Spoken Dialogue Processing Method Using Inductive Learning with Genetic Algorithm

Yasutomo Kimura,[1] Kenji Araki,[1] Yoshio Momouchi,[2] and Koji Tochinai[1]

[1]Graduate School of Engineering, Hokkaido University, Sapporo, 060-8628 Japan

[2]Graduate School of Engineering, Hokkaido-Gakuen University, Sapporo, 064-0926 Japan

SUMMARY

This paper describes a spoken dialogue processing, which includes the learning using spoken dialogue examples. Most of the spoken dialogue systems up to the present are task-oriented, where the processing is based on the prespecified generation rules and database. It is then difficult to handle various topics, such as miscellaneous talks in daily dialogue. In the proposed method, the dialogue between the system and the user is processed as the spoken dialogue example, and the rule is acquired based on the pair of system response and user utterance, through inductive learning using the genetic algorithm. In other words, it is not necessary to prepare the training data beforehand, but the response is composed using the rules acquired from the actual dialogue examples. With this approach, the learning can be executed based on the dynamic data, and the deviation depending on the data is reduced. It is intended in this paper to examine the usefulness of the proposed method. Miscellaneous talks are considered. ELIZA-type system, which is extended to the spoken dialogue, and the proposed method are compared through a comparison experiment and an experiment with multiple examinees. As a result, it is verified that the total ratio of the correct and almost correct responses is improved from 66.3% to 76.1%, and the effective responses can be constructed using the rules acquired from actual dialogue examples. This improvement of 9.8 points indicates that the proposed method is effective in handling miscellaneous talks. © 2004 Wiley Periodicals,

1. Introduction

Studies of dialogue have been conducted since the early stage of computer development. An example is the classical spoken dialogue system ELIZA [1]. ELIZA utilizes skillfully the keywords to acquire information from patients with mental diseases. In ELIZA, no constraint is imposed on the input sentence from the user. Consequently, some patients with mental disease, who used ELIZA, misunderstood that they got an actual human response. Since static rules are used, however, it is difficult to generate responses adapted to the user, and the user is sometimes disappointed when the system is utilized for a long time.

In contrast to ELIZA, in SHRDLU [2] the input sentence from the user is limited to the world of building blocks. SHRDLU can realize a secure response in the world of building blocks. Since, however, the knowledge is provided manually beforehand, the system cannot cope with an unexpected situation. Another point is that the extension of the range of application is difficult.

Among recent studies of non-task-oriented dialogue, there is a dialogue system [3] that aims at a new therapy for

depressive patients. Compared to ELIZA, the system can realize a more complex processing. The system generates responses using the production rule, based on the thesaurus, the semantic database, the topics tracking, and the query-response. Such processing is considered to be effective as a new therapeutic treatment for depressive patients. There is a problem, however, that text processing is used and the therapeutic process is emphasized. Consequently, all response rules are specified beforehand. In order to apply the system to a spoken dialogue system other than the system for depressive patients, new response rules must be prepared. This requires labor and cost, since construction of the response rules is most important.

There has been recent progress in speech recognition technology, and intensive studies are found regarding spoken dialogue. Nakagawa and colleagues constructed the "Mount Fuji sightseeing guide Japanese spoken dialogue system" [4], with the sightseeing guide as the task. Their system realizes high-quality response in the spoken dialogue, using the collaborative response of the system. However, the database for the sightseeing guide must be prepared. Consistent description and a database for each task are required.

Such problems can be solved if the spoken dialogue system can simulate human language acquisition [5]. Human language acquisition is based on daily dialogue. Daily dialogue contains not only the task-oriented dialogue [6, 7], but also various dialogues, such as miscellaneous talk, which progresses as the talk develops [8]. Kato and colleagues extended ELIZA to spoken dialogue, and constructed a 3DCG conversational robot "tea friend" [9]. They also investigated the system with a function for conversational learning in miscellaneous talk, but the system is still in a trail stage. Miscellaneous talk contains various topics, as well as nongrammatical utterances. Consequently, use of keywords, such as in ELIZA, may be effective in the first stage, but a dynamic adaptation based on actual dialogue becomes necessary, in order to satisfy the user.

We have been studying the Inductive Learning with Genetic Algorithm (GA-IL) to be applied to machine translation [10]. Machine translation using GA-IL acquires automatically the rules contained in the translation examples, and executes the translation processing. GA-IL executes learning with a smaller amount of data, compared to the statistical methods and the methods based on usage examples.

The statistical method determines the event based on the histogram. For example, in one method the frequency at which two words appear closely is examined statistically in terms of the conditional probability. Then, the associative inference is executed. In this approach, several problems are produced. Namely, when a large corpus is used in the statistical examination, a large amount of data is required, in order to realize an adequate association. The method is effective in handling a global scene, but it is difficult to adapt to a scene containing a large number of local features. The method may become deviated, depending on the given data.

The method based on usage examples provides processing based on the usage examples, but it is difficult to prepare all examples. Consequently, when a matched example does not exist, a similar example is utilized. In the conventional method, for example, the semantic closeness between the example pattern and the actual pattern is calculated in order to identify the pattern. In other words, similar patterns must be prepared beforehand, but a tremendous amount of labor is required to acquire a corpus of high quality and quantity.

In this paper, we propose a "Spoken Dialogue processing method using Inductive Learning with Genetic Algorithm" (GA-ILSD). The method executes learning by storing the actually observed dialogue examples. Using actual dialogue examples, effective responses can be generated using a smaller amount of data. Since the learning data are not required beforehand, manual preparation of the corpus is unnecessary, and the deviation that results from the data is eliminated.

In this paper, it is intended to acquire language from miscellaneous talks. In the evaluation experiment, the initial condition is set as the same, in order to examine the effectiveness of learning by GA-ILSD. As the initial condition, the experiment starts from the state in which the rule dictionary of GA-IL is empty. When the rule dictionary of GA-IL is insufficient, two methods are used for the response processing. When there is no adequate rule in the rule dictionary of GA-IL, the proposed system generates the response, using an algorithm based on ELIZA. This is called ELIZA-type response.

The ELIZA-type response is used in order to continue the dialogue between user and system and to acquire dialogue examples. The proposed system acquires the rules for the response from the dialogue examples, and generates the response, using five processes, including crossover, mutation, selection, and inductive learning. This is called GA-IL response. As the system continues the dialogue, GA-IL responses are gradually increased. We applied GA-IL to the spoken dialogue system, and verified that responses, which are effectively adapted to the user, can be generated from the learning of the actual dialogue examples.

In the comparison experiment to ELIZA-type system, the proposed method found that the total ratio of correct and almost correct responses is higher by 9.8%. By the correct response is meant the semantically correct response, and by the almost correct response is meant the response which is understandable, although not natural in representation. The details are discussed in Section 4.1.1.

Table 1.   Example of dialogue

A	1	$\alpha\delta\nu\beta\gamma\epsilon$
B	2	$\kappa\epsilon\zeta\vartheta\eta$
A	3	$\alpha\delta\Gamma\pi\gamma\epsilon$
B	4	$\xi\sigma\zeta\vartheta\eta$

Table 2.   Causal relationship of dialogue

System–user pair				
A			B	
$\underline{\alpha\delta}$	$\nu\beta$	$\gamma\epsilon$	$\kappa\epsilon$	$\underline{\zeta\vartheta\eta}$
$\underline{\alpha\delta}$	$\Gamma\pi$	$\gamma\epsilon$	$\xi\sigma$	$\underline{\zeta\vartheta\eta}$
Surface casual relation				
$\nu\beta$		$\kappa\varepsilon$		
$\Gamma\pi$		$\xi\sigma$		
Acquisition of rule for statement generation				
$\alpha\delta$	$@diff$	$\gamma\epsilon$	$@diff$	$\zeta\vartheta\eta$

$@diff$ represents the different part

2.   Basic Idea

It is intended to realize the human language acquisition process on a computer. In the dialogue processing that is generally used, the system is provided with the rules. Then, it is difficult to cope with an unexpected situation, when such occurs, and to generate a response. From such a viewpoint, we are attempting the engineering realization of a dialogue system that automatically acquires the rules for the dialogue. In order to realize such a language acquisition process, there should be a distinction between the ability to be prepared beforehand and the ability to be acquired later. In human language acquisition, knowledge and ability, which is already acquired, differ, depending on the stage of progress. The knowledge to acquire depends on what knowledge or ability is presumed for the user.

In this paper, the ability, which is provided before the learning, is assumed as the "ability to discriminate the common part and the different part, for two given events." It is further assumed in the character string in the dialogue between the user and the system that "there should be a certain relation between consecutive dialogues." Then, by discriminating the common part and the different part, the superficial causal relation or rules corresponding to the grammar are acquired. In the following, a general description is presented as to how the system acquires the unknown symbols from the dialogue example.

Assume that A and B are making a dialogue, as in Table 1.[†] In this case, B2 has a certain causal relation to A1, as well as B4 to A3. Then, the rules are estimated as in Table 2, which produces output B in response to input A. Among the rules, the common part is examined. When there appears a common part, as is shown underlined, it is estimated that the different part contains a strong causal relation. In addition, the sentence generation rule is acquired, with the substitution into the different part.

In this paper, A and B in Table 1 are considered to represent the user and the system, respectively. By using actual dialogue examples, it becomes possible to realize a dialogue matched to the situation and to acquire the inherent dialogue rules. On the other hand, the topic may easily change, and the response must be generated according to

various situations. With this viewpoint, GA is introduced corresponding to the human trial-and-error process in the acquisition of language. GA generates automatically the rules as well as the selection, by mutation. The detailed method is given in the section on procedure.

At the present stage of the study, miscellaneous talk (non-task-oriented dialogue) is used, and the processing of the general spoken dialogue is difficult. It is planned as a future study to handle task-oriented dialogue, by imposing constraints or other elaborations.

3.   Processing

3.1.   Outline

Figure 1 shows the flow of process in the proposed method. As the first step, the utterance of the user is converted to the text format through a commercial speech

RSG:Response Statement Generation
SSG:Surface Statement Generation

Fig. 1.   Flow of the process.

[†]The reason for using Greek characters for the character string is that the person may think that he has already experienced the string in Table 1, when the computer system handles the unknown symbol.
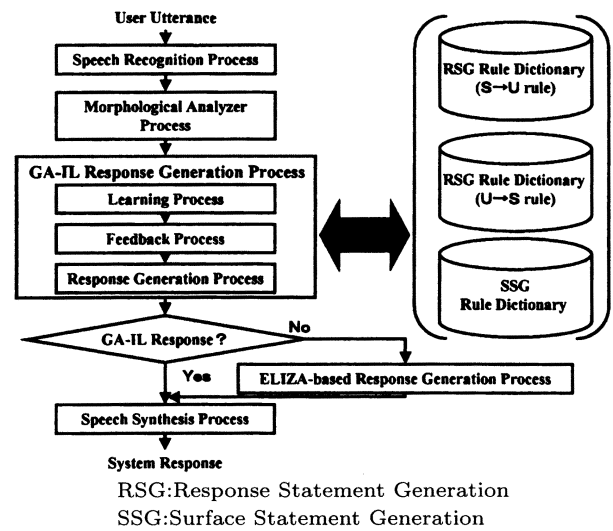
recognition device [11]. After conversion, morphological analysis is applied to the result of speech recognition for the user, through the morphological analysis tool JUMAN [12]. For the result of morphological analysis, the word classification information (part-of-speech information) and the segmentations are determined. The segmented word is utilized as the internal representation format for GA-IL rule.

GA-ILSD responds to the utterance of the user, by either GA-IL response or ELIZA-type response. GA-IL response generator acquires the rule from the dialogue example, using the learning unit. The feedback unit examines the adaptability of the acquired rule. Then, the response sentence generator generates the response statement, using the acquired rules. When GA-IL response generator cannot generate the response, ELIZA-type response generator generates ELIZA-type response. The commercial speech synthesis tool [13] is used to generate the speech output from the generated response statement. The system is programmed using C language, and is implemented on Windows98.

3.2. GA-IL response generator

3.2.1. Learning unit

(a) Rule acquisition

Figure 2 illustrates the method of rule acquisition from dialogue examples. There are two kinds of rules: the response statement generation rule and the surface statement generation rule. The response statement rule is acquired from the dialogue example between the user and the system, as a pair of independent word sequences. The
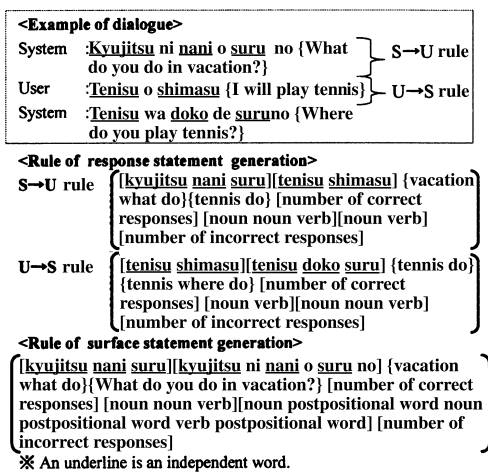

Fig. 2. Example of rules acquired from dialogue examples.

generation of the response statement using the acquired rules is described in Section 3.2.2.

In the morphological analysis of Japanese, the sentence is often divided into independent words and dependent words. A dependent word, such as auxiliary verbs and postpositional words, does not have a meaning until it is combined with an independent word. It does not have a meaning by itself. The response statement generation rule uses only independent words, which have their own meaning, in order to cope with incorrect speech recognition and to realize processing that is robust against subtle variation of expression.

Consider the case of the system uttering "Kyujitsu ni nani o suruno" {*What do you do in vacation?*}, and the user utters "Tenisu o shimasu" {*I will play tennis*}. Then, GA-ILSD acquires the response statement generation rule, which is composed of the independent word sequence pair "kyujitsu nani suru : tenisu shimasu" {*vacation what do : tennis do*} contained in the system response and the user utterance, as well as the word classification information, the number of correct responses, and the number of incorrect responses. The word classification information is described in detail in (b) below. The rule matched to the user can be acquired from the user utterance for the system response. If, to the user utterance "tenisu o shimasu" {*I will play tennis*}, the system responds "tenisu wa doko de suruno" {*Where do you play tennis?*}, it implies that the system is already empowered to respond.

In order to acquire further a new rule from this dialogue example, the pair of user → system (U → S) must be acquired. The response statement generation rule that is acquired from the system response to the user utterance is written as "U → S rule." The response statement generation rule that is acquired from the user utterance to the system response is written as "S → U rule."

The surface statement generation rule attaches dependent words to the independent words of the response statement generation rule, and constructs a natural response sentence. In the dialogue example of Fig. 2, the word classification information, the number of correct responses, and the number of incorrect responses, are acquired, in addition to "kyujitsu nani suru : kyujitsu ni nani o suru" {*vacation do what : in vacation do you what?*}.

(b) Word classification information

The result of speech recognition for the user utterance is divided into independent words and dependent words. Dependent words are "postpositional word," "space," "auxiliary verb," "decision word," "comma," "period," "parenthesis," "undefined word," and "conjunction." Independent words are "auxiliary numeral," "exclamation," "adjective," "verb," "symbol," "adverb," "noun," "indicative" and "participial adjective."

The morphological analysis tool JUMAN includes "human name," "numeral," "geographical name," "katakana character," "organization name," "noun-like noun auxiliary numeral," "others," and "alphabet." Those, however, are handled as nouns in this study, for the following reason: The adaptation to a small amount of data is considered in this study, and such a state should be avoided where the word classification is too detailed and the database is made sparse to prevent learning.

Additional processing is used as follows. When nouns continue, such as "research development," "research" and "development" are connected. The "prefix" and the "postfix" are connected to the independent word to form a new independent word. The word classification information is acquired at the recognition of user utterance and at the system response. The use of word classification information is explained in the description of each processing.

(c) Inductive learning

The processing of inductive learning is shown below.

(1) Selection of learning object rule

From the rule dictionary, a rule pair is selected, which has a common part on both sides, as well as one different part.

(2) Generation of rule

The rule is generated, extracting only the different part.

The rule is generated, replacing the different part by a variable.

(3) The above two processes are iterated until no other rule can be acquired.

In inductive learning, the rule pair with only one different part is considered as the object for learning. This is done because when two different parts exist, it is impossible to identify the causal relation for the different part. Consider the case where four response statement generation rules of (a) exist, as shown in Fig. 3. The number of correct responses and the number of incorrect responses are omitted. Comparing rules in "gakko nani suru : benkyo shimasu" {*school what do : study do*}, as well as "kyujitsu nani suru : tenisu shimasu" {*vacation what do : tennis do*}, there is a different part on the left-hand side and on the right-hand side. The rules extracting the pair of different parts "gakko : benkyo" {*school : study*} and "kyujitsu : tenisu" {*vacation : tennis*}, as well as the rule replacing the different part by the variable "@0 nani suru : @0 shimasu" {*@0 what do : @0 do*}, are acquired. The same procedure is applied to the other rules in (a), and the rules in (b) are acquired. From the bottom two rules in (b), the rules in (c) are further acquired.

(d) Genetic algorithm

GA is a technique that was introduced by J. Holland. It is an algorithm whose idea is derived from the principle of biological evolution [14]. In this method, each rule is considered as the individual (chromosome), and the words are considered to correspond to the genes. By applying GA, the rule with a low adaptability is eliminated, and new rules are acquired through crossover and mutation.
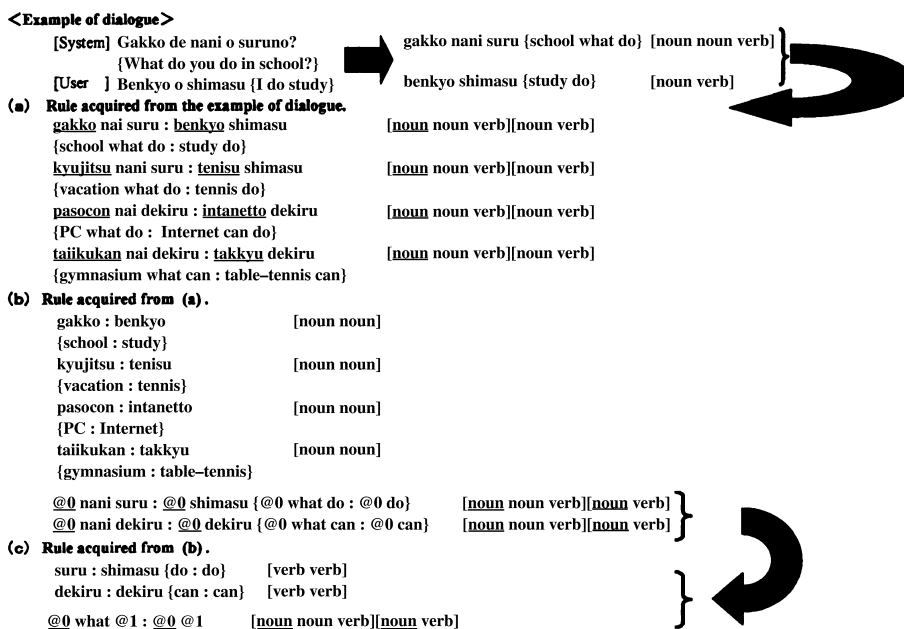

Fig. 3.　Example of rules acquired by inductive learning.

(1) Selection

Selection is the procedure by which the individual (chromosome) with a lower adaptability is deleted. The feedback unit described in Section 3.2.2 executes the selection. The adaptability is the evaluation of each individual to the environment, and is used in the selection process and rule selection. The adaptability of the rule is calculated using the following evaluation function:

$$\text{Adaptability (\%)} = \frac{\text{Number of correct responses}}{\text{Number of incorrect responses}} \times 100 \quad (1)$$

(2) Crossover

Crossover is the procedure by which two parent chromosomes are combined to form the child chromosome. The object rules are the rules with the "common word classification sequence" (composed of two or more words) on both sides. In order to avoid destruction of the sentence structure, the process is applied to the common part of the word classification sequences. For the common word classification sequence, the mask is defined at random (0: crossover is not applied; 1: applied), and uniform crossover is applied. The crossover is applied as shown in Table 3, to the object rules in the range of the common part of the word classification sequence.

(3) Mutation

Mutation is the procedure by which the gene is modified with the specified probability. In this study, the mutation is applied at random with a mutation rate of 2%.

(4) Generation

The procedure is applied for 10 generations. If a new rule cannot be generated, the procedure is terminated even if 10 generations are not covered.

3.2.2. Feedback unit

The system utilizes the error keyword, and evaluates the rule that generated the previous system response. The error keyword is the word that guides the system in the automatic evaluation of the rule. The procedure for determining the error keyword is described in Section 3.4.1. When the user utterance contains the error keyword, the number of incorrect responses is incremented by 1, in the rule used in the previous response. When the error response is not contained, the number of correct responses is incremented by 1. The system calculates the adaptability, using the number of correct responses and the number of incorrect responses. The condition for deletion is that the evaluation function of Eq. (1) not exceed 75%.

3.2.3. Response statement generator

The system compares the result of speech recognition for the user utterance and the left-hand side of each rule in the response statement generation rule dictionary, to extract the response candidate rules. If there exists a response statement generation rule satisfying the condition, the response statement is generated from the right-hand side of the rule. The response statement is generated by adding the dependent words to the independent words on the right-hand side of the response statement generation rule, using the surface statement generation rule. The priority in determining the rule is as follows.

(1) The rule with the highest match rate, being not less than 65%, is selected.

(2) If more than one rule with the same match rate exists, the rule with the higher adaptability is selected.

(3) If more than one rule with the same adaptability exist, the latest acquired rule is selected.

The match rate is calculated as follows. The number of independent words is compared between the independent word sequence in the result of speech recognition for the user utterance and the independent word sequence in the rule dictionary. The match rate is calculated, using the larger number of independent words as the denominator, and the number of matched words as the numerator.

Table 3.   Examples of rules acquired by crossover

	Independent word sequence on the left-hand side			Independent word sequence on the right-hand side		
Parent 1	gakko	nani suru	{school what do}		benkyo shimasu	{study do}
	noun	noun	verb		noun	verb
Parent 2	konshu skejuru {this week schedule}			getsuyobi shucho aru {Monday travel is}		
	noun	noun		noun	noun	verb
Mask	0	1		1	0	
Child 1	gakko skejuru suru {school schedule do}			shucho shimasu	{travel do}	
Child 2	konshu nani	{this week what}		getsuyobi benkyo aru {Monday study is}		

Table 4 shows examples of response statement generation. When the user utters "Pasokon de nani o suruno" {*What do you do with PC?*}, the system segments the sentence into independent words as "pasokon nani suru" {*PC what do*}. Using the response statement generation rule, the rule matched to the independent word of segmented user utterance is extracted as in (a). At this stage, the rule with variable is also extracted. If there exists a variable as in the third case of (a), the second rule is substituted into the variable, so that the word classification information is matched. Then, the fourth rule is constructed.

According to the priority order, the rule "pasokon nani suru : intanetto shimasu" {*PC what do : Internet do*} is selected, from the rules with match rate exceeding 65% with the user utterance. The surface statement generation rule agreeing with the right-hand side is selected. If two or more such rules exist as in (c), the adaptability is used for the decision. In this case, "Intanetto shimasu" {*I will play Internet*} is decided as the response of the system.

3.3. ELIZA-type response generator

When the system cannot provide GA-IL response, ELIZA-type response is generated. The purpose of ELIZA-type response generator is to continue the dialogue, and to acquire the dialogue example. In this process, the system reads-in the result for the user utterance, and tries to search the prespecified keyword. The determination of the keyword is described in Section 3.4.1.

The categories of generation rule in ELIZA-type response generator are as follows.

1. Greeting
2. Question
3. Error
4. Weather
5. PC
6. Eating and drinking
7. School
8. State
9. Sport

The precise response generation process for greeting is described as follows, using the rules of Table 5. @keyword is the keyword contained in the utterance. @random_adj and @adj represent adjectives. The rule related to greeting is prepared in @howareyou1st. When the user utters "Ohayo" {Good morning}, the response statement is generated using the greeting rules, since "ohayo" is already stored in the keywords of greeting category. Then, "@ keyword (ohayo)" {*good morning*}, "@ random_adj saikinno" {*recent*}, and "@ ad ii" {*good*}" are substituted into "@ keyword anatano @ random_adj @ adj koto o hanashite kuremasenka" {your}, {tell me your case}. Then, the response is constructed as "Ohayo anatano saikinno ii koto o hanashite kuremasenka" {*Good morning, tell me what was good in your case recently*}. If the user utters "Ohayo" {*good morning*} again, the response is generated by howareyou1st rule as "Choshi wa do desuka"

Table 4. Examples of generation of the response sentence

User: Pasocon de nani dekiruno {what do you do with PC?} → pasocon nani suru {PC what do}					
(a) Extraction of response statement generation rule					
1	Pasocon nani dekiru {PC what can}	intanetto dekiru {Internet can}	noun noun verb	noun verb	1 0
2	pasocon {PC}	intranet {Internet}	noun	noun	2 0
3	@0 nani suru {@0 what do}	@0 shimasu {do}	noun noun verb	noun verb	1 1
4	pasocon nani suru {PC what do}	intanetto shimasu {Internet do}	noun noun verb	noun verb	0 0
(b) Selection of response statement generation rule					
1	Pasocon nani dekiru {PC what can}	intanetto dekiru {Internet can}	noun noun verb	noun verb	1 0
2	pasocon nani suru {PC what do}	intanetto shimasu {Internet do}	noun noun verb	noun verb	0 0
(c) Selection of surface statement generation rule					
1	intanetto shimasu {Internet do}	Intanetto o shimasu {Internet do}	noun verb	noun postpositional word verb	2 0
2	intanetto shimasu {Internet do}	Intanetto ni shimasu {Internet do}	noun verb	noun postpositional word verb	1 2
System: Intanetto o shimasu {I do Internet}					

※ 4 of response statement generation rule is the result of substituting rule 2 into rule 3.

Table 5.   Examples of ELIZA-based rules

Keyword	Category
ohayo {good morning}	greeting
konnichiwa {how do you do}	greeting

Response rule of greeting category	
Rule 1	@keyword anatano {your} @ random_adj
	@adj koto o hanashite kuremasenka {tell me your case}
Rule 2	@howareyoulst

random_adj rule	
Rule 1	saikinno {recent}
Rule 2	kyomibukai {interesting}

adj rule	
Rule 1	ii {nice}
Rule 2	omo shiroi {interesting}

howareyoulst rule	
Rule 1	choshi wa do desuka {how does it go with you?}
Rule 2	genki desu ka {how is your health?}

{*how does it go with you?*}. Then, the same response is avoided.

There are other conversion processes. If the utterance does not contain two words and does not contain a keyword, the response "Wakariyasuku hanashite kudasai" {*Tell me in an easily understandable way*} is generated to invite a further utterance. There is prepared a conversion process to exchange pronouns "anata" {*you*} and "watashi" {*I*}. The word classification information acquired by JUMAN is utilized, in quoting the user utterance.

3.4.   Preliminary experiment

3.4.1.   Keyword

ELIZA-type response generator extracts the keyword and the error keyword as follows. The operation essentially follows the operation for ELIZA keyword. Since ELIZA contains many keywords, such as "jisatsu" {*suicide*} and "doraggu" {*drug*}, which are difficult to utilize in miscellaneous talk, the keywords are selected and corrected by a preliminary experiment. In the preliminary experiment, four males and three females engaged in dialogue for 30 to 50 turns with ELIZA-type system. Since dialogue processing is not the purpose of the ELIZA-type system, keywords

of the minimum necessary number are prepared. As the decision criterion for constructing ELIZA-type rule, the occurrence frequency and the dialogue content are emphasized. JUMAN is used in the word-wise segmentation.

The word frequency is examined only for the noun. The words that appeared two or more times and can be used, as well as the words that are related, are defined as the keywords. For example, since "sakka" {*soccer*} (two times) appeared and is defined as a keyword, "basuketto boru" {*basketball*} and "yakyu" {*baseball*} are included in the sports-related keywords. The same principle is applied to the error keywords, to be defined, depending on the frequency of occurrence. As a result, 196 ELIZA-type keywords, as well as 40 error keywords, are defined and used in the feedback unit.

3.4.2.   Match rate

The rule to be used by the system in its response is determined as follows. In determining the rule to be used, the match rate is emphasized most highly in the priority order. The system compares the "independent word sequence in the result of speech recognition for the user" and the "independent word sequence pn the left-hand side of the response statement generation rule," in terms of the match rate. The dialogue is adjusted by considering the match rate.

From the viewpoints of the spoken dialogue and the advantages of the proposed method, however, it is not necessarily true that 100% (exact match) is best. The reasons are as follows.

Table 6.   Results of the preliminary experiment

Speech recognition rate			
Match rate	○	△	×
50%	72.3%	4.0%	23.8%
65%	52.5%	6.5%	41.0%
80%	63.0%	6.5%	30.5%
100%	40.1%	8.4%	51.5%

Response rate						
	Correct response rate (%)		Almost correct response rate (%)		Incorrect response rate (%)	
Match rate	E	G	E	G	E	G
50%	10.9	25.2	3.5	11.9	12.4	36.1
65%	22.0	15.5	12.0	11.0	23.5	16.0
80%	32.0	7.0	19.0	7.5	25.0	9.5
100%	19.0	5.5	16.5	4.5	42.0	12.5

E: ELIZA–type G: GA-ILSD
○ : correct speech recognition rate
△ : correct speech recognition rate except for dependent word
× : incorrect speech recognition rate

Table 7. Total of the GA-IL response

Match rate	GA-IL response			Total
	Number of correct responses	Number of almost correct responses	Number of incorrect responses	
50%	51	24	73	148
65%	31	22	32	85
80%	14	15	19	48
100%	11	9	25	45

- The result of incorrect speech recognition is included in the rule.
- There can exist an effective response other than the one with 100% match.
- The opportunity of utilizing other acquired rules is lost.
- Acquire effective rules becomes difficult.

The experiment was conducted by setting the match rate to 50, 65, 80, and 100%. The already trained dictionary is used as the initial condition. The user executes 200 turns from the initial state, and the result is compared. Table 6 shows the results of speech recognition and the match rate for this experiment. Since the speech recognition rate varies, it is difficult to handle the data on a uniform basis. In order to realize a fair comparison, the "ratio of correct response when GA-IL response is used" is compared. Table 7 shows the "ratio of GA-IL response."

It is seen that GA-IL response increases for a match rate of 50%. In this case, new rules are acquired efficiently, but the ratio of the incorrect responses is the highest, being 49.3% (73/148). When the match rate is set high, such as 80 or 100%, the incorrect responses are fewer, but the correct responses are even fewer. It is verified that the ratio of the correct responses is the highest for 65%. In other words, a match rate of 65% is suited to handle the miscellaneous talk.

4. Evaluation Experiment

4.1. Comparison with ELIZA-type system

4.1.1. Evaluation method

The spoken dialogue is used in the evaluation experiment. The reason is that the "spoken language" we consider is not an extension of the "written language." The "written language" contains less redundant words, and is relatively easy to handle based on the grammar. In contrast, the "spoken language" starts with the noise of speech recognition, and contains inevitably the situation that cannot be avoided even if the user is aware, such as interjection,

reversal of position, restatement, and particular order of words inherent to the spoken language. In such a case, the learning based on the surface expression will be necessary as in the proposed method, which is the purpose of examination in this evaluation experiment using the spoken language.

The problem in the study of miscellaneous talk through spoken language is the difficulty of defining the object topic. The following items are considered as problems:

1. What miscellaneous talks should be handled.
2. How the result should be evaluated.

It should carefully be considered at which stage the miscellaneous talk is presently located in the dialogue. The dialogue can be divided into task-oriented dialogue and dialogue without a clear intention. Miscellaneous talk belongs to the latter. Then, the evaluation based on the achievement of the task or goal is impossible.

We define miscellaneous talk to have the following features:

1. The utterance is flexible.
2. A new topic can be developed.
3. The information can be acquired from the dialogue, to be utilized in a later topic.
4. The dialogue can be continued even if no knowledge is available.

The response can be evaluated if the examinee understands the above conditions and decides whether or not the response is natural and adequate. This is a subjective evaluation. Thus, miscellaneous talk should be handled differently from the case where the correct response can be defined clearly on an objective basis. Consequently, the decision criterion is defined, and the examinee gives a subjective decision for each response.

The examinee gives the evaluation of the system response as one of three, namely, the correct response, the almost correct response, and the incorrect response.

(1) Semantically correct response

The response is correct if the representation is natural, and is almost correct if the representation is unnatural.

(2) Semantically incorrect response

Such a response is an incorrect response.

Following this criterion, the examinee evaluates the system response given as the speech output. Table 8 is an example of response evaluation by an examinee.

Table 8. Example of response evaluation

	Correct response
USR	Omoshiroi koto o hanashimashoka {Shall I tell you something interesting?}
SYS	Omoshiroi koto wa iidesune {It's nice to hear something interesting}
	Almost correct response
USR	Kaeritai desune {I would like to go back}
SYS	Uchino desune {It's ours}
	Incorrect response
USR	Omoshiroi koto wa suki desuyo {I like interesting things}
SYS	Omoshiroi koto wa kotonarue {Interesting thing is different}

The result of speech recognition is shown as it is.

4.1.2. Method of experiment

In this experiment, the system in which GA-IL response generator is deleted and only ELIZA-type response is generated, and the system using GA-ILSD are compared. For this purpose, each examinee executes the evaluation experiments for the two systems. The examinee executed the dialogue for approximately 1000 turns.

The evaluation of the GA-ILSD response can be any of the following six:

① Correct response generated by ELIZA-type response generator

② Correct response generated by GA-IL response generator

③ Almost correct response generated by ELIZA-type response generator

④ Almost correct response generated by GA-IL response generator

⑤ Incorrect response generated by ELIZA-type response generator

⑥ Incorrect response generated by GA-IL response generator

The evaluation of the response from ELIZA-type system is one of ①, ③, and ⑤, since GA-IL response is not executed.

4.1.3. Initial condition

If GA-ILSD has a dictionary of sufficient power, the system can generate the response using only GA-IL response generator, without using ELIZA-type response generator. It is intended, however, in this paper to verify the effectiveness of the learning by GA-ILSD. Consequently, the initial condition is set as fixed. The experiment is started
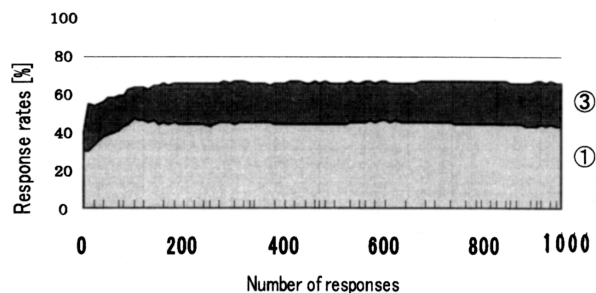

Fig. 4. Effective response rates of ELIZA-based system.

from the state where the GA-IL response generator dictionary is empty.

4.1.4. Result of experiment

Figure 4 shows the effective response rate (① and ③) of ELIZA-type system, Fig. 5 the effective response rate (①, ②, ③, and ④) of GA-ILSD, and Fig. 6 the ratio between GA-IL response and ELIZA-type response for each 50 responses in the experiment using GA-ILSD.

4.1.5. Discussion

By applying GA-IL, the ratio of correct and almost correct responses combined is improved from 66.3% to 76.1%. As seen in Table 9, although the speech recognition rate in the experiment using GA-ILSD is lower than in the experiment using ELIZA-type system, the effective response ratio is higher. We verified that the proposed method is effective in handling miscellaneous talk.

In this experiment, however, only one examinee tried the dialogue, and there is no evaluation concerning the individual difference. Consequently, we evaluated individual differences in the use of GA-ILSD, with multiple examinees.
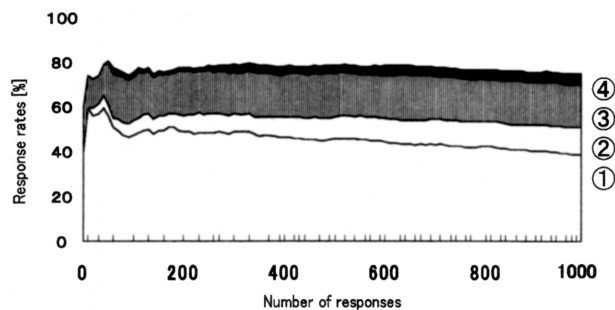

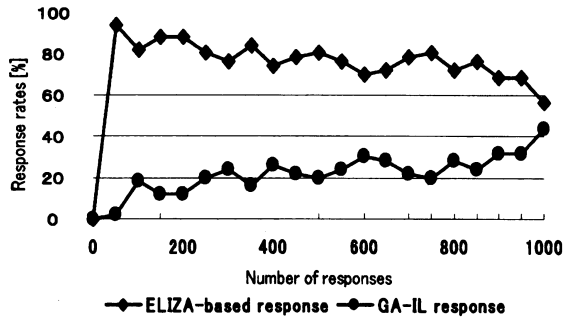Fig. 5. Effective response rates of GA-ILSD system.

Fig. 6.   Rates of the ELIZA-based response and the GA-IL response.

Table 10.   Results of multiple-user experiment

Speech recognition rate					
	A	B	C	D	E
○	53.5%	51.7%	48.9%	37.5%	61.2%
△	3.4%	4.6%	10.1%	13.1%	7.1%
×	43.0%	43.6%	40.9%	49.4%	31.5%
Response rate					
	A	B	C	D	E
①	19.3%	21.4%	36.0%	22.7%	14.1%
②	13.6%	14.1%	1.4%	7.9%	1.4%
③	20.6%	16.4%	25.9%	14.5%	38.1%
④	3.3%	6.5%	2.8%	7.2%	9.4%
⑤	31.7%	26.0%	25.5%	28.8%	25.2%
⑥	11.5%	15.7%	8.3%	18.9%	11.7%

4.2.   Evaluation of GA-ILSD by multiple examinees

The method of evaluation and the initial condition are the same as in Section 4.1.

4.2.1.   Method of experiment

The examinees are five students of the engineering school who had no experience in using the spoken dialogue system. Each examinee continued the dialogue up to 500 turns. The examinees are denoted A, B, C, D, and E, respectively.

4.2.2.   Result of experiment and discussions

Table 10 shows the results of speech recognition, as well as the response rate, of each examinee in this experiment. The effective response (② + ④) of GA-IL is more than 10%, except for examinee C. Since the experiment is started from the empty state of the dictionary, an effective response of GA-IL above 10% for 500 turns is a high value. It is seen that the learning is executed frequently in the latter half of the experiment, which increases GA-IL response.

This experiment is started from the state where the rule dictionary of GA-IL is empty, in order to evaluate rule acquisition from the actual dialogue examples. If an already

Table 9.   Results of comparative experiment

Speech recognition rate			
	○	△	×
ELIZA–type	61.6 %	8.1 %	30.3 %
GA-ILSD	57.0%	9.5%	33.5%
Response rate						
	①	②	③	④	⑤	⑥
ELIZA–type	43.3%	-	23.0%	-	33.7%	-
GA-ILSD	39.3%	12.0%	19.5%	5.3%	18.2%	5.6%

trained dictionary is given as the initial state, the effective GA-IL response will increase beyond the case where the experiment is started from the empty dictionary, and the response matched to the local dialogue can be realized.

In the case of examinee C, the effective response of GA-IL (② + ④) is lower. This seems due to the small number of independent words used in the dialogue. As seen in Table 11, examinee C used a larger number of independent words than other examinees. This includes the situation of fewer utterances using the same independent word iteratively. Unless similar expressions are iterated, it is difficult for the system to find a matched rule. Also, when the acquired rules are compared, it seems that progress of the learning is prevented due to the lack of common parts. Thus, examinee C is still learning. Excluding examinee C, it is seen that effective responses are generated for multiple examinees using GA-ILSD.

This does not imply that an effective GA-IL response is not generated for examinee C. It is seen that a dialogue, which is really a miscellaneous talk, is realized by including GA-IL response, as shown in Table 12.[†] In Table 12, "Kankei tsuzukete kudasai sugoi to nani" {*continue relation superb and what*} cannot be considered as a good response, but is judged an almost correct answer by examinee C. To this system response, examinee C uttered "Mata hitori dattai surundesuyo" {*Again a person will leave*}. The content of this utterance can be considered as an utterance to "Nani ga sugoinoka" {*What is superb*}. In the spoken dialogue, there is a large tolerance as to the occurrence order of words, since they are spoken. The content is understood based on the spoken language, and the decision may be different from that based on the text.

[†]The result of speech recognition is shown. The incorrect result of recognition is described, together with the correct user utterance in parentheses.

Table 11. Number of independent words in multiple-user experiment

Number of used independent words					
	A	B	C	D	E
Number of independent words	682	547	1054	549	651

In human-to-human miscellaneous talk, it is not natural that the third person decides whether the talk, for which the talkers are satisfied, is correct or incorrect. In contrast, in the experiment on human-to-computer talk, the human must make the decision. The evaluation experiment was conducted with multiple examinees, but the reliability for each evaluation does not seem to be very high. Consequently, the response in the experiment is evaluated by multiple examinees in the next section.

4.3. Evaluation of comparison experiment by multiple examinees

4.3.1. Method of evaluation and method of experiment

The data for evaluation are extracted from the result of comparison experiments for ELIZA-type system and GA-ILSD system. In order to achieve a fair comparison, the extraction condition is set so that data for 20 turns are extracted from the head for each talk of 100 turns. It is intended that the experiment should be as close to the actual environment as possible, and the response by the speech synthesizer should be examined. The result of speech recognition for the user utterance and the response statement generated by the system in the experiment are speech synthesized, and the results are evaluated.

Since the result of speech synthesis for the user utterance may contain an error, it is examined and corrected if necessary. There are four examinees in the evaluation: three undergraduate students and one graduate student in a university of science. They are designated F, G, H, and I, respectively. The evaluation criterion is the same as in Section 4.1.1.

4.3.2. Result of experiment and discussions

Table 13 shows the decision of each examinee, trying the dialogue with the system, as well as the match rate as decided by four examinees. By the match rate is meant the ratio of the number of matches between the "evaluation by the examinee who tried the comparison experiment" and the "evaluation by four examinees when listening to the dialogue in the comparison experiment" for the entire response.

As a result of evaluation, the match rate is low. One of the reasons for this is that there are fluctuations in the decision, not only by the examinee, executing the comparison experiment, but also by all four examinees. Consequently, the result is compared to the result of decision, which is the majority decision by four examinees. The result is 44.2% (ELIZA-type) and 66.5% (GA-ILSD), which are not high, although they are improvements.

The reason for this seems to be the difference in the experimental environment between the "evaluation of the concerned person" and the "evaluation by the third person observing the result of dialogue." It should also be noted that there is a difference between the decisions for ELIZA-type system and GA-ILSD system. This is considered to be due to the fact that the response of GA-ILSD system is easier to decide, but it is difficult to decide whether or not the response of ELIZA-type system is correct.

Table 12. Example of dialogue by subject C

Examinee C (Only English translation)		
USR	1	Do you know "Morning girls"?
SYS	2	Do you think "Morning girls" are superb?
USR	3	I think "Morning girls" are superb.
SYS	4	Please continue relation, superb and what?
USR	5	Again another girl is going to retire.
SYS	6	I do not quite remember, retire or not retire.
USR	7	"Morning girls" will be again seven.
SYS	8	Please tell me more what you think.
USR	9	I feel lonely.

Table 13. Match rates

ELIZA-type			
Examinee	F	G	H	I
Match rate	41.2%	38.1%	36.6%	33.1%
GA-ILSD			
Examinee	F	G	H	I
Match rate	52.5%	65.0%	59.5%	52.0%

In the whole "extracted data," the ratio of the correct and almost correct responses combined is 66.3% (ELIZA-type) and 76.1% (GA-ILSD). The ratio is 66.3% (ELIZA-type) and 80.0% (GA-ILSD) for the "result of 100 dialogues." Those results do not differ much from comparison experiment results. Consequently, the response is evaluated based on the above decision results. Table 14 shows the response rate.

As the average for four examinees, the sum of the correct responses and almost correct responses is 83.6% for GA-ILSD and 76.4% for ELIZA-type. The result is an improvement of 7.2%, and the result for GA-ILSD is improved as a whole. The ratio of correct responses tends to increase in the proposed system. Except for examinee H, the ratio is 80.8% for ELIZA-type and 83.2% for GA-ILSD. In other words, there is still room for improvement in the evaluation method. Based on those results, it is concluded that the response is generally improved, although there is a fluctuation in the evaluation of the individual response.

4.4. Discussion

4.4.1. Error keyword

In the experiment with multiple examinees, five examinees participated and a total dialogue of 2500 turns was executed. The error keywords were extracted 269 times. An error keyword is a word that is prespecified in constructing the system. Table 15 shows the number of occurrences, the number of correct uses, and the correct use rate, for the keywords in the experiment with multiple examinees.

Table 14.   The evaluation of the response

ELIZA-type

Examinee	F	G	H	I
Correct	44.2%	80.4%	73.4%	42.7%	39.2%
Almost correct	22.1%	14.1%	10.1%	20.6%	25.1%
Incorrect	33.7%	5.5%	15.1%	34.7%	35.7%

GA-ILSD

Examinee	F	G	H	I
Correct	57.0%	73.5%	83.5%	64.5%	42.5%
Almost correct	23.0%	20.0%	0.5%	20.5%	29.5%
Incorrect	20.0%	6.0%	15.5%	14.5%	27.5%

Table 15.   Correct use rates of erroneous keyword

Error keyword

	A	B	C	D	E
Number of occurrences	75	98	28	47	21
Number of correct uses	72	82	17	37	17
Rate of correct usages	96.0%	83.7%	60.7%	78.7%	81.0%

As seen from the rate of correct use, when there exists an error keyword in the user utterance, the processing is generally executed correctly. A problem is produced, however, due to use of the error keyword. In the case of examinee B, the examinee noted the existence of the error keyword, and iterated the utterance containing intentionally the error keyword to the incorrect response.

An example of processing that is incorrect as a whole is as follows. When the user utters "Chigau hanashi o shimasho" {*Let's talk on a different topic*}, since "chigau" {*it is not true*} is included in the error keywords, the system handles the previous system response as an incorrect response. A remedy for such a situation may be restriction of use, discriminating the cases where each keyword can be used or cannot be used. This, however, cannot be the essential solution.

When a person decides whether the response is correct or incorrect, the information may be acquired from data other than the utterance. In order to decide on the correctness or incorrectness of the system response, information acquisition from other sensations such as tactile sensation (tapping) will work dynamically and adaptively, independently of the language features of the user.

Table 16.   Classification of erroneous ELIZA-based response

	A	B	C	D	E
○	9.6%	6.3%	2.6%	9.7%	10.9%
△	0.6%	1.7%	1.2%	0.8%	1.6%
×	21.5%	18.0%	21.7%	18.3%	12.7%
Total	31.7%	26.0%	25.5%	28.8%	25.2%

○ : correct speech recognition
△ : correct recognition except for dependent words
× : incorrect speech recognition

4.4.2. ELIZA-type response

One of the properties of ELIZA-type response is that robust processing is not realized. Table 16 shows the results when an incorrect response of ELIZA-type is classified using speech recognition. When the speech recognition is incorrect, it often happens that ELIZA-type response is an incorrect response. When an incorrect response is generated in ELIZA-type response generator, the rule is acquired considering that it is the actual dialogue example. This can produce an incorrect GA-IL response. It can happen that ELIZA-type response is incorrect even if the speech recognition is correct. The reason is that it is difficult for the generation rule in ELIZA-type response generator to handle this situation. When keywords of the same category continue, as in Table 17, the same response is iterated in a short period, and is judged as an incorrect response.

4.4.3. GA-IL response

The difference from ELIZA-type response is considered as follows. Response statement generation rules can be $S \to U$ and $U \to S$. As described in Section 3.2.1, both rules can be acquired from the actual dialogue examples. It can also happen, however, that $U \to S$ rule is generated by ELIZA-type response generator. Among GA-IL responses, the response that can be generated by ELIZA-type response generator and the response that cannot are classified as follows. By the response that can be generated is meant the response composed of the same word sequence as in the situation in which $U \to S$ rule is acquired.

Table 18 gives the results of examination where the correct and almost correct responses in GA-IL responses can be generated or cannot be generated by ELIZA-type response generator. Of 500 turns, 2.8% to 10.2% are the responses inherent to GA-ILSD, which cannot be degener-

Table 17. Repetition of the same response (only English translation)

USR	1	I hate rain.
SYS	2	Weather may change.
USR	3	I am glad if it does.
SYS	4	Please continue.
USR	5	Snow is more often recently than rain.
SYS	6	How should I do if it rains?
USR	7	It will rain tomorrow.
SYS	8	It may be good to rain.
USR	9	I don't like it rain.
SYS	10	Weather may change.

Table 18. Classification of the GA-IL response

ELIZA-type response		A	B	C	D	E
Correct GA-IL	impossible	8	34	4	14	4
Correct GA-IL	possible	67	40	3	25	3
Almost correct GA-IL	impossible	6	17	11	26	31
Almost correct GA-IL	possible	12	17	3	10	15
GA-IL effective response total		93	108	21	75	53

Table 19. Example of the representation employed by a user (only English translation)

USR	1	Yes it is.
SYS	2	What makes it so?

Table 20. Example of the response adapted to the time (only English translation)

USR	1	Soon it will be one year, this year.
SYS	2	How about Christmas?
USR	3	It will come in several weeks.

Table 21. Example of the response of the paraphrase (only English translation)

USR	1	I will drink osaki with friend (note: osake = Japanese wine, osaki = future)
SYS	2	I guess friend is a nice guy.
USR	3	Friend is a nice guy.
SYS	4	Intimate friend.

Table 22. The example of dialogue (only English translation)

U		I have a friend.
S		Who is your friend?
U		Is my intimate friend
Acquired rule	Friend who matter	I intimate friend

ated by ELIZA-type response generator. This is the result of examination of the effective responses, which are inherent to GA-ILSD.

A response example that is adapted to the actual dialogue is described next. In Table 19, the system response is generated using the representation by the user. In Table 20, a question concerning Christmas is raised. It is a response following the timing of December.

"Shinyu" {*intimate friend*} in Table 21 is not an adequate response, but examinee C judged it to be almost correct. In this case, not the correct or incorrect response, but the rule acquisition process should be emphasized. The rule concerning "Shinyu" is not given as an ELIZA-type response generation rule beforehand. It is a rule that is acquired by learning from the dialogue example.

In the dialogue example of Table 22, there is only one user utterance, which includes "Shinyu." The rule, which is acquired from this dialogue example, is "Tomodachi dare koto : watashi shinyu" {*friends who matter : I intimate friend*}. Utilizing the acquired rule, the proposed system acquires "Tomodachi mujun, iikimochi : shinyu" {*friends inconsistency nice feeling : intimate friend*}.

Since GA is used, there is also the word "Mujun" {*inconsistency*}. Due to the match rate of 75% with the input sentence, however, the system responds as "Shinyu." Since the rule is acquired from the dialogue example, "Tomodachi kimochi iiyatsu : intimate friend" {*friends nice feeling guy : intimate friend*} is acquired. If, at this stage, there is a surface statement generation rule that adds a postpositional word, as in "Shinyu nano?" {*is he intimate friend?*}, there is a possibility that the response is judged a correct response. This is a response that does not depend on ELIZA-type response statement generation rule, and is considered to be effective in rule acquisition.

The above procedure is the rule acquisition from the actual dialogue examples, and has the feature that it can adapt dynamically to the user. Since the dialogue progresses in real time, the procedure must adapt dynamically to the dialogue. This study showed the possibility of such an approach.

4.4.4. Error inherent to speech recognition

At the current stage of technology, the error of speech recognition is inevitably included. Also, since correct or incorrect is decided through speech, an error in the result of speech recognition "Kyowa ii hidane" {*Today is a good fire source*} is interpreted as {*Today is a good day*} (note: hidane = *fire source*; hi dane = *is a day*), and is decided as correct.

As in Table 23, even if an incorrect speech recognition result is contained in the acquired rule, the response may be judged a correct response. Contrary to the above, "Konnichiwa" {*How do you do*} has the same Japanese characters, and may be spoken "Kyowa" {*Today is*}. In order to avoid such an incorrect response, the match between speech recognition and speech synthesis is important.

5. Conclusion

We intended to simulate human language acquisition in a spoken dialogue system. It is intended to learn from actual dialogue examples between the system and the user. A spoken dialogue processing method is proposed, using inductive learning by the genetic algorithm. In this paper, an evaluation experiment is run using miscellaneous talk, and it is verified that the response according to the representation and the seasonal expression, which the user is actually utilizing, can be realized through the learning from the actual dialogue example.

As a result of the evaluation example, it is seen that the total ratio of the correct response and the almost correct response is 76.1%. The effectiveness of the proposed system is evaluated by comparison with ELIZA-type system. It is seen as a result that the total ratio of the correct response and the incorrect response in ELIZA-type system is 66.3%. In other words, an improvement of 9.8% is verified by applying GA-IL. This result of experiment indicates that the proposed method is effective for miscellaneous talk (non-task-oriented dialogue).

This paper considered miscellaneous talk, and task-oriented dialogue processing is projected as a future study.

Table 23. Response containing erroneous recognition result (only English translation)

SYS	1	Gradually kucho is becoming worse (note: kucho = town master, speech tone)
USR	2	I do not think quite so

REFERENCES

1. Weizenbaum J. ELIZA—A computer program for the study of natural language communication between man and machine. Commun ACM 1966;9:36–45.
2. Winograd T. Understanding natural language. Academic Press; 1972.
3. Colby KM. Human–computer conversation in a cognitive therapy program. In: Wilks Y (editor). Machine conversations. Kluwer Academic Publishers; 1999. p 9–19.

4. Ito T, Kogure S, Nakagawa S. A spoken dialogue system with cooperative response and evaluation for the system. Trans Inf Process Soc 1998;39:1248–1257.
5. Kimmi M. A computer that acquires language. Kyoritsu Shuppan; 1998.
6. Smith RW, Gordon SA. Effects of variable initiative on linguistic behavior in human–computer spoken natural language dialogue. Assoc Comput Linguistics 1997;23:141–167.
7. Stent A, Dowding J, Gawron JM, Bratt EO, Moore R. The command talk spoken dialogue system. Annual Meeting of the Association for Computational Linguistics 1999;37:183–190.
8. Okada M. Science of unintentional action. bit 1998;30(12). Kyoritsu Shuppan.
9. Kato M, Tani Y, Kirigoe T, Sugata S. A study of ELIZA-type 3DCG conversation robot tea friend. 37th Natl Conv, Inf Process Soc, 6C-3, 1999.
10. Echizenya H, Araki K, Momouchi Y, Tochinai K. Application of genetic algorithm to machine translation technique in inductive learning based on actual example and its evaluation. Trans Inf Process Soc 1996;37:1565–1579.
11. Nishimura M, Ito N, Yamazaki K. Large vocabulary continuous speech recognition of Japanese with word as recognition unit. Trans Inf Process Soc 1999;40:1395–1403.
12. Kurohashi S, Nagao M. Japanese morphological analysis system JUMAN, version 3.61, 1999.
13. Saito T. Automatic construction of synthesis unit inventories using a reference unit inventory. IPSJ Tech Rep, SLP15-2, 1997.
14. Goldberg DE. Genetic algorithms in search, optimization, and machine learning. Addison–Wesley; 1989.

AUTHORS (from left to right)

Yasutomo Kimura (student member) graduated from Hokkai-Gakuen University in 1996, completed the master's program in 2001, and is now in the second half of the doctoral program. His research interest is natural language processing.

Kenji Araki (member) graduated from Hokkaido University in 1982 and completed the doctoral program in 1988. He was with Hokkai-Gakuen University from 1988 to 1998. He has been an associate professor at Hokkaido University since 1998. His research interests are machine learning and machine translation of natural language. He holds a D.Eng. degree, and is a member of Inf Process Soc., Jpn. Soc. Cogn. Sci., Soc. Artif. Intell., Lang. Process. Soc., IEEE, ACL, and AAAI.

Yoshio Momouchi (member) graduated from Hokkaido University in 1965, completed the master's program in 1967, and joined Hitachi Ltd. He fulfilled credit requirements and retired from the doctoral program at Hokkaido University in 1972. He has been a professor at Hokkai-Gakuen University since 1998. His research interests are understanding and generation of natural language. He holds a D.Eng. degree, and is a member of Inf. Process. Soc., Jpn. Soc. Cogn. Sci., Soc. Artif. Intell., Lang. Metrics Soc., and Lang. Process Soc.

Koji Tochinai (member) graduated from Hokkaido University in 1962 and completed the master's program in 1964. Presently, he is a professor at Hokkaido University. He is mostly engaged in research on speech information processing and natural language processing. He is a member of Inf. Process. Soc. and Acoust. Soc. Jpn.