A grammatical error detection method for dialogue-based CALL system

Oh-pyo Kweon,† Akinori Ito,‡ Motoyuki Suzuki‡‡ and Shozo Makino‡‡

This paper describes a method to detect grammatical errors from a non-native speaker’s utterance for a dialogue-based CALL (Computer Assisted Language Learning) system. For conversation exercises, several dialogue-based CALL systems were developed. However, one of the problems in conventional dialogue-based CALL systems is that a learner is usually assigned a passive role. The goal of our system is to allow a learner to compose his/her own sentences freely in a role-playing situation. One of the biggest problems in realizing the proposed system is that the learner’s utterance inevitably contains pronunciation, lexical and grammatical errors. In this paper, we focus on the correction of the lexical and grammatical errors. To correct these errors, we propose two methods to detect lexical/grammatical errors in an utterance. The conventional methods are to write a grammar that accepts the errors manually. The proposed methods 1 and 2 use the ‘error rules’ that are independent of the recognition grammar. The method 1 uses only correct system grammar and extends the recognition results using the ‘error rules’. The method 2 uses a general grammar (which does not consider the relationship between verb, particle and each noun) to recognize the learner’s utterance and check acceptance of each N-best result and searches the learner’s utterance. The grammar error detection experiment proved that the method 2 performs as well as the conventional method.

KeyWords: Grammar exercise, Detecting method of grammar error, Conversation exercise, CALL

1 INTRODUCTION

1.1 CALL system

With the advent of globalization, more and more people are studying foreign languages. Foreigners must learn correct pronunciation, vocabulary and grammar to communicate effectively with native speakers. It is ideal if a foreign language learner can be taught on a one-to-one basis with trained language instructors. Active usage of speech in such environment prepares the learner to participate effectively in ‘real conversation’. Unfortunately, this is actually difficult for most students due to economic reasons. In reality, most learners attend

† LG Electronics Inc., Mobile Handset Research Center
‡‡ Tohoku University, Graduate School of Engineering
classes in which they have to share their teacher with other learners. This reduces the amount of time each learner spends in producing foreign language speech. Conventional self-study methods such as tape recorder, VCR language material, and composition textbook force the learner to identify pronunciation and grammar mistakes by himself/herself (Toki and Murata 1988). This is a difficult task for a foreign language learner. If a learner could find his/her mistakes, they would have no need to learn them. The application of speech recognition technology can solve these problems.

1.2 Dialogue-based CALL system

In the development of speech recognition technology in recent years, the expectation of a new concept CALL system has been growing. Various CALL systems have been proposed so far (Kawai, Ishida, and Hirose 2001; Kawai and Hirose 1997). However, most of the conventional systems are practice systems of pronunciation and listening such as minimal pair exercise. Practicing pronunciation and listening only is insufficient in conversing effectively. To obtain communication ability for participating in ‘real conversation’, a learner also needs to practice grammar, sentence composition, oral exercise and so on. A dialogue-based CALL system can compensate for the lack of conversation exercises.

Dialogue-based CALL systems can roughly be divided into two types: whether a learner creates a response sentence actively or not. The first ones prepare ready-made answers when the learner is asked to answer a question (Rypa 1996; Auralog; Nakagawa, Reyes, Suzuki, and Taniguchi 1997). As a result, the learner has no practice in composing his/her own utterances. Auralog, for example, has produced an appealing language teaching system that feeds the learner’s pronunciation of one of three written sentences to the recognizer. The path of the dialogue depends on the learner’s choice. Nakagawa et al. 1997 proposed an English conversation CAI (Computer Assisted Instruction) system using speech recognition techniques. The feature of this system is similar to Auralog. However, the number of written sentences for a learner includes improper answers and grammatical errors. For example, the shown answers are [1: Sightseeing, 2: I am from Japan, 3: I am here for two weeks, etc.] for the question “What is the purpose of your visit”. While it has a certain degree of realism, one of the major problems in those CALL systems is that the learner is assigned a passive role. Therefore, the learner cannot do exercises in active production of speech which prepares the learner to participate in ‘real conversation’.

The second ones let a learner create his/her own utterances in a limited way (Eskenazi 1996, 1999; Yamamoto, Tagawa, and Miyazaki 1993; Yamamoto 2002). These systems give
a learner an active rather than passive role. For example, when FLUENCY (Eskenazi 1999) asked them [When did you find it?] with a given keyword [Yesterday], a learner constructs an answer using a keyword by themselves. Because a learner uses a keyword, the answer is easily predicted, which makes the speech recognition easier. Yamamoto et al. 1993 introduced a teaching method using conversation simulation. Using these systems, a learner firstly practices the defined simulated conversation step by step. Then, using the learned knowledge, a learner does the conversation exercise with the given keywords. This system also limitedly points out wrong expression such as [This is → My name is] in telephone environment and [Good afternoon Mr. Jack → Jack]. Both kinds of systems are useful for simple grammar and conversation practices. While these conventional systems are an important step in the right direction, other more complex and ambitious CALL systems are conceivable and no doubt desirable. For example, using a CALL system, a learner can do conversation practices as they do with human teachers.

According to the relationship between the retentive faculty and the human senses (Lee 1998), a person can remember only about 10% of things if he/she just read. If one reads them with speaking, he/she can remember about 70% of them. If one does them actually, he/she can remember about 90%. Therefore, to increase a learner’s communication capability, it is ideal to use a dialogue-based CALL system that allows the learner to do conversation exercises like ‘real conversation [actual doing]’. The long-term goal of our research is to develop such a system. There are two factors that are indispensable in order to realize the ‘real conversation’. One is that the system should allow a learner to practice in composing various utterances actively. For example, when the system asks [May I help you?], they can say [I want to buying a souvenir] or [What is most famous one?], [Are there Gyutan], [I wanna Haginotsuki] and so on. This means that lots of grammar errors in a learner’s utterances are also recognized and detected. The other is that the path of the dialogue should be changed through a learner’s response utterances.

1.3 Requirements to realize the proposed system

To realize the system, we have to solve lots of problems. Because the system allows a learner to produce his/her own speech, the learner’s utterances may contain pronunciation and grammar errors as well as system recognition errors. The problems to be solved are summarized as follows:

1 Pronunciation errors
2 Recognition errors of the recognizer
3 Grammar errors

1) Grammatical errors (incorrect usage of particles, incorrect inflection of a verb, etc.)
2) Lexical errors (incorrect wording)
3) Out-of-grammar utterances

A language learner has pronunciation problems. For example, a learner may pronounce a Japanese sentence ['hitotsu kudasai'] as ['hitotsyu kudasai']. We have discussed these pronunciation problems in our previous work and proposed a mispronunciation detecting method (Kweon, Suzuki, Ito, and Makino 2003). Another problem is the poor recognition performance. The recognition performance of freely produced speech is poor, especially utterances of foreigners because usually the speed of foreigners’ speech is slow and their pronunciation is ambiguous. To solve the problem of low recognition performance, Nakagawa et al. 1997 trained the acoustic model with non-native speakers’ speech. We (Kweon, Suzuki, Ito, and Makino 2004) also trained acoustic models with non-native speakers’ speech DB (Minematsu, Nishina, and Nakagawa).

The other problems are grammar issues. There are two kinds of grammar problems. One problem involves the grammatical errors ['koreni shimasu (hitohako onegai) → ichihako onegai']. The other problem arises from utterances that contain out-of-grammar expressions. When a learner produces his/her own speech, they may produce utterances which are outside of the recognition grammar.

In this paper, we will focus on grammar errors and propose several methods to solve this problem. This paper is organized as follows. First, the overview of the proposed system is described in section 2. Then the effect of the pre-exercise is briefly shown in section 3. Next, new methods to detect lexical and grammatical errors are proposed in section 4.

2 OVERVIEW OF THE PROPOSED SYSTEM

As we mentioned in section 1.2, our final goal is to construct a system which allows a learner to practice ‘real conversations’ actively. The overview of the system is shown in Fig. 1. When using this system, the system displays the situation and a role of the learner. For example, the system displays [You are a customer in a souvenir store, and you are going to buy a gift. I am a salesclerk of the store] and says [May I help you?]. The learner says what he/she wants to do, and no explicit choices are given. Instead, the system displays the detail of the situation. For the souvenir task, the system shows the list of items the store sells.
When the learner speaks, the system records the speech and performs speech recognition using a recognition grammar. The grammar is described as a finite state automaton and is written task-by-task by hand. After recognizing the input speech, grammatical and lexical errors as well as pronunciation errors are checked. Through these error detection processes, the errors in the input are recovered. Then the dialog controller decides the next utterance of the system. The system answers the learner, and the errors in the learner’s speech are listed in the display.

Using the proposed system, a learner can practice conversations as well as sentence composition. The system checks and returns grammar errors as well as pronunciation errors. Hence, a learner can practice pronunciation, grammar, sentence composition by doing the conversation exercises.

In this work, it is assumed that the first language (L1) is Korean and the second language (L2) is Japanese.
3 PRE-EXERCISE AND OUT-OF-GRAMMAR UTTERANCES

3.1 Out-of-grammar utterances

The CALL system uses the recognition grammar to recognize a learner's speech. As the recognizer assumes that the input utterance follows the grammar, the system cannot recognize an utterance that is outside of the grammar. As the proposed CALL system does not give the learner explicit choices of the utterance, there is always some risk that the learner utters an out-of-grammar utterance. Once an out-of-grammar utterance is given, it is very difficult to treat that utterance and recover the dialog. Therefore, it is desirable that the learner does not utter out-of-grammar utterances. Then we consider how to decrease the number of out-of-grammar utterances.

3.2 Introduction of pre-exercise

To reduce out-of-grammar utterances, a pre-exercise before doing conversation exercise with system is introduced (Yamamoto et al. 1993). First, the learner does pre-exercise of the vocabulary, grammar, and typical conversation examples. Then they do real conversation exercises using the system.

We predict that a pre-exercise will induce a learner to produce his/her own speech naturally which is inside of the system grammar, but we do not know how pre-exercise reduces out-of-grammar errors. Therefore, we investigated the influence of a pre-exercise upon the number of out-of-grammar utterances through an interview experiment.

3.3 Experimental setup

We did interviews without/with a pre-exercise. 13 Korean speakers were interviewed about a souvenir task and a hotel reservation task. The Japanese abilities of the learners were classified into the 'beginner' [Qualification : Have taken about 150 lesson hours of Japanese language, have mastered about 250 Kanji] and the 'intermediate' [About 300 lesson hours, 500 Kanji] levels (graded by International Student Center of Tohoku University). Without/with a pre-exercise, the Korean speaker freely composed his/her own utterances and made conversations with a Japanese native speaker. First, the situation was explained to the learner. For example, a native Japanese speaker explained the situation like "Here is a souvenir store. I am a staff of the store. You are a customer. ". Then, the learner was directly interviewed with the native speaker without any pre-exercise. The other way was that the learner was interviewed
Table 1  Total number of sentences and interjection

<table>
<thead>
<tr>
<th>Kind</th>
<th>Without pre-exercise</th>
<th>With pre-exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num. of sentences</td>
<td>291</td>
<td>203</td>
</tr>
<tr>
<td>Num. of interjection/sentence</td>
<td>0.787</td>
<td>0.645</td>
</tr>
</tbody>
</table>

Table 2  Result of Out-of-grammar utterances

<table>
<thead>
<tr>
<th>Num. of out-of-grammar utterances</th>
<th>Without pre-exercise</th>
<th>With pre-exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14%</td>
<td>5.4%</td>
</tr>
</tbody>
</table>

with the native speaker after they did a pre-exercise with typical conversation examples and vocabulary. Because our purpose is not to investigate listening ability, the native speaker wrote the sentences if the learners had listening problems.

3.4 Results of the interview experiment

The experimental results are shown in Table 1.

With the pre-exercise, the total number of sentences decreased by about 31%. The reason is that the learners naturally used the conversation flow in the pre-exercise. The other observation is that the number of interjection is large in the ‘without pre-exercise’ condition because the learners had problems constructing sentences immediately without a pre-exercise. With the pre-exercise, the number of interjection per sentence decreased by about 18%.

Results of the out-of-grammar utterances are shown in Table 2. According to Table 2, 14% of total sentences is out-of-grammar utterances in case where pre-exercise are not conducted. However, in the case where pre-exercises are conducted, 94.6% of total sentences is inside of the system grammar. Comparing the results of ‘without pre-exercise’ to ‘with pre-exercise’, the number of the out-of-grammar utterances dramatically decreased.

4  DETECTION OF GRAMMATICAL AND LEXICAL ERRORS

4.1 Treatment of grammatical and lexical errors

As shown in the last section, the out-of-grammar utterances can be reduced by the pre-exercise. However, as shown later, the grammatical and lexical errors cannot be reduced even if the pre-exercise was introduced. Therefore, we have to develop new methods to treat these errors in the input utterance.
In this section, first we analyze the utterances in the interviews that were explained in the last section and observe what kind of errors occurred. Then, we propose two kinds of new methods to detect the grammatical and lexical errors. These methods are compared by the experiment.

4.2 Investigation of grammatical and lexical error tendency

We investigated the grammatical and lexical errors in the transcription of the interview data gathered in the last section. Tendency of grammatical and lexical errors is shown in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>Grammatical errors</th>
<th>Lexical errors</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without pre-exercise</td>
<td>0.134</td>
<td>0.089</td>
<td>0.223</td>
</tr>
<tr>
<td>With pre-exercise</td>
<td>0.113</td>
<td>0.064</td>
<td>0.177</td>
</tr>
</tbody>
</table>

The grammatical and lexical errors of learners were decreased by only 0.046 point when the pre-exercise is introduced. This result shows that the pre-exercise have little influence on the number of lexical and grammatical errors. The reason why the influence of pre-exercise was small is that the learners tried to change the verb and vocabulary in the syntax and grammar which were used in pre-exercise incorrectly. For example, after they learned the sentences [英語で書いても大丈夫ですか、これでお願いします (Can I write it down in English? This one please)] in pre-exercise, they produced [韓国語でもいいですか、これでします (Can I use Korean? I take it — the Japanese expression これで (korede) should be これに (koreni) whenお願い (onegai) is not used)].

Tendency of the grammatical errors was related to Japanese particle and verb conjugation errors. Fig. 2 shows the examples of the grammatical and lexical errors. As Fig. 2 shows, the learners were confused between the particles [で] and [に].

We classified the detected grammatical/lexical errors in the transcription and made grammar error rules. These rules consist of production rules and vocabularies. A production rule has the form of

\[ \langle \text{cat} \rangle \langle \text{part} \rangle \langle \text{opt} \rangle^* \langle \text{verb} \rangle \leftrightarrow \langle \text{cat} \rangle \langle \text{part} \rangle \langle \text{opt} \rangle^* \langle \text{verb} \rangle \]

where \( \langle \text{cat} \rangle \) is a category of nouns, \( \langle \text{part} \rangle \) is a list of one or more particles, \( \langle \text{opt} \rangle^* \) is zero or more optional elements such as a numeral, and \( \langle \text{verb} \rangle \) is a verb. The left side of \( \leftrightarrow \) is the correct expression and the right side is the erroneous expression. A vocabulary is a category...
Grammatical errors

erroneous sentence → correct sentence

食べ物 を 大丈夫 です → 食べ物 で 大丈夫 です
これ します → これ に します
それ で 2つ お願いします → それ を 2つ お願いします
値段が 違うですか → 値段が 違いますか（違うのですか）

Lexical errors

ふた 箱 → に 箱
違って ますね → 迷いますね

Fig. 2 Examples of grammatical and lexical errors

name along with nouns that belong to the category.

Fig. 3 shows the examples of the grammar error rules.

Grammar rule

<table>
<thead>
<tr>
<th>Correct grammar</th>
<th>Wrong grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>[事物：固有名詞] を・で・sil お願い ⇔ [事物：固有名詞] は・の・でも お願い</td>
<td></td>
</tr>
<tr>
<td>[手段: 名詞] で お願い ⇔ [手段: 名詞] を お願い</td>
<td></td>
</tr>
</tbody>
</table>

Vocabulary

<table>
<thead>
<tr>
<th>[事物：固有名詞]</th>
<th>[手段: 名詞]</th>
</tr>
</thead>
<tbody>
<tr>
<td>お土産</td>
<td>カード</td>
</tr>
<tr>
<td>牛タン</td>
<td>現金</td>
</tr>
<tr>
<td>ひよこ</td>
<td>...</td>
</tr>
</tbody>
</table>

Fig. 3 Examples of the grammar error rules

For example, according to Fig. 3, there can be a grammatical error that the verb [お願い] is preceded by the object [事物：固有名詞] and the particle [は, の, でも], which should be one of [で, を].
4.3 Detection methods of grammatical and lexical errors

4.3.1 Conventional systems

There are a couple of grammatical error detection systems (Atsumi; Yano, Ogata, Sakakibara, and Wakita; Hasimoto and Shimada). However, all of the systems detect grammatical errors from a learner’s text input.

These methods cannot be directly applied to our proposed system, because our proposed system must recognize a learner’s utterances which include grammatical and lexical errors. As mentioned before, a speech recognizer recognizes the input speech using the recognition grammar. This means that the recognizer never output a sentence that has grammatical errors, as long as the recognition grammar is described to accept only correct sentences. Therefore, we must develop a method to recognize the errors in the learner’s utterance.

To detect grammatical errors from a speech input, a couple of the conventional methods (Yamamoto et al. 1993; Yamamoto 2002; Nakagawa et al. 1997) were proposed. These methods extend a system grammar with grammatical errors simply by hand. If these methods only treat a few grammatical errors, it is good enough. However, as the number of grammatical errors increases, it is more and more difficult to extend a system grammar with grammatical errors (The reason is explained below).

4.3.2 Manual(Referenced) method

The simplest method to treat the grammatical/lexical errors is to extend the recognition grammar to accept a learner’s grammatical and lexical errors. An example of such grammar is shown in Fig. 4.

The advantage of this method is that the optimum system grammar can be constructed for each task. However, the disadvantage is that it is difficult to extend system grammar automatically with grammatical and lexical errors. We need to extend each grammatical error for each verb, grammar, and particle, because the correct particle is different for each verb and grammar. For example, particles 「で、を」 is correct for a verb 「お願い」 and grammar 「牛タン で お願い」, but 「で」 is a grammatical error for grammar 「牛タン で 一つ お願い」. 「で」 is also a grammatical error for a verb 「下さい」 like 「牛タン で 下さい」. For a verb 「します」, correct/wrong particles are also different from verbs 「お願い」 and 「下さい」. As a result, one has to write grammars with errors task-by-task manually.
4.3.3 Proposed method 1

The feature of this method is that the original correct grammar which does not include any grammatical errors is used. As we mentioned, the disadvantage of the manual method is that it is difficult to extend a system grammar with grammatical errors. To overcome that disadvantage, the concept of this method is to use correct grammar (compared to the manual method) to recognize a learner's utterance and N-best recognition results are extended by the grammar error rule and a learner's utterances is searched.

This method is performed in two steps. Fig. 5 shows the diagram of the method 1. The first step is that the system uses the grammar that accepts only correct sentences to recognize a learner’s utterances. N-best lists $N(i)$ where $i$ indicates $i$-th candidate of N-best results is output as a result of recognition. Then, the system extends recognition results of the first step using the grammar error rules, and generates hypotheses $N(i,j)$ that contain grammatical errors, where $j$ indicates $j$-th hypothesis of $i$-th candidate. Then, the acoustic likelihoods are calculated for all candidates. Finally, the candidate with the maximum likelihood is chosen as the recognition result.

The advantage of this method is that the recognition grammar does not need to be extended with grammatical errors for each task. One disadvantage is that recognition time is longer than the manual method, because of the likelihood calculation. The other is that the
Fig. 5 Overview of proposed method 1

recognition performance can decrease if the corresponding sentence of the real input is not recognized within the N-best list.

4.3.4 Proposed method 2

The feature of this method is that a general grammar which does not consider the relationship between verb, particle and each noun is used. As we have mentioned earlier, the disadvantage of the manual method is that it is difficult to extend grammatical errors. To overcome this disadvantage, this method uses a general grammar (which does not consider the relationship between verb, particle and each noun) to recognize a learner's utterances, whereby the N-best recognition results are calculated and the utterances of the learner is searched.

This method performs the detection in 3 steps. Fig. 6 shows the diagram of the method 2. Let the initial recognition grammar (that accepts only correct sentences) be $G$. The first step is that the system generate the general error grammar $G'$, which allows any particle to occur where a particle appears. In Fig. 6, the second part of the grammar (を, で, に, は, が, の) is extended to accept any particle, while $G$ only accepts [を] and [で] (Fig. 5).

The system recognizes the input speech using $G'$ and outputs an N-best list $N'(i)$ as the result. Then, $N'(i)$ in the N-best list is verified using the following criteria:

(1) If $N'(i)$ is accepted by $G$, $N'(i)$ is valid.
(2) If $N'(i)$ can be recovered by any of the grammar error rules and the recovered sentence ($N(i)$) is accepted by $G$, $N'(i)$ is valid.

Then the invalid candidates are removed from the N-best list, and the remaining candidate
A grammatical error detection method for dialogue-based CALL system

with highest likelihood becomes the final recognition result. For example, if $N'(i)$ is [牛タンでします], the recovered sentence $N(i)$ is [牛タンにします] by the grammar error rules. Then, the system checks whether $N(i)$ is accepted by $G$ or not.

Fig. 6 Overview of proposed method 2

The advantage of this method is that the system grammar can be simply extended automatically. However, the disadvantage is that the recognition performance may be decreased as compared to the manual method, because the system uses the general grammar which is shown in Fig. 6.

4.3.5 Experiment

We checked the recognition results and the grammatical error detection rate for each method. We used 397 utterances (includes 140 grammatical and lexical errors: particle errors 99, verb errors 12, lexical errors 29) from the interview data by 20 Korean speakers. As we mentioned in Fig. 2, we used 29 grammar rules like "「事物: 固有名詞」を・で・sil お願い ← は・の・でも お願い 「手段: 名詞」で お願い ← 「手段: 名詞」を お願い 「事物: 固有名詞,手段: 名詞」に します ← 「事物: 固有名詞,手段: 名詞」で します".
manual method, the proposed method 1 and 2. The perplexity of the correct grammar of the souvenir task is 8.44.

The sentence recognition rate, word accuracy and grammar error detection rate are shown in Table 4 and Table 5. The recognition rate and grammar error detection rate decreased by about 8% and 33% in method 1, as compared to the manual method, because the correct system grammar could not output the candidate that corresponds to the utterances with grammatical and lexical errors.

One example of the recognition result recognized of the method 1 is shown in Fig. 7. In this example, grammatical error “に” is not correctly recognized. As shown in Fig. 7, the candidates do not correspond correctly to the utterances with grammatical errors “に お願い”. Instead of “に お願い”, the word “お肉 お願い” was recognized. Therefore, even though the grammar error rule was applied to those candidates and the most likelihood of those results is searched, the correct recognition result can not be obtained. The final recognition sentence is “牛タン お肉 お願い します” instead of “牛タン に お願い します”.

Most of the recognition errors happened in the input utterances which include grammatical errors. Therefore, the error detection rate by the proposed method 1 is much lower than the manual method.

However, the recognition results of the manual method and the proposed method 2 for input speech “牛タン に お願い します” are as follow:

<table>
<thead>
<tr>
<th></th>
<th>Recognition rate</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sentence</td>
<td>Word</td>
<td></td>
</tr>
<tr>
<td>Manual (reference) method</td>
<td>90.43%</td>
<td>95.55%</td>
<td></td>
</tr>
<tr>
<td>Proposed method 1</td>
<td>82.62%</td>
<td>87.87%</td>
<td></td>
</tr>
<tr>
<td>Proposed method 2</td>
<td>87.66%</td>
<td>92.71%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Error detection rate</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Particle</td>
<td>Verb</td>
<td>Lexical</td>
</tr>
<tr>
<td>Manual (reference) method</td>
<td>92.92%</td>
<td>91.66%</td>
<td>79.31%</td>
</tr>
<tr>
<td>Proposed method 1</td>
<td>63.64%</td>
<td>66.67%</td>
<td>51.03%</td>
</tr>
<tr>
<td>Proposed method 2</td>
<td>92.29%</td>
<td>83.33%</td>
<td>79.31%</td>
</tr>
</tbody>
</table>
The manual (referenced) method

牛タン に お願い します

The proposed method 2

N-best : 1 You have Grammar Errors
Your speech: 牛タン に お願い します
Error modified: 牛タン に お願い します.

The expression of “に（を）” in the manual method means that grammatical error word “に” is in input speech instead of the correct word “を”.

As shown in Table 4, the recognition rate and grammar error detection rate decreased by 2.77% and 1.43% as compared to the manual method. For the proposed method 2, another example is shown in Fig. 8.

As shown in Fig. 8, N-best results are calculated in the first step. Then, each N-best is modified to form proper sentences in the second step like “この を いち箱 で Fillar いくら です か → これ を ひと箱 で いくら ですか”. Finally, the acceptance is checked to find learner’s utterance using correct grammar.

For the recognition error in the proposed method 2, another example is shown in Fig. 9. Because the general grammar which does not consider relationship between each noun, verb and so on was used, the 1-best result may include recognition error and still be accepted in
Input speech “この いち箱 で いくら ですか”

N-best in the first step
First 1: この を いち箱 で 「Fillar」 いくら ですか
First 2: 「Fillar」 この を いち箱 で 「Fillar」 いくら ですか

First 9: この いち箱 で 「Fillar」 いくら ですか

Generated correct sentence in the second step
Second 1-1: これ を ひと箱 で いくら ですか
Second 2-1: これ を ひと箱 で いくら ですか

Second 9-1: これ を ひと箱 で いくら ですか

Accept/reject in the third step and recognition result
Second 1-1 reject, Second 2-1 reject, —, Second 9-1 accept
N-best: 9 You have Grammar Errors
Your speech: この いち箱 で いくら ですか
Error modified: これ を ひと箱 で いくら ですか

Fig. 8 Example of recognition result of the proposed method 2

Input utterance: それ で 下さい

The manual method: それ で (を) 下さい

The proposed method 2:
N-best: 1 You have Grammar Errors
Your speech: もの それ で 下さい
Error modified: もの それ を 下さい

Fig. 9 Example of recognition error by the proposed method 2

third step.
However, the recognition rate and the grammatical error detection rate were similar to the manual method. We point out the recognition improvement and grammar error detection rate of the manual method is not significant at the significance level of 0.05 as compared to the proposed method 2.

We also checked the processing time (2.6 GHz Xeon CPU) for 50 utterances. Comparing to the manual method, the proposed method 2 took 1.28 times longer (225 second in the manual method, 290 second in the proposed method 2 for 50 utterances).
From the experimental result, the proposed method 2 was performed similarly to the manual method. Moreover, the method 2 can be applied automatically to the system without extending the recognition grammar. From these results, we can say that the proposed method 2 is better than the manual method.

5 CONCLUSION

In this paper, we proposed new methods to detect grammatical and lexical errors from an utterance for a dialogue-based CALL system.

To realize the dialogue-based CALL system in allowing a learner to exercise 'real conversation', grammatical/lexical errors as well as out-of-grammar utterances have to be treated. The out-of-grammar problems were solved by introducing a pre-exercise. The number of out-of-grammar sentences decreased if a learner does pre-exercise. In order to detect a learner's grammatical and lexical errors, we examined several methods. According to the experiment results, the proposed method 1 had a lower grammar error detection rate and recognition rate, as compared to the manual (reference) method. However, we confirmed that the proposed method 2 was performed similarly in both the grammar error detection rate and the recognition rate. Considering the disadvantage of the manual method, the proposed method 2 was the most effective.

This study showed a new concept of dialogue-based CALL system, but there still remain other subjects. As a future work, we need to develop a total courseware of dialogue-based CALL system and verify the effectiveness of the integrated system.

Reference


Hasimoto, T., and Shimada, S. “Grammatical Errors Of Using Particles in Japanese Sentences Written by Foreigners.” In SIG Note of IPS Japan, NL, No. 117.


**Oh Pyo Kweon:** He was born in Korea in 1972. He received the B.E. in Yeungnam University in Korea in 1999 and M.E. and Ph. D. degrees from Tohoku University, Sendai, Japan, in 2002, and 2005 respectively. Since 2005, he has worked with Mobile Handset Research Center, LG electronics
Inc. in Korea, as a Senior Engineering. He has engaged in spoken language processing, speech coding and acoustics. He is a member of the International Speech Communication Association

**Akinori Ito:** He was born in Yamagata, Japan in 1963. He received the B.E., M.E. and Ph. D. degrees from Tohoku University, Sendai, Japan, in 1984, 1986 and 1992 respectively. Since 1992, he has worked with Research Center for Information Sciences and Education Center for Information Processing, Tohoku University. He joined Faculty of Engineering, Yamagata University from 1995 to 2002. From 1998 to 1999 he worked with College of Engineering, Boston University, MA, USA, as a visiting scholar. He is now an Associate Professor of Graduate School of Engineering, Tohoku University. He has engaged in spoken language processing and statistical text processing. He is a member of the Institute of Electronics, Information and Communication Engineers, the Acoustical Society of Japan, and the Information Processing Society Japan.

**Motoyuki Suzuki:** He was born in Chiba, Japan in 1970. He received the B.E., M.E. and Ph. D. degrees from Tohoku University, Sendai, Japan, in 1993, 1995 and 2004 respectively. Since 1996, he has worked with the Computer Center, and Information Synergy Center, Tohoku University, as a Research Associate. He is now a Research Associate of Graduate School of Engineering, Tohoku University. He has engaged in spoken language processing, speech coding and statistical pattern recognition. He is a member of the International Speech Communication Association, the Institute of Electronics, Information and Communication Engineers, and the Acoustical Society of Japan.

**Shozo Makino:** He was born in Osaka, Japan, on January 3, 1947. He received the B.E., M.E. and Dr. Eng. Degrees from Tohoku University, Sendai, Japan, in 1969, 1971 and 1974, respectively. Since 1974, he has worked with the Research Institute of Electrical Communication, Research Center for Applied Information Sciences, Graduate School of Information Science, Computer Center and Information Synergy Center, as a Research Associate, an Associate Professor and a Professor. He is now a Professor of Graduate School of Engineering, Tohoku University. He has been engaged in spoken language processing, speech corpus, music information processing,
image recognition and understanding, natural language processing, semantic web search and digital signal processing

(Received September 30, 2004)
(Revised December 22, 2004)
(Rerevised January 20, 2005)
(Accepted April 23, 2005)