Grammatical error detection from English utterances spoken by Japanese

Takuya Anzai*, Seongjun Hahn*, Akinori Ito†, Masashi Ito† and Shozo Makino‡

* Tohoku University, Sendai 980-8579 Miyagi Japan
E-mail: {takuya,branden65,aito}@makino.ecei.tohoku.ac.jp
† Tohoku Institute of Technology, Sendai 982-8577 Miyagi Japan
E-mail: itojin@tohtech.ac.jp
‡ Tohoku Bunka Gakuen University, Sendai 981-0943 Miyagi Japan
E-mail: makino@ait.tbgu.ac.jp

Abstract—This paper describes methods to recognize English utterances by Japanese learners as accurately as possible and detects grammatical errors from the transcription of the utterances. This method is a building block for the voice-interactive Computer-Assisted Language Learning (CALL) system that enables a learner to make conversation practice with a computer.

A difficult point for development of such a system is that the utterances made by the learners contain grammatical mistakes, which are not assumed to happen in an ordinary speech recognizer. To realize generation of accurate transcription including grammatical mistakes, we employed a language model based on an N-gram trained by generated texts. The text generation is based on grammatical error rules that reflect tendency of grammatical mistakes made by Japanese learners. The experimental results showed that the proposed method improved recognition accuracy compared with the conventional recognition and error detection method.

I. INTRODUCTION

With the progress of globalization in recent years, population of English learners has been increased. Among various English learning methods, Computer-Assisted Language Learning (CALL) system is one of the most promising learning methods. Most CALL systems focus on training for reading, writing and listening, and few commercial CALL systems provide with training method for speaking. Recently, CALL systems for speaking practice have been developed. However, conventional CALL systems with speaking practice focus on training of pronunciation or intonation [1,2,3].

To improve conversation skills, it is necessary for learners to practice conversation in a real dialogue. So, several voice-interactive CALL systems for conversation practice have been developed [4,5].

One of the biggest problems on realizing such CALL systems is how to recognize learners’ utterance correctly. In most cases, utterances spoken by a foreign language learner contain pronunciation errors and grammatical mistakes. However, a voice-interactive CALL system must recognize a learner’s utterance correctly including grammatical errors to progress dialogue smoothly and to give pertinent feedbacks. Therefore, to develop a CALL system for grammar or expression practice, we have to improve recognition accuracy of utterances containing grammatical errors.

In this paper, we propose a method for recognizing learner’s utterances and detecting grammatical errors. This work is an improvement of recognition method based on text generation and N-gram trained from the generated text [6]. We revised the rules of generation of sentences with errors for improving recognition performance.

On developing the CALL system, we assume that learners first do a pre-exercise on the fundamental vocabulary and grammatical expressions used in a specific situation, and then make a conversation with the system on the topics. The learner’s utterance is recognized by the system using a speech recognizer. If the utterance contains any errors, the system gives feedback to the learner. Therefore, learners progress a conversation practice [6].

Because learners already know keywords through the pre-exercise before the conversation with the system, we can expect learners to respond to the system using the same expressions as those appearing in the pre-exercise [5]. We call a correct sentence expected to be uttered by learners “a target sentence”. However, in reality, not all utterances made by a learner match the target sentences. A sentence actually uttered by a learner is referred to as “an uttered sentence”. And then, the recognition results by the speech recognizer are often different from the uttered sentences because of recognition errors. We refer to a sentence obtained from the speech recognizer as “a recognized sentence”.

II. RECOGNITION OF LEARNERS’ UTTERANCE USING AN N-GRAM FROM GENERATED TEXT

In this section, we describe the approach by Ito et al. [6], which is our previous work. An N-gram LM is powerful and flexible with regard to unpredictable errors, but it requires a large amount of training data. For training an N-gram without large training data, we developed a method to train an N-gram using artificial sentences generated from the target sentences. First, we prepare grammatical error rules that are frequently made by Japanese learners. Then, the rules are applied to the target sentences, and a sufficient amount of sentences are generated. Finally, a back-off N-gram is trained from the generated sentences.

482


10-0104820485©2010 APSIPA. All rights reserved.
In this text generation method, two kinds of error rules are used: corpus-based error rules and generic error rules. The first ones are extracted from transcriptions of English utterances by native Japanese speakers. The SST (Standard Speaking Test) corpus was used as a source of the error rules [7]. The SST corpus consists of transcriptions of interviews with native Japanese speakers in English. The grammatical and lexical errors in this corpus are manually annotated with the correct expressions. Table I shows the errors most frequently observed in the corpus. The symbol $\epsilon$ denotes the absence of a corresponding word. The error types DEL, INS and SUB indicate a deletion error, an insertion error and a substitution error, respectively. The second ones are manually created to cover a wider range of error types. These error rules include insertion of articles, confusion between singular and plural, adjective comparative and verb tense errors.

On applying these two kinds of rules, the generic error rules were first applied, and the corpus-based rules were applied next. The probabilities of choosing a corpus-based rule are in proportion to the frequencies of the errors in the SST corpus. Conversely, the probabilities of applying a generic error rule are given a priori. The final recognition performance was not greatly affected by the probabilities of the generic error rules. In the corpus-based rules, only substitution and deletion error rules were applied to one word. However, errors applied to more than one word (e.g., Correct: office worker, Uttered: work man) and insertion errors also exist in the SST corpus. Therefore, the error rules concerning multiple words should be treated properly.

### III. Revision of Sentence Generation Rules

#### A. Contraction rules

In the previous work, generic error rules include contraction rules such as “I am $\rightarrow$ I’m” or “He’s $\rightarrow$ He is”. However, for a contraction is not a mistake, it is inappropriate to regard a contraction as a grammatical error. Therefore, we generate sentences with and without contractions, and all of the generated sentences are used as new target sentences.

#### B. Corpus-based rules

Ito et al. [6] did not use insertion error as corpus-based error rules because contextual information is indispensable for applying such rules. Instead, only insertion of articles was involved in the generic error rules. In this paper, we address various insertion errors including article insertions considering POS (parts-of-speech) of the words before and after the insertion words.

It is necessary to determine the POS of all words that appeared in the SST corpus to extract the insertion error rules. The first ones are extracted from transcriptions of English utterances by native Japanese speakers. The SST (Standard Speaking Test) corpus was used as a source of the error rules [7]. The SST corpus consists of transcriptions of interviews with native Japanese speakers in English. The grammatical and lexical errors in this corpus are manually annotated with the correct expressions. Table I shows the errors most frequently observed in the corpus. The symbol $\epsilon$ denotes the absence of a corresponding word. The error types DEL, INS and SUB indicate a deletion error, an insertion error and a substitution error, respectively. The second ones are manually created to cover a wider range of error types. These error rules include insertion of articles, confusion between singular and plural, adjective comparative and verb tense errors.

On applying these two kinds of rules, the generic error rules were first applied, and the corpus-based rules were applied next. The probabilities of choosing a corpus-based rule are in proportion to the frequencies of the errors in the SST corpus. Conversely, the probabilities of applying a generic error rule are given a priori. The final recognition performance was not greatly affected by the probabilities of the generic error rules. In the corpus-based rules, only substitution and deletion error rules were applied to one word. However, errors applied to more than one word (e.g., Correct: office worker, Uttered: work man) and insertion errors also exist in the SST corpus. Therefore, the error rules concerning multiple words should be treated properly.

### III. Revision of Sentence Generation Rules

#### A. Contraction rules

In the previous work, generic error rules include contraction rules such as “I am $\rightarrow$ I’m” or “He’s $\rightarrow$ He is”. However, for a contraction is not a mistake, it is inappropriate to regard a contraction as a grammatical error. Therefore, we generate sentences with and without contractions, and all of the generated sentences are used as new target sentences.

#### B. Corpus-based rules

Ito et al. [6] did not use insertion error as corpus-based error rules because contextual information is indispensable for applying such rules. Instead, only insertion of articles was involved in the generic error rules. In this paper, we address various insertion errors including article insertions considering POS (parts-of-speech) of the words before and after the insertion words.

It is necessary to determine the POS of all words that appeared in the SST corpus to extract the insertion error rules. The first ones are extracted from transcriptions of English utterances by native Japanese speakers. The SST (Standard Speaking Test) corpus was used as a source of the error rules [7]. The SST corpus consists of transcriptions of interviews with native Japanese speakers in English. The grammatical and lexical errors in this corpus are manually annotated with the correct expressions. Table I shows the errors most frequently observed in the corpus. The symbol $\epsilon$ denotes the absence of a corresponding word. The error types DEL, INS and SUB indicate a deletion error, an insertion error and a substitution error, respectively. The second ones are manually created to cover a wider range of error types. These error rules include insertion of articles, confusion between singular and plural, adjective comparative and verb tense errors.

On applying these two kinds of rules, the generic error rules were first applied, and the corpus-based rules were applied next. The probabilities of choosing a corpus-based rule are in proportion to the frequencies of the errors in the SST corpus. Conversely, the probabilities of applying a generic error rule are given a priori. The final recognition performance was not greatly affected by the probabilities of the generic error rules.

In the corpus-based rules, only substitution and deletion error rules were applied to one word. However, errors applied to more than one word (e.g., Correct: office worker, Uttered: work man) and insertion errors also exist in the SST corpus. Therefore, the error rules concerning multiple words should be treated properly.
output probability of error words. The error rules are applied word by word in the following order.

- Insertion errors considering POS context
- Multiple-word substitution errors
- Corpus-based substitution and deletion errors
- Generic error rules or thesaurus-based word substitution rules

If a word is as is after the corpus-based error rules are applied, generic error rules or thesaurus-based word substitution rules are applied according to the probability of applying thesaurus-based word substitution rules.

IV. EXPERIMENT

A. Data collection

To carry out an experiment under the condition similar to the assumed CALL system, we collected utterances by Japanese native speakers as follows.

1) English sentences (the target sentences) containing key-
words were prepared.
2) Learners were asked to remember the target sentences. No time limitation was set for the memorization task.
3) Japanese translations of the target sentences were pre-
tened to the learners, and the learners were asked to

Table IV shows an example of uttered sentences corresponding the target sentence “I’m an office worker. I work at a car company.”

B. Effect of the contraction rules

We carried out an experiment to examine the effect of revision of the contraction rules. In this experiment, we use two sets of target sentences: one contains the original target sentences, and the other one includes the sentences generated by the contraction rules. The system performance was measured using the word accuracy of the recognized sentences. The test data consisted of 126 sentences spoken by 15 students, each of which includes contraction or contractible expression. Table V shows the experimental conditions.

Word accuracy using the sentences generated by the proposed contraction rules was 84.88%, while that of the conventional method was 83.46%. The performance of the proposed method that used multiple target sentences was better than the method where the contraction rules were included in the generic error rules.

C. Effect of revision of error rules

Next, we investigated the effect of revision of other types of error rules: the corpus-based rules, the generic rules and the thesaurus-based rules.

According to the experiment, word accuracy was not very affected by the revision of the corpus-based and generic rules. However, 6 words decrease in OOV was observed by applying the corpus-based rules because of the insertion rules and the multiple word substitution rules.

Next, we examined the effect of thesaurus-based word substitution rules upon word accuracy and the number of OOV. The test data included 441 sentences spoken by 15 students, and other experimental conditions are same as shown in Table V. Applying the thesaurus-based word substitution resulted in decreasing word accuracy by 0.2. However, the number of OOV decreased by 4 words.

D. Comparison of total performance

Next, we compared the conventional error rules and the revised error rules. The performance was compared using word accuracy as well as recall and precision of grammatical error detection. The probability of applying thesaurus-based word
Next, we evaluated the same results from the aspect of recall and precision of error detection. A detection result was regarded as a misdetection when any of the following were wrong: word position, recognized word, or type of error. The relationship between the recall and precision rate is shown in Fig. 4. The results showed that the revised rules produced better performance than conventional rules.

V. CONCLUSIONS

We have developed a method to recognize English utterances spoken by Japanese for development of a voice-interactive CALL system. We prepared target sentences with and without contractions. Then we applied the revised error rules, which involved insertion errors or rules applied to multiple words. The generated text using the proposed method improved the recognition performance and recall/precision ratio. Although we could improve the performance of recognition and error detection, the absolute performance of the method is still low for applying an actual CALL system. We need further improve ment of the performance of the system by using more corpus data and improving the error rules.

ACKNOWLEDGMENT

The present study was supported in part by a JSPS Grant-in-Aid for Scientific Research (B) 20320075.

REFERENCES