Tracking Verb Phrases for Formative Feedback in Foreign Language Writing

Shuai Shao, Kazuhiro Ohtsuki, Hidenari Kiyomitsu, Min Kang
Graduate School of Intercultural Studies
Kobe University
Kobe, Japan
e-mail: samveldf@stu.kobe-u.ac.jp, ohtsuki@kobe-u.ac.jp,
kiyomitsu@carp.kobe-u.ac.jp, kang@kobe-u.ac.jp

***Abstract*—Providing feedback is crucial in the language learning process. In time, formative feedback can help both learners and teachers confirm the ongoing acquisition of language. However, as of today, there is still little research not only on manual formative feedback but also on automated formative feedback. Thus, in this paper, we elaborate on how to track learners' acquisition for formative feedback by developing a system for foreign language writing. The system is implemented based on the results of an analysis of data collected from conventional face-to-face classrooms in Chinese learning.**

Keywords-foreign language writing; automated formative feedback; phrase extraction; dependency relation; change detection.

I. INTRODUCTION

Feedback plays an important role in foreign language learning [1]. The effectiveness of corrective feedback has been clarified by a number of researchers [2]-[4]. Meanwhile, formative feedback became an indispensable component as well [5]. Because providing feedback can be time consuming and costly, automated feedback has also drawn much attention [6].

Previous research has shown the advantage of the automated feedback system over the paper-based feedback [7]. Educational Testing Service (ETS) has developed a Web-based writing evaluation service, *Criterion*, and Vantage Learning has created *My Access*, which are both programs that combine a scoring engine with a separate editing tool, which provides grammar, spelling and mechanical feedback [8]. On the other hand, Warschauer and Grimes [8] conducted a mixed-methods case study to evaluate the use of those programs in classrooms. They pointed out that although the programs saved teachers' grading time and learners tended to edit their writings more, the editing was usually superficial and no iterative process was observed. These automated systems are just designed to improve the writing quality in the current document by finding errors, which is different from a teacher's goal which is to improve the learners' writing ability to produce better new documents [9]. Thus, research on the long-term usage of automated systems becomes a necessity.

Simone and Christian implemented a Web-based feedback system in their lectures and analyzed the effects of the system which provided automated formative feedback throughout the semester [10]. They found that students who received feedback achieved higher scores and became more motivated and confident. Formative feedback can not only help learners but also help teachers improve their instructional strategies [11]. Nevertheless, additional research on automated formative feedback in foreign language writing is still rare.

The aim of our research is to design an automated formative feedback environment to facilitate the writing process. In this paper, we report the partial results concerning the on-going research. We focus on the writing process in Chinese for Japanese learners, especially on Japanese-Chinese translation process. Furthermore, verb-object (V-O) phrases are chosen as the targets of providing feedback because V-O phrases are basic sentence structures expressing the meanings of sentences and appear frequently in teaching materials for beginners.

In Section 2, we look at the V-O phrases in several translation exercises conducted in face-to-face classrooms and first analyze the translations manually to observe how learners translate the corresponding Japanese phrases to the Chinese phrases in time-series. In Section 3, based on the results of the analysis, we build a prototype system to track changes of learners' translations concerning the phrases and provide this to learners to help them confirm their progression in the learning process. The tracking results will also be provided to teachers to give them an overview of the learners' acquisition of the material. The conclusion will be given in Section 4.

II. ANALYZING LEARNING LOG DATA

A. Data from Classrooms

68 sophomore students (2 classes of 34 students) taking "Intermediate Chinese" at Kobe University, Japan, whose overall Chinese proficiency level was empirically considered to be intermediate were subjects for this research. The students were asked to translate Japanese sentences into Chinese as a class exercise every week. One specific word "花見" (hanami or cherry-blossom viewing) is chosen as the target to providing feedback. We designed three Japanese sentences containing the word "花見" (hanami or cherry-blossom viewing) and put them respectively into three exercises over eight weeks: the interval between the first two exercises was one week, and the interval after the 2nd exercise was six weeks. In the 1st week, the Chinese translation of the phrase "花見に行く" (go to see cherry blossoms) was presented as a hint along with the exercise paper for Class 1 while not Class 2. In the following week,

 ISBN: 978-1-61208-619-4

the students from both classes did the 2nd exercise without any hint. Then in the interim, in the 3rd week, the teacher thoroughly explained about the various translations of "cherry-blossom viewing" and told the students of both classes that "看樱花" (see cherry blossoms) was the most appropriate answer. Five weeks later, the 3rd exercise containing "cherry-blossom viewing" was conducted. The three Japanese sentences are listed below.

S1. "もし明日雨が降らなければ，私たちは花見に行くつもりです．"

(If it doesn't rain tomorrow, we are going to see cherry blossoms.)

S2. "もし花見に行くなら，京都が一番いい．"

(If you go to see cherry blossoms, Kyoto is the best place.)

S3. "来年 3 月末に私は神戸に来る予定だが，花見に来るのではなく，出張に来るのだ．"

(I plan to come to Kobe at the end of March next year for business trip not for cherry blossom viewing.)

B. Analysis and Results

TABLE I. CORRECT ANSWER RATE OF "花見" (CHERRY-BLOSSOM VIEWING)

	Week 1	Week 2	Week 8
Class 1	100%	94.1%	94.1%
Class 2	73.5%	88.2%	100%

TABLE II. PERCENTAGE OF STUDENTSCHANGING ANSWERS BETWEEN EXERCISES

	Week 1- 2	Week 2- 8	Week 2- 8(G2-G1)
Class 1	85.3%	64.7%	23.5%
Class 2	52.9%	64.7%	23.5%

Our analysis focuses on the changes of translation of the specific word "cherry-blossom viewing" found in all three exercises. There were four main translations, "看樱花" (see cherry blossoms), "看花" (see flowers), "赏花" (admire flowers), and "观赏樱花" (admire cherry blossoms). Although "cherry-blossom viewing" is a word in Japanese, it should be translated as a verb phrase in Chinese, with one verb and one noun. The Japanese word "花見" (cherry-blossom viewing) refers to the tradition of sitting under blooming cherry trees to appreciate the beauty of the cherry blossoms. Thus, even though the kanji/Chinese character "花" (flower) exists in both Japanese and Chinese, in the original Japanese word it specifically refers to cherry blossoms. However, in translations such as "看花" (see flowers), "赏花" (admire flowers), "花" (flower) means flowers without explicitly referring to cherry blossoms. Hence "樱花" (cherry blossoms) is considered as a more appropriate translation. In addition, all three sentences come from everyday conversations, "观赏樱花" (admire cherry blossoms) seems too formal in this context. Therefore, we divided the different translations into three groups: Group 1 (G1: most appropriate): "看樱花" (see cherry blossoms); Group 2 (G2: correct but flawed): "看花" (see flowers), "赏花" (admire flowers), and "观赏樱花" (admire cherry blossoms), as well as Group 3 (G3: mistakes). We then calculated the percentage of correct answers for each exercise and also the percentage of students changing answers overtime according to descriptive statistics methods.

Table 1 shows the percentage of correct answers (G1 & G2) of the word "花見" (cherry-blossom viewing) in the three exercises. As we can see, students in Class 1 achieved 100% accuracy because of the hint, however, Class 2 only achieved 73.5%. However, it is noteworthy that in the following week, the accuracy of Class 1 fell while that of Class 2 increased. In week 3, the teacher explained about the exercises conducted previously and emphasized the most appropriate translation. In week 8, the accuracy of Class 2 exceeded that of Class 1, which suggested that giving students answers without any explanation was not as effective as one might think. This kind of input may lead students to just use the answer without any active thinking or reflection involved.

Table 2 shows the percentage of students who changed answers between the exercises. In week 1, all students in Class 1 used the most appropriate translation owing to the hint. However, 85.3% of Class 1 changed their answers in week 2, which indicated that the hint had not been properly memorized. In week 2, the percentage of G1 was 14.7% in Class 1 and 8.8% in Class 2. The percentage of students changing answers in both classes between exercise 2 and 3 were identical, and there were over 20% of students in each class who changed their answers from G2 to G1. These percentages reveal that students' self-reflection can improve their accuracy but explanations by a teacher can further facilitate the learning process.

Based on the above results, it is revealed that by tracking the changes of translation, teachers could confirm the effects of the provided hint and explanation; students may benefit from the formative feedback to find out the weak points in the learning process.

III. SYSTEM DESIGN AND OUTPUT ANALYSIS

A. Formative Feedback System

We propose an approach to provide feedback for learners' time-serial data in Japanese-Chinese translation process, with a focus on tracking changes on V-O phrases. The key idea in the approach is the utilization of the dependency relation between two words that consist of a phrase for tracking. By using a dependency parser, we can obtain the structural information of input sentences in which the phrase should be contained; based on that information we can extract the phrase within and then detect whether a learner has changed the phrase or not by comparing the extracted phrase with that from the sentences in previous translation exercises. If the phrase cannot be extracted, there are two possible reasons. One is that the learner used an incorrect phrase. The other is that the learner used a different correct phrase with different dependency relation.

 ISBN: 978-1-61208-619-4

Currently, the approach is presented for the simple sentences with one V-O phrase. The approach can be divided into two phases, as shown below.

1) The preparation for, and extraction of verb phrase:

a) A teacher or a learner chooses the V-O phrase (We call it intended phrase or IP.) to which he/she wants to confirm the acquisition.

b) Learners' translations which should contain the IP (based on the source language sentence) will be processed by a Chinese parser and the V-O phrase (We call it learner's phrase or LP.) will be extracted based on the dependency parser's result.

2) The formative feedback:

a) The LP will be extracted along with the information of the time when it was submitted (timestamp). As a result, extracted LPs will be in time series, and LP submitted later can be compared with the previous LP.

b) The extraction and comparison will provide not only the information of the phrases but also the detection of whether the learners have changed their translations or not. Subsequently, the results of all the exercises will be presented to both learners and teachers.

B. Algorithm for Prototype System

Following the above approach, we built a prototype system. In this system, we utilized the Stanford Parser [12] through Python NLTK (Natural Language Toolkit) interface to analyze the input data. The input data should be simple sentences which are produced by learners in an exercise and should contain an IP. The steps of the algorithm for the prototype system are listed below.

1) The input will be segmented and the part of speech (POS) information will also be generated by exploiting the segmentor and POS tagger of the Stanford Parser.

2) The dependency parser of the Stanford Parser will provide the structual information of the segmented input and the LP within will be extracted depending on the existense of a "dobj" (direct object) tag. If a "dobj" tag exists, the contents, as well as their POS tags will be extracted, otherwise the output will be "*".

3) Since the extracted LPs will have timestamps, the system will compare the latter LP with the former one to detect whether the learner has changed the translation or not .

C. Output Analysis and Discussion

In order to evaluate the approach, we used the previously collected translations from Class 1 as our testing data. We also chose the V-O phrase "看樱花" (see cherry blossoms) as the target to provide feedback as described in the Section 2. The translations for S1, S2 and S3 are the raw data to input into the system. Because the raw data was compound sentences, we did manual pre-processing to obtain the input sentences which should contain the IP. Examples are shown in Table 3.

TABLE III. EXAMPLES OF PRE-PROCESSED INPUT

Original	如果去看樱花，京都是最好。 (If going to see cherry blossoms, Kyoto is the best.)
	如果我去看樱，京都最好。 (If I go to see cherry, Kyoto is the best.)
Pre-processed	如果去看樱花 (If going to see cherry blossoms)
	如果我去看樱 (If I go to see cherry)

TABLE IV. EXAMPLES OF EXTRACTED PHRASES

No.	verb	vPOS	object	oPOS	V&O
S1	去 (go)	VV	看樱 (see cherry)	NN	去看樱 (go to see cherry)
S1	看 (see)	VV	樱花 (cherry blossoms)	NN	看樱花 (see cherry blossoms)
S2	看 (see)	VV	樱花 (cherry blossoms)	NN	看樱花 (see cherry blossoms)
S2	*	*	*	*	**

StudentNumber	VO in S2	Detection	VO in S8	Detection
Student1	去看樱	changed	看樱花	ok
Student2	看樱花	ok	**	changed
Student3	去赏花	changed	看樱花	ok
Student4	看樱花	ok	来看樱	changed
Student5	赏樱花	changed	来赏花	changed

Figure 1. Examples of change detection

TABLE V. ABOUT EXTRACTION OF THE SYSTEM

	Extraction Rate of All Inputs	Usage of "看樱花"
Week 1	100%	100%
Week 2	100%	14.7%
Week 8	76.5%	38.5%

There are two kinds of outputs provided by the prototype system that are shown in Table 4 and Figure 1. Table 4 shows some examples of the output obtained by the step 1) of the algorithm. If a LP was extracted, then the verb and object, as well as the V-O phrase will be provided to users. It can be observed that not all inputs can be extracted with a V-O phrase. Just like the last example in the table, if the system couldn't find a "dobj" tag, then the output would be "*".

For examining the validity of the feedback in Table 4, we calculated the extraction rate that describes how many V-O phrases there are in the input sentences. The extraction rates are presented in Table 5. Meanwhile, we calculated the percentages of the IP "看樱花" (see cherry blossoms) used by students in the raw data and showed the percentages in the same table. From Table 5 it is clear that in week 1, all students of Class 1 translated "花見" (cherry-blossom viewing) into the IP "看樱花" (see cherry blossoms) because of the hint. Consequently, all inputs were successfully extracted. Apart from "看樱花" (see cherry blossoms), other variations were also extracted in week 2 and week 8, as long as the input contained a V-O phrase. In week 2, although every input contains a V-O phrase, the usage of the IP decreased to 14.7%. Thus, if students have grasped the

 ISBN: 978-1-61208-619-4

basic sentence structure, e.g. the V-O structure, all LP would be extracted and Table 4 would provide teachers a visual feedback to confirm what different phrases or wrong phrases are used by students. On the other hand, the extraction rate in week 8 was only 76.5%. In the case, this results from that the two-character words in G2: "看花" (see flowers) and "赏花" (admire flowers) were treated as nouns instead of V-O phrases in the system. It remained a problem when automatically dealing with the 24.5% sentences without a "dobj" tag.

Figure 1 shows examples of some change detection outputs. With the input parsed and extracted, the system then compared the extracted LP of the latter two inputs (S2, S3) with the first one (S1) respectively and generate the outputs showing whether students had changed their translations or not. If a student did not change the translation, then "ok" will be shown in the "Detection" column, otherwise it will be "changed".

According to the results in Figure 1, there are 85.3% of the LPs that have been changed in week 2 and 73.5% that have been changed in week 8 in comparison with the phrases used in week 1. Despite the inputs with no extracted V-O phrases, the percentages are agreements with those manually calculated from the raw data.

Thus, Figure 1 can provide teachers an overview of students' progress in learning process. It can be readily noted that there is only one student who did not changed the translation in all three exercises. There were 7 students who changed their answer in week 2 but then changed back to the most appropriate "看樱花" (see cherry blossoms) in week 8, which demonstrated the effectiveness of the teacher's detailed explanation in week 3. The other 26 students failed to change back to the correct translation they submitted in week 1. This information can help teachers improve instructional strategies, and facilitate individual students to comprehend whether the required grammatical knowledge had been mastered or not as well.

IV. CONCLUSION

In this paper, we first analyzed learning log data from two face-to-face classrooms in Chinese learning. The analysis results revealed that tracking the changes of translation on V-O phrase could help teachers confirm the effects of the provided hint or explanation; and students may benefit from the formative feedback to find out the weak points in the learning process. Thus, we proposed an approach for providing formative feedback and developed a prototype system to test the approach.

It is suggested that the system is effective in providing automated formative feedback for both learners and teachers. Learners can confirm their acquisition throughout the learning process. The feedback on V-O phrases would help teachers grasp the whole picture used by learners and confirm the effects of the current strategies. Because the system focuses on the extraction and comparison of V-O phrases by using the Stanford Parser, thus it is expected to be applied to other languages as long as similar structures can be identified by the parsers.

There still remained some problems in the approach. In the prototype system, the input sentences should be simple sentence so that a pre-processing is needed. Developing an automatically pre-processing function will be our next work soon. The extraction method still needs improvement. As we have explained, it is an important issue to deal with the phrases without a "dobj" tag. In addition, further practical use in classrooms needs to be investigated.

ACKNOWLEDGMENT

This work is supported by JSPS KAKENHI Grant Number JP17K01081.

REFERENCES

[1] H. Kimura, T. Kimura, and O. Shiki, Theory and Practice in Reading and Writing: Nurturing Independent Learning, pp. 124–134, 2010.

[2] P. Duppenthaler, "The effect of three types of feedback on the journal writing of EFL Japanese students," JACET Bulletin, (38), pp. 1-17, 2004.

[3] K. Oi, T. Kamimura, T. Kumamoto, and K. Matsumoto, "A Search for the Feedback That Works for Japanese EFL Students: Content-based or Grammar-based," JACET Bulletin, (32), pp. 91-108, 2000.

[4] A. M. F. Yousef, U. Wahid, M. A. Chatti, U. Schroeder, and M. Wosnitza, "The Impact of Rubric-Based Peer Assessment on Feedback Quality in Blended MOOCs," Communications in Computer and Information Science Computer Supported Education, pp. 462–485, 2016.

[5] I. Goldin, S. Narciss, P. Foltz, and M. Bauer, "New Directions in Formative Feedback in Interactive Learning Environments," International Journal of Artificial Intelligence in Education, vol. 27, no. 3, pp. 385–392, 2017.

[6] P. D. Ware and M. Warschauer, "Electronic feedback and second language writing," Feedback in second language writing: Contexts and issues, pp. 105-122, 2006.

[7] S. W. Yeh and J. J. Lo, "Using online annotations to support error correction and corrective feedback," Computers & Education, 52(4), pp.882-892, 2009.

[8] M. Warschauer and D. Grimes, "Automated writing assessment in the classroom," Pedagogies: An International Journal, 3(1), pp.22-36, 2008.

[9] C. Leacock, M. Chodorow, M. Gamon, and J. Tetreault, "Automated grammatical error detection for language learners," Synthesis lectures on human language technologies, 7(1), pp.109-112, 2014.

[10] S. van Kol and C. Rietz, "Effects of Web-Based Feedback on Students' Learning," International Journal of Teaching and Learning in Higher Education, 28(3), pp.385-394, 2016.

[11] K. Ludvigsen, R. Krumsvik, and B. Furnes, "Creating formative feedback spaces in large lectures," Computers & Education, 88, pp.48-63, 2015.

[12] C. D. Manning, et al, "The stanford corenlp natural language processing toolkit," In ACL (System Demonstrations) pp. 55-60, 2014.

 ISBN: 978-1-61208-619-4