all. After 300 time steps, the child preserved the number of the features, whereas the adult began to acquire the features gradually. Finally, both the lines came to parallel roughly after 1,000 time steps.

The number of common features among all agents is shown by a broken line in Figure 6. The number of the features increased more slowly than that of the child's ones, and kept on taking the fixed value after 1,300 time steps. The number of the rules which were assigned these features was 3 when the simulations were finished. These rules generated or parsed 70 percent of all the spoken sentences at that time.

As the result of the experiment, the child and the adult acquired nearly 8 and 4 features respectively. Since the features that the child acquired were only a half as many features as he/she did in experiment 1, he/she spoke more ungrammatical sentences for an adult's initial grammar. The features that the adult acquired caused generating the same ungrammatical sentences as the child. In other words, the adult loosened their grammar in order to accept what the child said tolerantly. We regard this as a process of forming

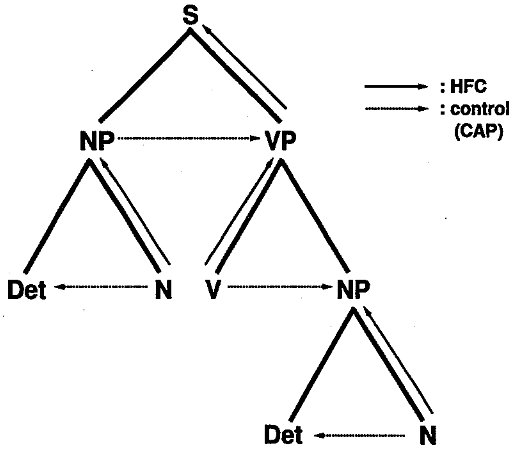

Figure 3: HFC and CAP

When an agent receives the spoken sentence "dogs come," he/she parses it first. The agent refers to a dictionary that is used by both the child and the adult, and he/she selects an appropriate rule for parsing. Here, a rule 'S* → NP* VP*' is applied. The agent acquires an instantiated rule, that is, each feature in the rule is assigned some value. The features in the left-hand side of the rule are filled with the values by a head feature convention (HFC) in GPSG. The HFC means that the features of the left-hand side inherit from those of the head of the right-hand side in the rule (Figure 3). As VP is the head in this rule, the features of S inherit from those of VP. Next, the agent computes a conditional probability between features, in which he/she uses a control agreement principle (CAP) in GPSG. The CAP means which word controls which (Figure 3). As NP controls VP, the agent can compute the conditional probability $p(F_{NP_k}|F_{VP_j})$ in the reverse direction (Figure 4). The reason why we use the reverse direction of the CAP is that this means is useful to compute the conditional probability exactly in this model. The adequateness of the CAP will be discussed in the final section. As the result of the computation, when the three probabilities become larger than the threshold θ, a new rule is generated, being assigned values to the features F_1 = '00,' F_2 = '01' and F_4 = '01.'

As this procedure shows, the objective of this method is to discover regularity of the relation between features on the form of words. In this model, we expect that the child mainly refines his/her own rules from *general* forms to *specific* ones through discovering features and determining their values.

3.3.2 Weight Assignment on Rules

The weight assignment on rules is a method for adapting rules to a common language $\mathcal{L}_C$ used in a community. The weight of the rules varies in the following three conditions:

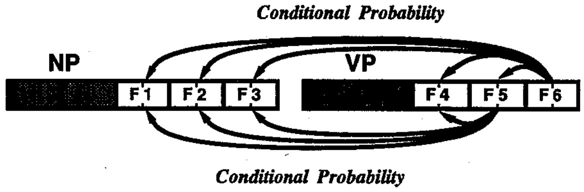

Figure 4: Computation of Conditional Probability

- Completion of parsing
- Generation of new rules
- Rule merging

Completion of parsing. If an agent can parse the spoken sentence completely, the weights of all the rules which are used for the processing are added a score p. The score p is defined as the number of used rules for this parsing process.

If several applicable rules compete in the parsing process, the one that has the largest weight is chosen among them. In this model, the specificity of a rule, that is the number of bit '1' or '0' in a rule, is not considered in the rule competition, because those specific rules would acquire larger weights naturally through the refinement of rules.

In this method, the rule that is used for parsing a longer sentence gets a higher score. In other word, the longer the sentence an agent recognizes or speaks, the better the agent learns.

Generation of new rules. If a new rule is generated by the refinement of rules, the rule is given the weight of $q+r$ that is larger than that of the previous rule q, where the added margin r is a given positive constant.

Although this method is very simple, we can prove its robustness for different environments in the following experiments.

Rule merging. This method may happen to produce the same rules with different weights. In this case, one rule is merged to the other with larger weight, and the merged rule becomes the sum of previous two rules.

An example of this process is shown in Table 3. If an NP1 " 'a' + singular noun (sg)" is used frequently at time 1, then a new rule 1 (R_1) is generated from R_0 at time 2, and at this time $w_1 > w_0$. Next, if an NP2 " 'the' + plural noun (pl)" is used repeatedly, R_0 is applied each time and its weight w_0 increases. At time 3, R_2 is generated from R_0, and at this time $w_2 > w_0 > w_1$. Thereafter, if NP1 is used again, R_0 is applied [1]. At time 4, R_3 is generated from R_0, and at this time $w_2 > w_3 > w_0 > w_1$. As the same rule, R_1 and R_3, exists after all, two rules are merged into R_4, and at this time $w_4 > w_2 > w_0$.

[1] Notice that the rule which is applied to NP1 is not R_1.

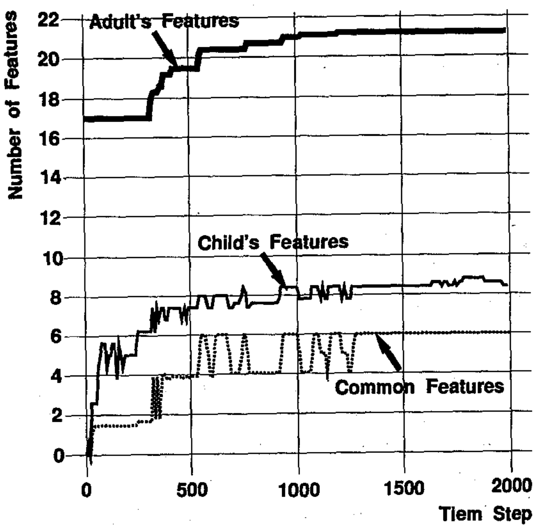

Figure 6: Result of Experiment 2

a common language in the community.

This result, furthermore, means that the grammar changes in the community dynamically by communication between the agents who have the ability of self-modification. Our proposed model realized adaptability which was one of the important characteristics that a multi-agent system had, and it was the aim of this paper.

5 Discussion and Conclusion

In this paper, we have studied common language acquisition by multi-agent model. We proposed the multi-agent model which consisted of the child and the adult agents. Each agent had the grammar based on GPSG, and two learning methods. The child agents modified their grammar and learned a refined one, imitating adult's sentences, while the adult agents also loosened their grammar in order to accept what the child agents said tolerantly. We implemented a multi-agent system based on our model, and experimented on the computer. The results of the experiments showed that our model could realize the process of the formation of the common language, and the adaptability which is one of the important features of natural language.

First, we discuss a point of difference between the related work and our research. The research to construct a common language (or protocol) based on evolutionary and biological interests has been. However, it is difficult for them to raise the level from those poor languages to the human language by artificial evolution alone. Generally speaking, the language acquisition process of a human being itself has never yet been made clear. Accordingly, it is necessary to set up some assumptions and constraints in order to model the language learning process. We, therefore, adopt the assumptions based on GPSG, for example, CAP, HFC and so on. To use these assumptions facilitates constructing a computational model for natural language. Because GPSG is expressed by CFG (Context Free Grammar) without a transformation rule, and a feature system including CAP and HFC can briefly describe the relation between categories in a rule, and between rules in a parsed tree. We believe that our position is sound both from the scientific point of view, and from engineering research.

Next, why do we need a multi-agent model for grammar acquisition? Some researches of grammar acquisition try to abstract the grammar statistically from many examples of sentences, for example, the acquisition of the phrase structure from a corpus [Brill and Marcus 92]. The viewpoint of these researches and ours have basically much in common, where the grammar of natural language is regarded as statistical a phenomenon. However, those researches do not consider another characteristic of the grammar, that is a dynamic change in accordance with the change of a community. The multi-agent model can realize this characteristic. We realized its adaptability which was one of the important features of natural language.

To explain linguistic phenomena, it is important to study the fusion process between the two different grammars. We proposed that the framework of this research is to solve the problem of second language acquisition and the autonomous formation of a dialect. This framework also suggests a useful view to engineering applications, for example, machine translation. Therefore, we regard the *foreigner agent* model, where each agent that has different grammar learns a common language in the community, as a more meaningful research.

References

[Brill and Marcus 92] Brill, E. and Marcus, M. (1992): Automatically Acquiring Phrase Structure using Distributional Analysis, *Proceedings of the DARPA Speech and Natural Language Workshop*, p.155-159, Morgan Kaufmann.

[Gazdar et al. 85] Gazdar, G., Klein, E., Pullum, G.K. and Sag, I.A. (1985): *Generalized Phrase Structure Grammar*, Harvard University Press.

[Hashimoto and Ikegami 95] Hashimoto, T. and Ikegami, T. (1995): Evolution of Symbolic Grammar Systems, *Third European Conference on Artificial Life*, p.812-823, Springer.

[Koza 91] Koza, J.R. (1991): Genetic Evolution and Co-Evolution of Computer Programs, in C.G. Langton et al. (Eds.), *Artificial Life II*, p.603-629, Addison-Wesley.

[Werner and Dyer 91] Werner, G.M. and Dyer, M.G. (1991): Evolution of Communication in Artificial Organisms, in C.G. Langton et al. (Eds.), *Artificial Life II*, p.659-687, Addison-Wesley.

[Winograd 83] Winograd, T. (1983): *Language as a cognitive process*, Addison-Wesley.