A dialogue-based CALL system for Japanese conversation

Oh Pyo Kweon, Motoyuki Suzuki, Akinori Ito and Shozo Makino
Graduate School of Engineering, Tohoku University
05 Aramaki aza Aoba Aoba-ku, Sendai 980-8579 Japan
(kweop, moto, aito, makino)@makino.ecei.tohoku.ac.jp

Abstract
This paper describes a dialogue-based CALL (Computer Assisted Language Learning) system. One of the major problems in CALL systems is that learners are usually assigned a passive role. Learners have no practices in composing their own utterances. The other major problem is that lots of the proposed CALL systems are the pronunciation exercise systems such as minimal pairs. The pronunciation exercise is an unrealistic task, if the goal of the learner is to obtain the ability to participate actively in a conversation. We proposed the dialogue-based CALL system with which learners can practice in making a conversation and in composing utterances actively.

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 effectively communicate with native speakers. It is ideal if a foreign language learner can be taught on a one-to-one basis. Unfortunately, this is actually difficult for most students due to economic reasons. Conventional self-study methods such as tape recorder based learning force learners to identify pronunciation and grammar mistakes by themselves. This is a difficult task for a foreign language learner. If a learner could find their mistakes he/she would have no need to learn them. The application of speech recognition technology can solve this problem.

1.2. Previous CALL system
As development of speech recognition technology in recent years, the expectation of a new concept CALL system has been growing. Various CALL systems are proposed [1, 2]. However, almost all conventional systems are practice systems of pronunciation and listening such as minimal pairs exercise. The pronunciation exercise is insufficient to obtain the ability of conversation.

A dialogue-based CALL system can roughly be divided into two by whether a student creates a response sentence actively or not. The one is that the answers are ready-made with vocabulary chosen and syntax assembled if learners are asked to answer a question [3]. As a result, students have no practice in composing utterances on their own. AuraLog [4], for example, has produced an appealing language teaching system that feeds the user’s pronunciation of one of three written sentences to the recognizer. The path of the dialogue is dependent on which of these sentences is elicited. While a certain degree of realism is attained, students do not actively composing any of the utterances they produce.

The other is that a system lets learners create limitedly their own utterances [5, 6]. That system gives learners an active rather than passive role. They acquire automatic reflexes in this way that are useful later, when they need to build an utterance during a real conversation. Then they should be able to maintain a conversational tempo rather than searching for correct structures and words. For example, when the system asked them “When did you find it?” with a key-word “Yesterday”, learners construct answers using a key-word by themselves. However, the path of the dialogue is not changed. Because learners use a key-word, the answer is easily predicted. Learners can not actively compose various sentences for the question of the system also.

1.3. Object
In order to raise a student’s communication capability, it is ideal to develop the system which allows learners to make a conversation and construct various sentences near to real conversation. The purpose of research is constructing the dialogue-based CALL system near real conversation, as shown in Figure 1.

System: May I help you?
Student: I want to buy a souvenir.
System: The most famous one in Sendai city is Gyutan.
Student: How much is it?
System: ¥ 1000 for one box.
Student: I want to buy two boxes.
System: Would you like to pay by cash or credit card?
Student: Credit card, please.
System: Thanks.

Figure 1: Conversation of souvenir task
It is the system which allows learners to practice in making a conversation and in constructing utterances actively. For example, when the system asks [May I help you?], they can say [I want to a souvenir] or [What is the most famous?] and so on. The path of the dialogue is also changed by learners’ response utterances.

2. Proposed dialogue-based CALL system

Our object is to construct the system which allows learners to compose various sentences. However, if the system gives learners the flexibility of sentence creation, recognition accuracy will become poor and learners can not make a conversation with a system. We defined the dialogue situation as a particular task like a souvenir task because of the poor recognition rate. Hence, a system can easily predict what learners can say. In the limit of the task, learners can construct their own utterances. The path of the dialogue is also changed by learners’ utterance.

2.1. Problems of dialogue-based CALL system

When learners create various utterances, they make various errors. For example, they try to pronounce [hitotsuyu kudasai] instead of the sentence [hitotsu kudasai: one please] in a souvenir task. They also pronounce [ni kudasai: second please] instead of [hutatsu kudasai: two please]. They may construct sentences using the unknown words. The system can not recognize correctly, especially utterances of foreigners because of poor recognition rate. In that case, learners can not exercise a conversation and conversation is failed. The problem of a dialogue-based CALL system is below.

1. Mispronunciations
2. Grammatical errors
3. Lexical errors
4. Problems of unknown words
5. Recognition error

2.2. Proposed method

To solve the problems explained in section 2.1, we proposed below methods.

- To solve the mispronunciations, vocabulary in the dictionary is extended to be robust for the mispronunciation using error pattern. Using difference of likelihood, the system evaluates mispronunciation [7].
- To solve grammatical errors and lexical errors, we extended vocabulary in the dictionary using previous works [8].
- To decrease the problems of unknown words, lexical and grammatical errors, we also used the study flow which is often used for a VTR teaching materials. Firstly, learners study the verb, noun, grammar, dialogue situation, and typical conversation examples. Then, they make a conversation with the system.

- When system recognize wrong, learners can input text alternatively. They exercise next conversation and learn the path of conversation even though the system fails to recognize utterances of learners correctly.

3. Investigation of grammatical errors and lexical errors

To solve lexical and grammatical error problems, we extended the vocabulary and used the study flow often used for a VTR teaching materials. However, we not exactly know influence of the study flow. For example, if learners study grammar and typical conversation example, we do not know how much lexical and grammatical errors decrease comparing to when learners make a conversation without pre-exercises.

3.1. Experiment

We investigated how lexical and grammatical problems change according to a difference of pre-exercise methods. According to Figure 2, first, situation is explained to learners. For example, native Japanese speaker explains the situation like [Here is a souvenir store. I am a staff of the stores. You are a customer]. Then, foreigners are interviewed with native speaker without any pre-exercise (Interview 1). The other way is that foreigners are interviewed with native speaker after they do pre-exercise such as typical conversation examples and vocabulary (Interview 2). Because our purpose is to investigate grammatical and lexical errors, native speaker writes the sentences if foreigners have difficulty to understand. For a native speakers, we prepared the manual which indicate how to make questions and answers corresponding to utterances of foreigners.

As the way Figure 2 shows, two Korean speakers were interviewed about souvenir task and hotel reserva-
tion task. One Korean has studied Japanese about 1 year and the other has done about 1.6 year.

3.2. Result of interview

One of the transcriptions of interview 1 is in Figure 3. As Figure 3 shows, learners used a lot of [a-, e-, e: to].

Native: いらっしゃいませ. 東北ホテルです. 今日から泊まりますか.
Foreigner: はい.
Native: いつまで泊まる予定ですか.
Foreigner: あの. いつかまで.
Native: お1人で泊まりますとシングルになりますけど. シングルでよろしいですか.
Foreigner: あのシングルがいいと思います.
Native: はい. 名前と住所と電話番号をこちらに書いてください.
Foreigner: はい. 分かりました.
Native: はい. シングルルーム1つかまで3万円になりますけども. 支払いは現金でよろしいですか.
Foreigner: あの. カード. いいですか.
Native: お支払いはカードで. 部屋の鍵です. 部屋の番号は612室です.
Foreigner: 612. はい. 分かりました. はい. ありがとうございました.

Figure 3: One of examples of Interview 1

Because any pre-exercises were not done, learners have problems to construct sentences immediately to make a conversation. Almost all sentences which learners constructed are possible to predict. However, there are a few sentences to be difficult to predict. For example, one of foreigners asked [May I try to eat?] corresponding to the recommendation of native speaker [Gyutan is famous in Sendai city].

One of the transcriptions of interview 2 are shown in Figure 4. Underlined sentences in Figure 4 mean that sentences which learners constructed are exactly same as the sentences in pre-exercises. Bold sentences are similar to sentences in pre-exercises. Vocabulary and sentences which learners used in interview 2 are the same vocabulary and sentences which are used in pre-exercise of vocabulary and typical conversation. As results, the same and similar utterances which are used in pre-exercise are 17 and 6 out of 26 utterances in interview 2. Difficult utterances to predict (Outside of grammar) are decreased from 23.5% utterances in interview 1 to 11.5% in interview 2. Grammatical errors of foreigners are also decreased from 0.26 (the number of grammatical errors per each utterance) to 0.19 such as [これでします]. As comparng results of interview 1 to 2, in case of using pre-exercise, we can control the rate of using grammatical error and unknown word.

Native: いらっしゃいませ. 東北ホテルです. 今日から泊まりますか.
Foreigner: はい. 泊まります. 今日から泊まります.
Native: いつまで泊まる予定ですか.
Foreigner: あの. 明後日までです.
Native: はい. 部屋はどちらにしますか.
Foreigner: あの. シングルルームとダブルルームはいいですか.
Native: シングルが泊5千円. ダブルルームが7千円になります.
Foreigner: シングルルームでお願いします.
Native: じゃ. こちらに名前と住所と電話番号を書いてください.
Foreigner: はい. 住所は英語で大丈夫ですか.
Native: はい. 大丈夫です. お支払いはどうしますか.
Foreigner: あの. 現金ですか.
native: はい. 大丈夫です. 合計15000円になります. 部屋の鍵です. 部屋の番号は612室です.
Foreigner: 分かりました. ありがとうございました.

Figure 4: One of examples of Interview 2

4. Recognition rate of utterances

In the present system, we defined the dialogue situation as a particular task like a souvenir task because of the recognition rate as we mentioned in section 2. Hence, we need to survey the recognition rate of learners and we need to find a way to increase recognition rate if it is poor.

4.1. Experiment method

The details of experiment are shown in Table 1. First, to check the recognition rate of learners, we did phoneme recognition using Japanese phone models (baseline models) which is constructed by Japanese speech DB, ATR B and C sets. There are ambiguous pronunciations in utterances of foreigner. For example, a Korean can hardly distinguish the voiced and unvoiced sound. They only have a distinction of aspiration and un-aspiration. If those ambiguities are trained, the recognition rate could be increased. Hence, baseline models are trained using Foreigners’ speech DB (trained models). We used 8 male and 7 female for training (about 100 sentences per each speakers). We did not use grammar (possibility of any phoneme after any phoneme is the same).

4.2. Result of Recognition rate

The result of recognition rate is shown in Figure 5. M-Base and F-Base indicate the recognition rate using baseline models of male and female. M-training and F-training indicate the recognition rate after baseline models are trained by Foreigners’ speech DB. According to Figure 5, the phoneme correct rates of baseline models are about 57.97% and 61.78% for male and female. Respectively, the accuracy is 10.23% and 15.10%. Because
Table 1: Experiment Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Features</th>
<th>Total Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>Monophone HMM</td>
<td>34 features</td>
</tr>
<tr>
<td></td>
<td>MFCC+∆MFCC+Pow+∆Pow,</td>
<td></td>
</tr>
<tr>
<td>Baseline Model</td>
<td>Japanese Phoneme Model from</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATR B, C sets</td>
<td></td>
</tr>
<tr>
<td>Training data</td>
<td>8 Korean male, 7 Korean female</td>
<td></td>
</tr>
<tr>
<td>Testing data</td>
<td>Korean male, 4 Korean female (about 100 sentences per each Korean speakers)</td>
<td></td>
</tr>
</tbody>
</table>

the speaking speed of foreigners is slower than the one of native speakers, there are lots of insertion errors. After training the baseline models, the correct/accuracy rate of male are increased from 57.97/10.23% to 66/13.37%. The rate of female is also increased from 61.78/15.10% to 66/19.96%.

We also check the recognition tendency using about 20 sentences which are derived from interview experiment in section 3. For example, the recognition result of baseline models is [Hai, ka:dodemo desuka] for learner’s utterance [a-,ka:dodemo ii desuka]. However, the recognition result of trained models is [hai, ka:dodemo ii desuka]. In all case except one sentence, recognition results are better when we used trained models instead of baseline models. Word correct rate is increased from 79.7% to 81.1% after training the baseline models.

![Figure 5: Recognition rate](image)

5. Conclusions

In this paper, we discussed about a dialogue-based CALL system. We proposed the dialogue-based CALL system to allow learners to construct various sentences and change the path of dialogue depending on the constructed sentences by learners.

We also discussed about the flow of exercise. We know that grammatical errors, lexical errors and unknown words are decreased if learners do pre-exercise of vocabulary and typical dialogue exercise before making a conversation with the system. We also checked the recognition rate of foreigner. As the ambiguous pronunciation is trained, the recognition rate is increased.

6. Future work

We discussed the grammar error and lexical error in section 3. The flow of exercise can help to decrease grammatical and lexical errors. However, there still remain lots of problems to be discussed. First, interview data is too small to cover grammatical and lexical errors. For future work, we need to correct more interview data and analyze the grammar errors more correctly. Grammatical error and lexical errors are also influenced by language abilities of learners. We also need to analyze the grammar error and the range depending on language abilities of learners.

7. References


