A Verb Phrase Tracking System 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, both manual and automated can help both learners and teachers confirm the ongoing acquisition of language. However, little research has been done on either manual or automated formative feedback in actual foreign language classroom. 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. We have previously reported on a part of the system, and in this paper we extend our study, especially concerning the pre-processing algorithm.**

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

I. INTRODUCTION

In this paper, we propose a verb phrase tracking system for formative feedback based on our previous work [1].

Feedback plays an important role in foreign language learning [2]. The effectiveness of feedback has been clarified by a number of researchers [3]-[5]. It is considered that effectiveness depends on how it is best delivered. Feedback type, feedback timing etc. are important factors which have been shown to influence feedback performance [6, 7]. Recently, formative feedback has been seen as an indispensable component as well [8, 9]. According to Shute [8], formative feedback is defined as information communicated to learners which supports the learning process and enhances learning. It consists of a variety of different types and is managed at various times during the learning process. Although there are many advantages, providing feedback can be time consuming and costly in classroom learning. Gradually, automated feedback has drawn much attention [10].

Previous research has shown the advantages of an automated feedback system over the paper-based feedback [11]. Educational Testing Service (ETS) has developed *Criterion*, a Web-based writing evaluation service, and Vantage Learning has created *My Access*, both programs combine a scoring engine with a separate editing tool, which provides grammar, spelling and mechanical feedback [12]. Source language has also been taken into consideration to generate feedback based on error detection [13]. Warschauer and Grimes [12] have evaluated the use of *Criterion* and *My Access* through a mixed-methods case study 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 of 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 documents over time [14]. Thus, research on the long-term usage of automated systems to improve writing ability becomes a necessity.

Simone and Christian implemented a Web-based feedback system in their lectures and then analyzed the effects of the system which provided automated formative feedback throughout the semester [15]. They found that the students who received feedback achieved higher scores and became more motivated and confident. Computer-based formative feedback not only helps learners but also assists teachers in improving their instructional strategies [16]. AI techniques have already been used in intelligent tutoring systems to understand formative feedback, and a lot of the researches focus on how to build general models or frameworks of systems [9, 17]. McNamara, Crossley and Roscoe proposed an excellent intelligent tutoring system, the Writing Pal, based on a natural language processing (NLP) algorithm, to support adolescent native writers in English [18], but additional research on automated formative feedback in foreign language writing is still rare [19].

The aim of our research is to design an automated formative feedback environment which will facilitate the writing process. In this paper, we report on our efforts to further develop a system for formative feedback in classroom learning. We focus on the writing process in Chinese for Japanese learners, especially on the Japanese-Chinese translation process. Verb-object (V-O) phrases are chosen as the targets of feedback because V-O phrases are basic sentence structures expressing the meanings of sentences and appear frequently in teaching materials for beginners.

In Section II, we look at the V-O phrases used in several translation exercises conducted in face-to-face classrooms. We first analyze the translations manually to see how learners translate the corresponding Japanese phrases to the Chinese phrases in time-series. In Section III, based on the results of the manual analysis, we propose an approach to provide feedback for learners' time-series data in the Japanese-Chinese translation process. In Section IV, we elaborate the methodology in details about how to track changes of

learners' translations concerning the phrases. The results will be provided to teachers to give them an overview of the learners' acquisition of the material. In Section V, we will analyze the output of our system to examine the validity. The discussion and conclusion will be given in Section VI and Section VII.

II. Analyzing Learning Log Data

In this section, we analyzed the translations collected from face-to-face classrooms to observe how the learners translate V-O phrases under different circumstances.

A. Data from Classrooms

The subjects for this research were 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. All students had taken two levels of "Basic Chinese" in the 1st academic year. They learned the basic pronunciation and basic grammar, such as the sentence structure of simple sentences, the use of auxiliary verbs and prepositional phrases in simple sentences, etc. They also mastered some basic knowledge about compound sentences and complex sentences.

In "Intermediate Chinese" in the 2nd year, the students are required to understand relatively complicated sentences with two or more clauses. An important part of the course is the ability to use conjunctions and specific sentence structures in writing sentences with two or more clauses. Weekly composition exercises are given to assess the students' use of conjunctions and specific sentence structures learned.

With respect to this research, the students were asked to translate Japanese sentences into Chinese as a class exercise every week. One specific word "花見" (cherry-blossom viewing) was chosen as the target to provide feedback. We designed three different Japanese sentences containing the word "花見" (cherry-blossom viewing) for three exercises to be given 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 for Class 1 but wasn't given for Class 2. In the following week, the students from both classes did the 2nd exercise without a hint. Then in the interim, during 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 and the reference translations in Chinese are listed below.

S1.

Japanese: "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．"

Chinese: "如果明天不下雨，我们就去看樱花。"

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

S2.

Japanese: "もし花見に行くなら，京都が一番いい．"

Chinese: "如果去看樱花，京都是最好的。"

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

S3.

Japanese: "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．"

Chinese: "明年 3 月底我打算来神户，但不是为了来看樱花，而是来出差。"

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

In fact, the three Japanese sentences were composition exercises mentioned above. S1 and S2 are provided to help students understand the use of the conjunction "如果…，就…" (if…, then…), and the target of S3 is the sentence structure "不是…而是…" (for… not for…).

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 Students Changing 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 variations, "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 generally 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 over time according to descriptive statistics methods.

Table I 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, in contrast, 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 II 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 improves their accuracy but explanations by a teacher can further facilitate the learning process.

Based on the above results, it is suggested that by tracking the changes of translation, teachers will be able to confirm the effects of the hints and explanations provided; students should benefit from the formative feedback which enables them to assess their weak points in the learning process.

III. System Design

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

This approach can be divided into two phases, as shown below.

Preparation and extraction of verb phrase:

A teacher chooses a V-O phrase, we call it the intended phrase (IP), which is used to confirm the appropriate acquisition.

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

Formative feedback:

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

This extraction and comparison will provide not only information about the phrases, but also detect whether the learners have changed their translations or not. Subsequently, the results of all the exercises will be reviewed by the teacher.

IV. Methodology

Following the above approach, we implemented a system with four stages: 1) pre-processing for extracting simple sentences, 2) segmentation and part of speech (POS) generation of extracted simple sentences, 3) V-O phrases extraction based on dependency parsing and 4) comparison between extracted V-O phrases. This system utilizes the Stanford Parser [20] through Python NLTK (Natural Language Toolkit) interface to analyze the input data, while the system interface is based on PHP.

A. Overview of the System

Figure 1 shows an overview of the system. Currently, the system is mainly designed for teachers. The elements depicted by solid lines are completed, while those in dash lines are still in development.

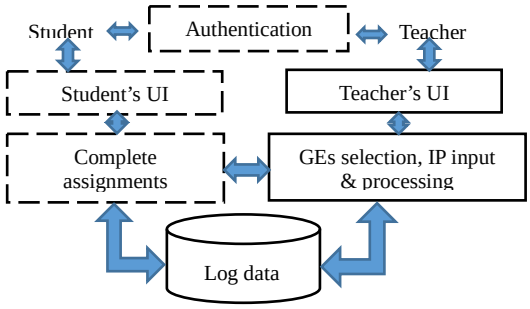

Figure 1. Overview of the system

In Student's UI, students can input answers for assignments and receive feedback generated by the system. The "Complete assignments" is designed to deal with data from students and save it as log data. In Teacher's UI, teachers can select the assignments that they want to check and input corresponding reference answers. The "GEs" in "GEs selection, IP input & processing" means "grammatical elements". In this part, teachers can choose GEs and input IPs. LPs will be extracted automatically and saved as log data. The details will be described in the next sub section.

B. Algorithm

The pre-processing in stage 1) is designed to extract a single clause with the target V-O phrase from a sentence with several clauses. As we indicated in Section II, students in the "Intermediate Chinese" course are required to master how to

use conjunctions or specific sentence structures for composing compound sentences, complex sentences and compound-complex sentences. To help them consolidate their knowledge, all the exercise sentences previously introduced consist of two or more clauses and many may contain multiple V-O phrases. This makes it difficult for the system following the stage 2) to automatically identify the phrase the teacher chose to confirm acquisition of. Therefore, it is necessary to pre-process the sentences to ensure that the system can parse and extract the clause containing the LP. In other words, the input data for stage 2) should be a single clause which is produced by learners in an exercise and should contain an IP.

1) Pre-processing for extracting simple sentences:

It is important to be able to automatically analyze sentence structures in natural language processing. Parsers developed for a variety of languages have made it easy to separate a sentence with plural clauses into single clauses. However, analyzing a sentence, even a simple sentence with errors automatically, continues to be a challenging issue. Although many approaches have attempted to solve this problem, most of the approaches deal only with English and few addressed Chinese sentence structures. In addition, most of the approaches are based on learner corpora and few learner corpora on Chinese have been reported. Hence, through an empirical observation of teachers, we present a method based on punctuation and the grammatical elements such as conjunctions, adverbs and flag words for specific sentence structures to separate the target sentences. For sentences with plural clauses composed by students, we first try a punctuation position step and then a grammatical element step.

Empirically, students at the intermediate level tend to grasp grammar that they just learned in the classroom very well but easily make mistakes on what they have learned after several weeks or months. Because the main grammar elements that students learn in the "Intermediate Chinese" are conjunctions and specific sentence structures, there are few errors related to these but the same does not hold true for previously learned preposition phrases or VO phrases etc. This means that most students can master the use of conjunctions when they have translation assignments. Therefore, our sentence separation in the empirical observation is restricted.

For sentences by students that contain two clauses, the sentences generally consist of two clauses separated by punctuation. As a result, it is extremely possible that the position for the punctuation in the sentence is correct. In this case, it is easy to extract a single clause with the IP according to the punctuation position. We call this the punctuation position step. Table III shows the percentages with a correct punctuation position in the sentences composed by the students in the "Intermediate Chinese" course.

In the punctuation position step, a reference answer is used to determine the position of the single clause that contains the IP. For example, the reference answer of S1 is "如果明天不下雨，我们就去看樱花。" (If it doesn't rain tomorrow, then we are going to see cherry blossoms.) Since the IP is "看樱花" (see cherry blossoms), the desired clause will be the latter part of the whole, e.g., "我们就去看樱花" (then we are going to see cherry blossoms). Based on this information, we can extract the single clause from sentences by students by initially finding the comma, and then retrieving the latter part of the sentence.

TABLE III. PERCENTAGES OF STUDENTS WITH CORRECT PUNCTUATION AND GRAMMATICAL ELEMENTS

	Correct Punctuations		Correct GEs		Correct Subject	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
S1	100%	91.2%	14.7%	5.9%	88.2%	76.4%
S2	100%	100%	100%	100%	91.2%	88.2%
S3	82.4%	70.1%	85.3%	79.4%	100%	97.1%

For sentences by students that consist of two or more clauses without punctuation or with plural punctuation, we adopt a grammatical element step. In this step, instead of relying entirely on the position of the IP in the reference answer, we add conditions based on grammatical elements such as conjunctions, adverbs or flag words used in specific sentence structures just learned by students as explained above. The grammatical elements usually correctly appear in each student's translation. Thus, the grammatical elements should be considered as helpful factors when separating the sentences. In addition, similar to English, Chinese sentences usually contain subjects, which should also be regarded as an important factor to help us separate the sentences. Table III shows the percentages of the grammatical elements correctly used and the percentages of the subjects that appear in the sentences composed by the students in the "Intermediate Chinese" course.

In the grammatical element step, the grammatical element and the subject included in the clause with an IP in a reference answer are first determined according to both the grammar and the IP chosen by the teacher. The grammar is that which is needed to compose sentences with two or more clauses. The element could be a conjunction or an adverb. The Stanford Segmentor will segment the reference answer and the teacher can choose the grammatical elements concerning the grammar. By searching the IP in the reference answer, we can easily obtain the clause that contains the IP and its position in the reference answer so that we can determine the grammatical element included and the subject within. If there is no grammatical element or no subject in the clause, the grammatical element step will go to the end of the pre-processing stage.

Therefore, for sentences by students that consist of two or more clauses without punctuation or with plural punctuation, the part with the element or the subject is extracted as the clause that should contain the IP, otherwise the whole sentence will be used in the next stage. Figure 2 shows the flow of the step. In Figure 2, "RA" and "PI" refer to "reference answer" and "positional information" respectively. The sentence "如果明天不下雨，我们就去看樱花" (If it doesn't rain tomorrow, then we are going to see cherry blossoms.) is an example of a reference answer where the IP is "看樱花" (see cherry blossoms) and the grammatical elements are respectively "如果"(if) and "就"(then) that correspond to the grammar "如果…，就…"(if…, then…).

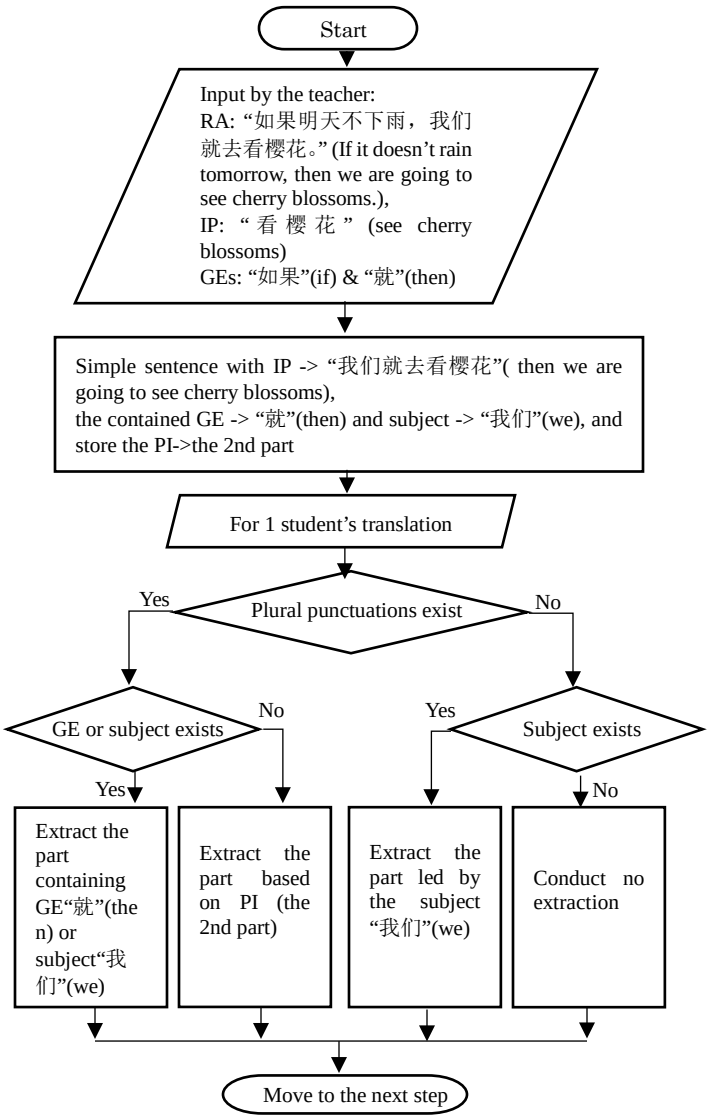

Figure 2. The flow of the grammatical element step of pre-processing for extracting simple sentences

Since "就" (then) should be in the clause with the IP, it becomes the key to extract the desired simple sentence in students' translations. In addition, the Stanford Dependency Parser is also used to parse the simple sentence from the reference answer, and if "nsubj" (nominal subject) exists, then the referred word will be extracted as a keyword to help locate the desired simple sentence. In this case, the subject "我们" (we) is also found and marked. After obtaining the information about the grammatical element and the subject, and the position of the simple sentence in the reference answer (the 2nd part), we can finally extract the desired simple sentence that should contain the IP in a student's translation.

The extraction algorithm first breaks down the students' translations into parts based on the punctuation. For translations which contain plural punctuation, we try to find whether the extracted grammatical element or subject exists. If one or both exist, e.g., "如果明天，不下雨，就去看樱花。" (If tomorrow, it doesn't rain, then go to see cherry blossoms.), then the part which contains the grammatical element "就" (then) will be extracted. If no grammatical element or subject exists in the translation, e.g., "如果明天，不下雨，观赏樱花。" (If tomorrow, it doesn't rain, admire cherry blossoms.), then the part based on the positional information (the 2nd part), which will be "不下雨" (it doesn't rain) is extracted. If the translation cannot be split but contains the extracted subject, e.g., "如果不下雨我们打算去观赏樱花。" (If it doesn't rain, we are going to admire cherry blossoms.), then we search for the extracted subject "我们" (we) and since the subject usually appears at the beginning of a clause, we extract the part led by the subject. Otherwise, we conduct no extraction and use the original translation in the following step.

2) Segmentation and POS generation:

The extracted simple sentence (e.g., single clause) will be segmented and the POS information will also be generated by exploiting the segmentor and the POS tagger of the Stanford Parser.

3) Dependency parsing and V-O phrase extraction:

The dependency parser of the Stanford Parser will provide the structual information of the segmented input and the LP within will be extracted if there is 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 "*".

4) Comparison between extracted V-O phrases:

Since the extracted LPs will have timestamps, the system will compare the later LP with the earlier to detect whether the learner has changed the translation or not .

C. System Practice

We first designed the interface to help teachers confirm students' acquisition of V-O phrases. In the pressent situation, students' translations have already been collected, and from these, we prepare an SQL database to store the translations. In order to evaluate the approach, we use the collected 204 translations from Class 1 and Class 2. We also choose the V-O phrase "看樱花" (see cherry blossoms) as the target to provide feedback as described in Section II. The translations of S1, S2 and S3 are the raw data to input into the system. The description of the user interface is shown as follows.

Step 1: Sentence selection and reference answer input

Figure 3 shows the first step of our system. As the translation assignments and students' answers have already been stored in the database, teachers need to choose a sentence or several sentences they want to check the acquisition of and then input the corresponding reference answers. In Figure 3, all the exercise sentences are selected so the teacher needs to input all the reference answers at this point, as shown in the lower part of the figure.

Step 2: Grammatical element selection and IP input

Figure 4 shows the second step. In this step, the reference answers have been segmented by the Stanford Parser, and the teacher chooses the grammatical element they taught in class when the assignment was given. The IP should be input at this step as well.

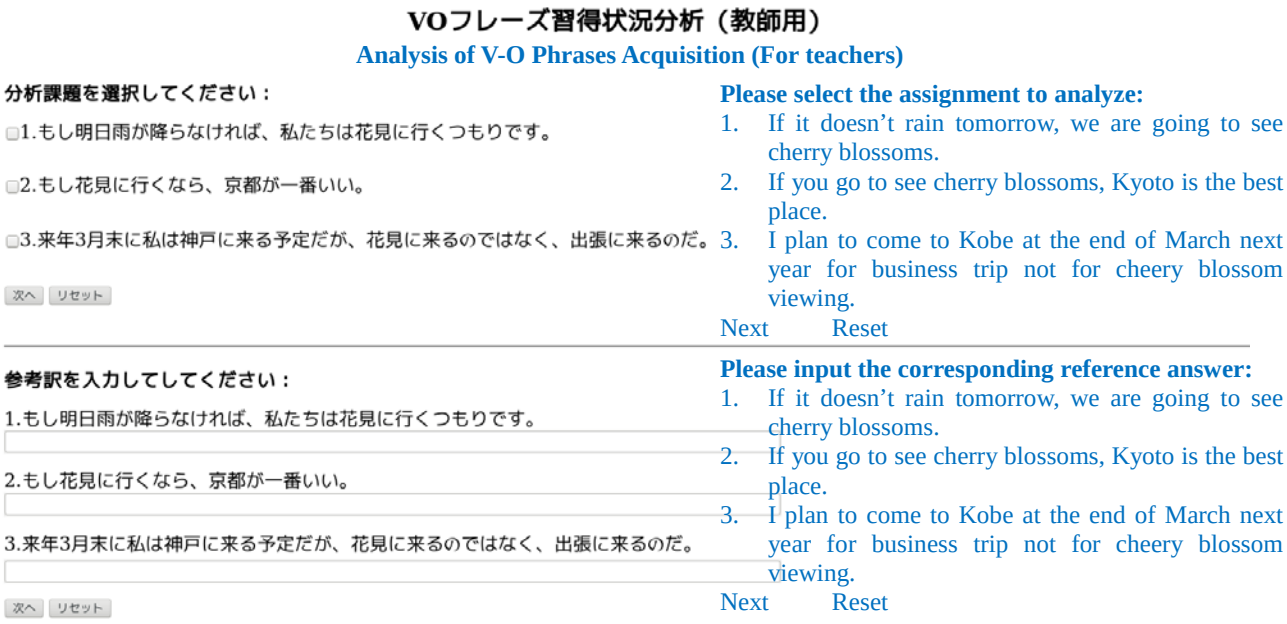

Figure 3. Sentence selection and reference answer input

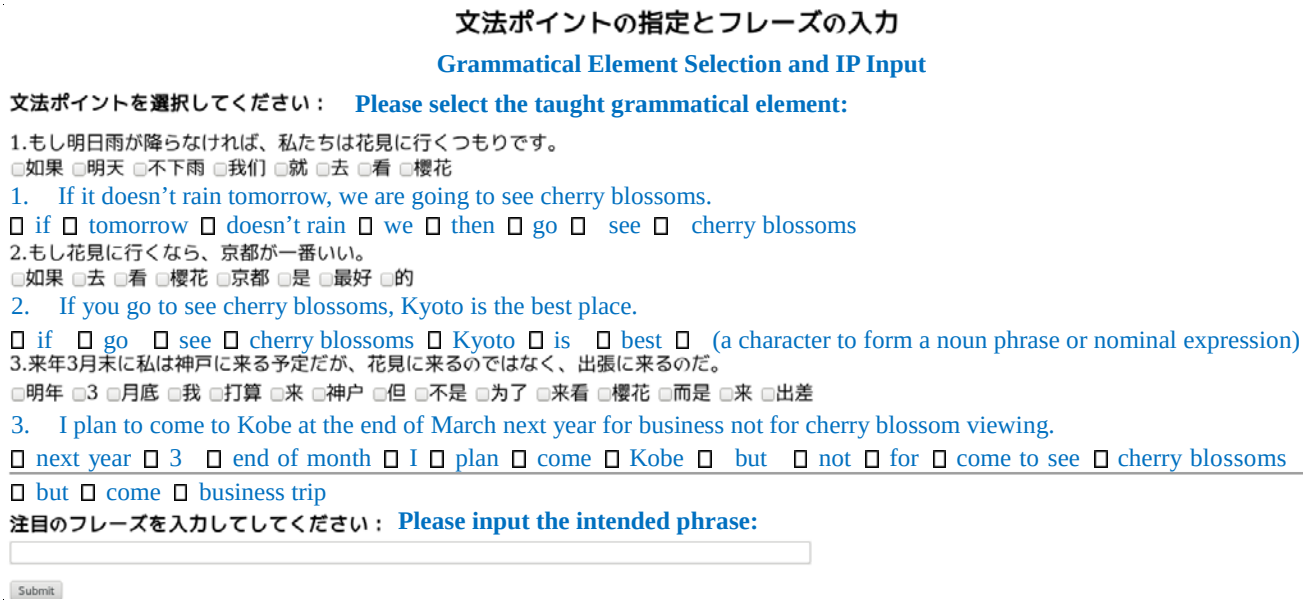

Figure 4. Grammatical element selection and IP input

Step 3: Simple sentence extraction

After completing the previous 2 steps, reference answers, their segmentation results, as well as the IP will be written to the database. With this information, the system first extracts the simple sentence containing the IP in the reference answer by splitting the sentence according to its punctuation into several parts and searching for the part that includes the IP. Then, the information and the position of the simple sentence are used to extract the simple sentence from the students' translations. Using this method, we correctly extracted 203 simple sentences out of 204 translations. Examples of the pre-processing results are listed in Table IV.

ID	課題1	VO in S1	課題2	VO in S2	変更(S1->S2)	課題3	VO in S3	変更(S1->S3)	変更(S2->S3)
1	如果明天雨不下，我们想去看樱花。	看樱花	如果我去看樱，京都最好。	去看樱	看樱花=>去看樱	明年3月底我打算来到神户，但不是为看樱花，而是为出差来。	看樱花	No	去看樱=>看樱花
2	如果明天不下雨，我们打算去看樱花。	看樱花	如果去看樱花，京都是最好。	看樱花	No	明年三月终我打算来到神户，为了不是赏花，而是出差。	**	看樱花=>**	看樱花=>**
3	如果明天不下雨，我们打算去看樱花。	看樱花	如果去赏花，京都是最好。	去赏花	看樱花=>去赏花	明年三月底我打算来神户，不是来看樱花，而是来出差。	看樱花	No	去赏花=>看樱花
4	明天不雨，我们还去看樱花。	看樱花	如果去看樱花，那京都是好很多。	看樱花	No	我打算来神户三月末，不是来看樱，而是来出差。	来看樱	看樱花=>来看樱	看樱花=>来看樱
5	如果明天没淋雨，我想去看樱花。	看樱花	如果去赏樱花，京都最好。	赏樱花	看樱花=>赏樱花	但是明年三月份我有予定来神户，不是来赏花，而是来出差。	来赏花	看樱花=>来赏花	赏樱花=>来赏花
6	如果明天不下雨，我们打算去看樱花。	看樱花	如果去赏樱花，京都最好。	赏樱花	看樱花=>赏樱花	明年3月末我将来神户，不过，不是来视赏樱花，来出差。	赏樱花	看樱花=>赏樱花	No

Figure 5. Examples of system output

TABLE IV. EXAMPLES OF SIMPLE SENTENCE EXTRACTION

Students' translations	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)

Step 4: V-O phrase extraction and change detection

Figure 5 shows some examples of our system's output. All the V-O phrases extracted from the students' translations, and the detection results of whether they changed their answers or not across all the three exercises are shown. If the extracted phrase is different from that in the reference answer, then it will be shown in red. If the student's answer changed between exercises, then the detection result will also be displayed in red. In this way, the teacher can easily notice the noteworthy parts as well as any changes over time.

The first column of Figure 5 shows the students' ID. For the protection of privacy, we replaced them with numbers. The students' translations of S1, S2 and S3 are shown in column 2, 4 and 7 respectively. In column 3, 5 and 8, the extracted LPs are displayed. Here, the symbol "**" means that no V-O phrase was extracted in the translation. In column 6, the information that whether the students changed their usage of the V-O phrase between S1 and S2 is shown here. If the system detected a change, then the contents of the corresponding cell will in the following format: "the extracted V-O phrase in S1=>the extracted V-O phrase in S2", otherwise the contents will be "No". Similarly, in column 9 and 10, the change detection results of whether the students changed their usage of V-O phrases between S1 and S3 as well as between S2 and S3 are shown.

From the information, the teacher can readily check the acquisition of the chosen V-O phrase. The system output not only shows whether the students used the intended phrase or not, but also provides the information on students' changes in the usage of V-O phrases.

V. OUTPUT ANALYSIS

Figure 5 shows the final system output stage. In order to reach that stage, pre-processing plays a significant role. Our pre-processing algorithm for extracting simple sentences achieved an overall 99.5% correct extraction rate, with only 1 out of 204 extractions incorrect. In contrast, before applying the algorithm, we had also extracted simple sentences from the same 204 translations only based on the punctuation position step. The result is shown in Table V.

TABLE V. CORRECT SIMPLE SENTENCE EXTRACTION RATE OF DIFFERENT METHODS

	Solely by position	Based on several conditions
S1	97.1%	100%
S2	100%	100%
S3	76.5%	98.5%

The absent comma resulted a null extraction for 2 students' translations of S1, which caused the correct extraction rate to be 97.1%. However, the correct extraction rate was only 76.5% in S3. S3 was the longest and most complicated sentence among the three. There are 2 commas in both the Japanese sentence and the reference answer. Due to this comparably complicated sentence structure, students often missed or added extra commas/periods, which caused the false extraction of simple sentences within those translations. The reference answer of S3 is "明年三月底我打算来神户，但不是来看樱花，而是来出差。" (Literal translation here to show the structure of the Chinese reference answer clearer: At the end of March next year I plan to come to Kobe, not to see cherry blossoms, but for a business trip.). Based on the IP "看樱花" (see cherry blossoms), the second part of the reference answer can be extracted. However, because of the mother-tongue interference, students tended to add a comma after the "明年三月底" (at the end of March next year), which resulted in an extraction of corresponding part of "我打算来神户" (I plan to come to Kobe) in the work of 14 students. And another 2 students added an extra comma to a different place.

The grammatical element step made the detected subject and grammatical element as key to extract, that perfectly solved the null extraction caused by absent punctuation. The only false extraction is from the following sentence: "明年 3 月末我将来神户，不过不是来，观赏樱花，来出差。" (I will come to Kobe at the end of next March, but not come to, see cherry blossoms, come for business trip.). In this sentence, the student added an extra comma between "不是来" (not come to) and "观赏樱花" (admire cherry blossoms). Since our system first searches for the taught grammatical element "不是" (not for), and in the reference answer and most students' answers, the V-O phrase follows without any

punctuation, the simple sentences were correctly extracted. However, in this student's translation, he made a rare error that separate the simple sentence and resulted in the false extraction.

TABLE VI. 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	*	*	*	*	**

The extracted V-O phrase shown in Figure 5 is provided to teachers. But before that, the components that form the phrase and their POS information were also generated and stored into the database. Table VI shows some examples. If a LP was extracted, then the verb and object, as well as the V-O phrase will be stored. It can be observed from both Figure 5 and Table VI 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 "**".

TABLE VII. EXTRACTION RATE OF LP IN CLASS 1

	Extraction Rate of All Inputs	Usage of IP "看樱花"
Week 1	100%	100%
Week 2	100%	14.7%
Week 8	76.5%	26.5%

For examining the validity of the output of the system (Figure 5), we calculated the extraction rate that describes how many V-O phrases there are in the input sentences. The extraction rates of Class 1 are presented in Table VII. 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. Since the output is displayed in different colors, the difference is quite clear and the calculation is easily done. From Table VII 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 basic sentence structure, e.g., the V-O structure, all LP would be extracted and Figure 5 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. However, we also observed that in translations from week 2, 14 students used the two-character words in G2: "看花" (see flowers) or "赏花" (admire flowers). Nevertheless, in those 14 translations, all the V-O phrases were extracted because the students all used another verb "去" (go) before the two-character words. While the two-character words were regarded as noun by the parser, the noun and the verb "去" (go) together form another phrase which was determined as a V-O phrase and extracted. Other variations extracted from week 2 translation include G2 "(观)赏樱花" (admire cherry blossoms) and G3 "去看樱" (go to see cherry).

TABLE VIII. EXTRACTION RATE OF LP IN CLASS 2

	Extraction Rate of All Inputs	Usage of IP "看樱花"
Week 1	91.2%	0.03%
Week 2	97.1%	0.09%
Week 8	82.4%	26.5%

Table VIII shows the extraction rate of LP from Class 2 students' translation. In week 1, V-O phrases were not extracted in 3 translations. The two-character word in G2: "看花" (see flowers) exists in 2 of the 3 translations, and another null extraction is due to the mistake of using "看去花" (see go cheery blossoms). In translations from week 2, the only translation didn't have an extracted V-O phrase used the same two-character word in G2: "看花" (see flowers). In week 8, V-O phrases were not extracted in 6 students' translations. One of them is due to the false extraction of the simple sentence we explained before. All of other 5 students used the two-character words in G2: "看花" (see flowers) or "赏花" (admire flowers) without another verb nearby, which caused the null extraction.

From Figure 5, besides the extracted V-O phrases, the change detection results were also provided to teachers. The correct detection rate of all the 204 detections achieved 94.6%, which proved the high accuracy and the possibility to be utilized. To sum up, there are mainly two patterns in the eleven incorrect detections. The first one is that the students used the two-character words in G2: "看花" (see flowers) and "赏花" (admire flowers) in multiple translations. In this case, some of them were not extracted because of the lack of verb before the word and others will be extracted along with the verb. As a result, even the student used the same two-character word between two translations, in six students' translations of two exercises, one was extracted and the other was not, then the change detection turned out to be wrong. And in four students' translations, both of them were extracted along with the verb, however, the verbs were different, in week 2 they used "去" (go), but in week 8 they used "来" (come), so the system detected the change as a result. The second pattern is the null extraction of LP due to the mistakes across exercises. In one student's translation, the mistake caused one null extraction in week 1, and usage of two-character words caused null extraction in week 2 and week 8, so there is no extracted phrase from all the translations. Even he changed the phrases across exercises, the system couldn't detect.

VI. DISCUSSION

As our results show, the extraction rate of LP and the change detection over time were both found to have a high correct percentage. This means that the system can provide teachers with an almost immediate overview of a student's progress in the learning process. The color differentiation allows teachers to readily note that there was only one student in each class who used the IP in all three exercises in this instance. From the system output of Class 1 translations, it can be observed that eight students 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. However, the other 25 students failed to change back to the correct translation they submitted in week 1. From the output of Class 2, we found that even without the teacher's detailed explanation, two students changed their answers to the IP in week 2, and in week 8, nine students used the IP which was the same number of students who used the IP in Class 1 of week 8. As showed in Section II, these results clearly illustrate that a provided hint cannot improve a student's long-term performance. But with the system, the progress of the students can be promptly perceived and are more intuitive. This information can help teachers improve their instructional strategies, and facilitate individual students to comprehend whether the required grammatical element had been mastered or not. At the current stage, only the interface for teachers has been developed. Our pre-processing algorithm helped us extract virtually all of the translations so that it is a practicable method to adapt empirical observations of teachers to process students' translations that contain errors.

On the other hand, the incorrect change detection result caused by the parser's POS determination remains a problem. Incorrect outputs may not cause much trouble for teachers to distinguish, however, they may confuse the students. The interior algorithm of the parser is difficult to alter, so we may need to investigate the possibility and effects of adding exterior rules or using different segmentors. Segmentors do not seem to be able to separate the two-character words determined as V-O phrases by humans, so the POS tagger and dependency parser are unable to provide the desired information. In addition, it is noteworthy that we found that the system output changed significantly just by using different trained parser models concerning Chinese grammar. There are 5 different Chinese parser models trained on data from the Penn Chinese Treebank provided by the Stanford Parser. According to the official document, the PCFG parsers are smaller and faster, but the Factored parser is significantly better for Chinese. In the practical use, however, the output generated by using the xinhuaFactored grammar model was much worse than the result based on the xinhuaPCFG grammar model. Even when IP exists in translations, the extraction cannot achieve 100% by using the xinhuaFactored grammar model. Thus, in this specific context, choosing the appropriate Chinese grammar model should be considered of high importance.

VII. 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 phrases could help teachers confirm the effects of the provided hints or explanations; and students may benefit from the formative feedback to find out their weak points in the learning process. Thus, we proposed an approach for providing formative feedback and developed a system to test the approach.

We designed an interface for teachers to confirm students' acquisition of a specific grammatical element. Since the raw data contains a lot of complex sentences, which creates a barrier for our system to locate the desired part within the translation, we made an effort to improve the pre-processing method. The system achieved a high correct percentage in both the extraction rate of LP and change detection.

It is suggested that the system is effective in providing automated formative feedback to teachers. The feedback on V-O phrases would help teachers grasp the overall situation of learners and confirm the effects of the current instructional strategies. Although the V-O phrase is limited in the description in the methodology and the evaluation of the system, the system can certainly be used to track other V-O phrases as well. Because the system focuses on the extraction and comparison of V-O phrases by using the Stanford Parser, it is thus expected that it can be applied to other languages as long as similar structures can be identified by the parsers. Moreover, unlike Chinese, English sentences are separated by space, which makes it much easier to be correctly segmented.

There still remain some problems in the approach. In the system, the two-character words cannot be determined as phrases as we have explained. Furthermore, it is important to deal with the phrases without a "dobj" tag, which suggests that the extraction method still needs improvement. Currently we have only developed an interface for teachers, developing an interface for learners will be the next step. In addition, further practical use in classrooms needs to be investigated.

ACKNOWLEDGMENT

This work is supported by JSPS KAKENHI Grant Number JP17K01081.

REFERENCES

[1] S. Shao, K. Ohtsuki, H. Kiyomitsu, and M. Kang, "Tracking Verb Phrases for Formative Feedback in Foreign Language Writing," Proceedings of The Tenth International Conference on Mobile, Hybrid, and On-line Learning, pp. 58-61, 2018.

[2] H. Kimura, T. Kimura, and O. Shiki, "Theory and Practice in Reading and Writing: Nurturing Independent Learning," Tokyo, Japan: Taishukan, 2010.

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

[4] 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.

[5] 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.

[6] Y. Attali and F. van der Klei, "Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving", Computers & Education, (110), pp. 154-169, 2017.

[7] F. M. van der Kleij, T. J. Eggen, C. F. Timmers, and B. P. Veldkamp, "Effects of feedback in a computer-based assessment for learning", Computers & Education, (58), pp. 263-272, 2012.

[8] V. J. Shute, "Focus on Formative Feedback", Review of Educational Research, 78(1), pp. 153-189, 2008.

[9] 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, 27(3), pp. 385-392, 2017.

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

[11] 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.

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

[13] K. Kawamura, H. Kashiwagi, and M. Kang, "An Approach toward Automatic Error Detection in Learners' English Writing Based on the Source Language," Proceedings of The Tenth International Conference on Mobile, Hybrid, and On-line Learning, pp. 62-65, 2018.

[14] 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.

[15] S. V. 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.

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

[17] M. W. Easterday, D. R. Lewis, and E. M. Gerber, "Designing crowdcritique systems for formative feedback," International Journal of Artificial Intelligence in Education, 27(3), pp. 623-663, 2017.

[18] D. S. McNamara, S. A. Crossley, and R. Roscoe, "Natural language processing in an intelligent writing strategy tutoring system," Behavior research methods, 45(2), pp. 499-515, 2013.

[19] E. M. Golonka, A. R. Bowles, V. M. Frank, D. L. Richardson, and S. Freynik, "Technologies for foreign language learning: a review of technology types and their effectiveness," Computer assisted language learning, 27(1), pp. 70-105, 2014.

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