A Japanese CALL System based on Dynamic Question Generation and Error Prediction for ASR

Hongcui Wang and Tatsuya Kawahara

School of Informatics, Kyoto University,
Sakyo-ku, Kyoto 606-8501, Japan
wang@ar.media.kyoto-u.ac.jp

Abstract

We have developed a new CALL system to aid students learning Japanese as a second language. The system offers students the chance to practice the Japanese grammar and vocabulary, by creating their own sentences based on visual prompts, before receiving feedback on their mistakes. Questions are dynamically generated along with sentence patterns of the lesson point, to realize variety and flexibility of the lesson. Students can give their answers with either text input or speech input. To enhance speech recognition performance, a decision tree-based method is incorporated to predict possible errors made by non-native speakers for each generated sentence on the fly. Trials of the system have been conducted with a number of foreign students in our university, and positive feedbacks were obtained.

Index Terms: CALL, Second language learning, ASR, Error prediction, Japanese.

1. Introduction

Computer Assisted Language Learning (CALL) systems can provide many potential benefits for both learners and teachers [1-2]. And there is significant interest in the development of CALL systems recently. Many research efforts have been done for improvement of such systems especially in the field of second language learning [3-6].

Some of CALL systems focus on practicing and correcting pronunciation of individual vowels, consonants, words, such as the system in [3], FLUENCY [7], WebGrader [8], and EduSpeakTM [9]. Some concentrate on vocabulary or grammar learning[10]. And also some allow training of an entire situation-based conversation, such as the Subarashii system [4]. However, little has been done to improve learners' communication ability including sentence generation skill. Motivated by these, we have designed and developed a new CALL system called CALLJ to aid students learn the elementary Japanese grammar and vocabulary via a set of dynamically generated sentence production exercises.

We have previously presented the system based on text-input via a keyboard in [11]. In its evaluation, a number of students asked for speech-input capability, that is, they prefered to practice uttering their answers. Thus, we have investigated the incorporation of ASR (Automatic Speech Recognition) technology to this system. Whereas there is no problem for the text input system to detect errors that are out-of-vocabulary or out-of-grammar for any input, this is not the case in the speech input system. Since ASR relies on the constrained grammar and limited vocabulary, the biggest challenge was to predict errors made by students without the degradation of ASR performance.

To solve this problem, we have proposed a decision tree-based error prediction method [12]. In this work, we incorporate it to the CALLJ system and evaluate with a number of subjects. Section 2 gives the system design and some implemented modules. Then, Section 3 presents the evaluation results, findings and feedbacks from students. At last, section 4 concludes with a summary.

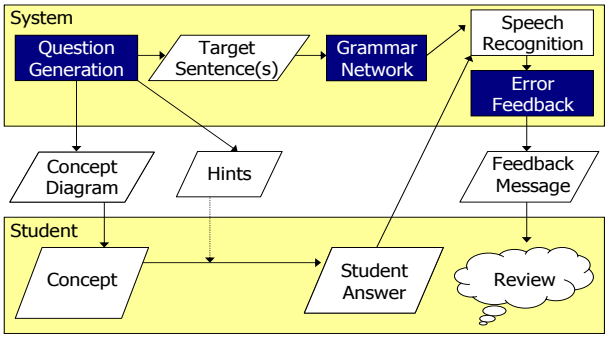

Figure 1: System overview

2. System Overview

CALLJ system is organized in lessons. A lesson is a collection of related questions (sentences) connected to some key sentence patterns (grammar points), such as "like to do something". A process flow of the system is depicted in Figure 1. The system generates questions on the fly, based on a key sentence pattern that the students are to practice. Each question involves the students being shown a "Concept Diagram", which is a picture representing a certain situation or scene. The students are then asked to describe this situation with an appro-

Accepted after peer review of full paper

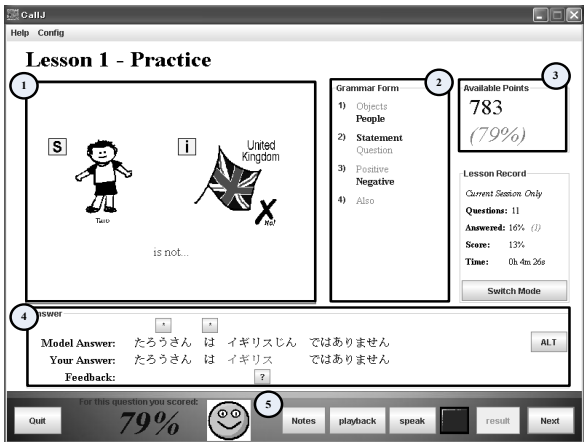

Figure 2: Question practice screen; 1: Concept diagram, 2: Desired form guide, 3: Score, 4: Answer area and feedback display buttons, 5: Control button panel

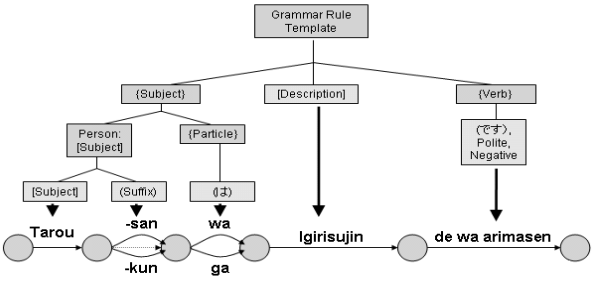

Figure 3: Grammar-based sentence generation

priate Japanese sentence using text input or speech input. Thus, the system allows students the freedom to create their own sentences. If the answer is given via a microphone, ASR is done using a dynamically generated language model in the form of a grammar network for the target sentence. Errors will be detected and feedback information is generated for the students. Figure 2 shows the user practice interface. In the followings, we describe further details regarding the main modules of the system, namely question generation, grammar network generation, and error feedback.

2.1. Dynamic Question Generation

In order to reduce the repetitiveness of the questions offered by the system, we dynamically generate each question at run time from the set of vocabulary and grammar rules available. This involves creation of three main components: a concept or situation that the students must describe, target sentence instances that the students are expected to produce, and a diagram that expresses this situation.

A template is prepared to cover a range of related situations. It defines the semantic components or slots that are required, optional or to be omitted when defining a specific situation. Then, the target sentences are

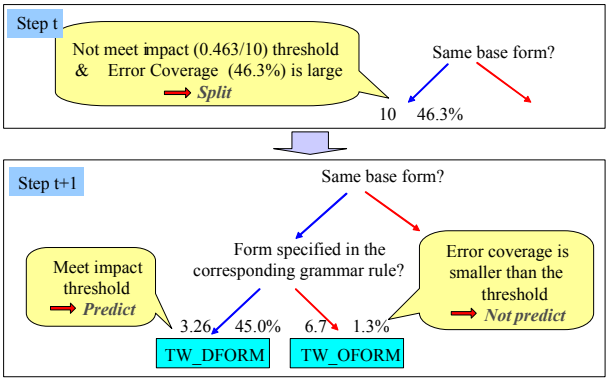

Figure 4: Example of decision tree training process

created by taking the information in the concept instance and applying a set of grammar rules (sentence patterns), as shown in Figure 3. In the user interface, the diagram is created by combining a number of smaller sub-images, each representing a component of the concept instance.

2.2. Grammar Network Generation for ASR

As the system has an idea of the desired target sentences, and patterns of possible sentences for the given picture are limited, it is reasonable to perform ASR based on a finite state grammar network representing each target sentence. To be an effective CALL system, the grammar network should cover errors that non-native learners tend to make. However, considering all possible errors would significantly increase the perplexity of the network, thus degrade the ASR performance. Therefore, a decision tree-based error classification algorithm is proposed [12].

2.2.1. *Error Classification*

The error classification is done by comparing the features of the observed word to those of the target word. The features include same POS, same base form, similar concept, wrong inflection form etc. Coverage-perplexity (=impact) criterion is introduced to find an optimal decision tree that balances the tradeoff of the error coverage and perplexity. It is used to expand a certain tree node from the root node (containing everything), and partition the data contained in the node according to some feature. For a given error pattern, it is defined as below:

$$impact = \frac{\text{increase in error coverage}}{\text{increase in perplexity}}$$

The larger value of this impact, the better recognition performance can be achieved with this error prediction. Our goal is reduced to finding a set of error patterns that have large impacts. If a current node in the tree does not meet this criteria (threshold), we expand the node and partition the data iteratively until we find the effective subsets and mark “to predict”, or the subset’s coverage becomes too small and mark “not to predict”. Figure 4 shows an example of one step of the tree training for

Table 1: Error patterns being predicted for verbs

Pattern	Class	Description
TW_DForm	grammar	same base form with the target word, but not the desired verb form
DW_SForm	lexical	a similar concept word, having the same verb form with the target word
DW_DForm	lexical	a similar concept word, with a different verb form available in its grammar rule
TW_WIF	grammar	wrong inflection of the target word

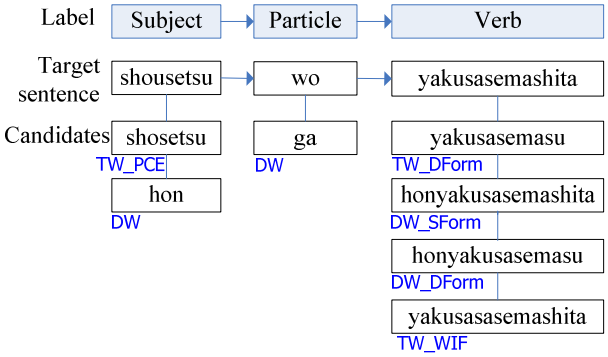

Figure 5: Prediction result for given sentence

verbs. In each node, perplexity and error coverage of the node is labeled from left to right.

The training data for the decision tree learning were collected through the trials of the prototype CALLJ system with text input [11]. They consist of 880 sentences. Table 1 lists all error patterns for verbs which are chosen for prediction by the decision tree learning.

2.2.2. Error Prediction Integrated to Language Model

As we identified the errors to predict, we can exploit this information to generate a finite state grammar network. Given a target sentence, for each word in the surface form, we extract its features needed such as POS and the base form, and compare the features with error patterns to predict using the decision tree. Then, we generate potential error patterns with the prediction rules and add them to the grammar node. Figure 5 shows an example of a recognition grammar based on the proposed method for a sentence "*shousetsu wo yakusasemashitaka*".

2.3. Feedback to Learners

The CALL system should provide pertinent corrective feedback of errors made by students. The feedback in our system consists of a number of pieces of information. Firstly, it includes some basic information about the error class, extracted directly from the features used in the error prediction. In addition, a short text is also displayed to outline the error including why they may have made it, and what they should do to correct it. This text is prepared for each error pattern.

3. Experiments and Evaluation

3.1. Experiment Setup

Ten foreign students of Kyoto University took part in the trials of the system. They are from seven different countries including China, France, Germany and Korea. They had no experience with the CALL system before the trial, but were briefly introduced before undertaking the task. Each student ran through a set of lessons, answering a set of generated questions by speech input before seeing the correct answers and feedback for errors they made. ASR based on a grammar network was executed at run time. After the trials, all utterances were transcribed including errors by a Japanese teacher.

3.2. ASR Performance

Comparing to the transcription of utterances, the WER (Word Error Rate) of ASR is 11.2%, which is quite lower compared with the case (28.5%) using the baseline grammar for the text-input system. And up to 62.9% of errors made by students were correctly detected, though 85.7% of errors were covered by the grammar network and could be recognized in theory.

3.3. Error Analysis for System Improvement

Figure 6 shows the distribution of different types of errors detected during the trials. The error rate is calculated by dividing the total occurrence of each error type by the number of components observed on which that error type may occur. It is observed that the most frequent form of problems is lexical errors. This result suggests that the lexical errors were more important than the grammatical errors, and that the students had more problems with vocabulary deficiencies. For better learning effect, the system is revised to select a relevant concept rather than generating randomly. That is, the system keeps track of the words and forms (mainly for verbs) erroneously replied in the previous question, and try to use them in the next question, until the learners correct them.

3.4. Error Correction by Considering Communication Aspect

We investigate the word errors of ASR that were covered by the the grammar network but could not be detected, which amounts to about 23%. It is observed that the majority of such errors by the system belong to "TW_PCE" type, which means the word is pronounced erroneously by adding or omitting a single double consonant, long vowel or voiced pronunciation, for example, "kipu" instead of "kippu". Actually, most of these errors do not cause difficulty for people to understand in a context of a whole sentence. Thus, we offer students an option to

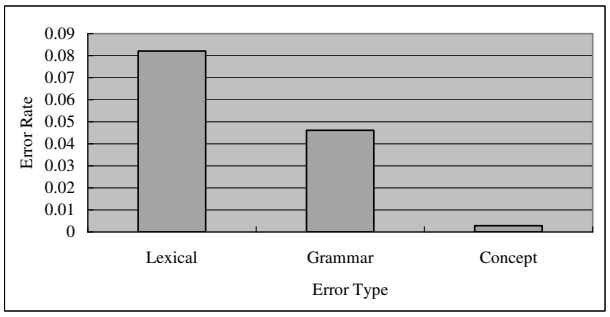

Figure 6: Frequencies of observed error types

Table 2: System assessment

The key point of each lesson was clear	4.2
I could clearly understand the Concept Diagram	3.3
I found that the diagram became easier to understand over time	4.7
In general, did you feel you experienced a lot of problems with the ASR in CALLJ	10%
I prefer to speak my answer than to type it	3.8
I prefer to type my answer as opposed to speaking it	3.0
I would always like to be able to choose whether to speak or type	4.2
I would enjoy using such a system	90%
I would like to have used such a system before coming to Japan	90%

weigh "communication more than pronunciation details". If students choose this option, the system will automatically correct the above-mentioned errors (either by the students or by the system) before displaying the ASR results as they are. This would improve the robustness of the system.

3.5. System Assessment by Students

After the trial, students were asked to evaluate the system with a questionnaire. Some questions and statistics were listed in Table 2. In the table, percentage is the ratio of students who selected each statement as appropriate and the score is from 1 (strongly disagree with statement) to 5 (strongly agree with statement).

It is confirmed that the key grammar point of each lesson is clear and the concept of the scene represented by a picture is easier to understand over time. Most students could tolerate the ASR problems and would enjoy using such a system, especially before coming to Japan. It is also observed that more students like to have the choice of using text input or speech input, which is now available. Some suggestions were given and adopted, for example, adding a function of listening to what students have said to help them find pronunciation errors by themselves.

4. Conclusion

We have completed a new CALL system CALLJ for studying the elementary Japanese grammar and vocabulary and improving their communication ability. We have given an overview of the fully implemented system, and the evaluation of the system with the trials by a number of students. It is confirmed that they enjoy using the system and find it useful for language learning.

5. References

[1] Witt S.M, "Use of Speech Recognition in computer-assisted Language Learning", PhD's thesis, November, 1999.

[2] Zinovjeva, N., "Use of Speech Technology in Learning to Speak a Foreign Language", Speech Technology, Autumn 2005.

[3] Kawai,G., Hirose,K., "A Call System using Speech Recognition to Train the Pronunciation of Japanese Long Vowels, the Mora Nasal and Mora Obstruent", Proc. Europspeech,657-660, 1997.

[4] Bernstein, J., Najimi, A., Ehsani, F. "Subarashii: Encounters in Japanese Spoken Language Education", CALICO Journal, 1999.

[5] Tsubota,Y., Kawahara,T., and Dantsuji,M., "Practical Use of English Pronunciation System for Japanese Students in the CALL Classroom", ICSLP, 2004.

[6] Abdou,S.M., Hamid, S.E., Rashwan, M., et al., "Computer Aided Pronunciation Learning System Using Speech Recognition Technology", Interspeech, 2006.

[7] Eskenazi M., and Hansma S., "The Fluency Pronunciation Trainer", Proc. STiLL Workshop on Speech Technology in language learning, Marhallmen,1998.

[8] Neumeyer L., Franco H., Abrash V., "WebGrader: A Multilingual Pronunciation Practice Tool", Proceeding of ICSLP,566-569, 2000.

[9] Franco H., Abrash V., Precoda K., "The SRI EduSpeakTM System: Recognition and Pronunciation Scoring for Language Learning", Proceeding of STILL, 2000.

[10] Nagata,N., "Japanese Courseware for Distance Learning", AILA, 2000.

[11] Waple,C., Wang, H., Kawahara,T., Tsubota,Y., and Dantsuji,M., "Evaluating and Optimizing Japanese Tutor System Featuring Dynamic Question Generation and Interactive Guidance", ICSLP, 2007.

[12] Wang, H., Kawahara, T., "Effective Error Prediction using Decision Tree for ASR Grammar Network in CALL System", ICASSP, 2008.