Effects Comparison between English and Chinese Speakers in Learning Japanese Role Words with Comic Scenes

Yoko NISHIHARA*, Kohei MATSUOKA** and Ryosuke YAMANISHI*

* College of Information Science and Engineering, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu-shi, Shiga 525-8577, Japan nishihara@fc.ritsumei.ac.jp, ryama@media.ritsumei.ac.jp
** Graduate School of Information Science and Engineering, Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu-shi, Shiga 525-8577, Japan

Abstract: The role word is one of the expressions for letting people imagine one’s character. It is difficult for people to learn the role words from conversation examples written in textbooks of foreign language because few examples are shown to learn the role words on the textbooks. Comic scenes have the examples of conversation in their line texts of character’s speech. We have proposed a support system for learning Japanese role words with comic scenes. The system classifies comic scenes according to a role word included in the scenes. System users can learn how to use role words by watching comic scenes and lines. We have verified the efficiency of the system for Chinese speakers whose language culture is similar with Japanese language culture; both of them use Kanji in writing. The good results might be caused by the similarity of language culture. In this paper, we report results of an experiment with English speakers. The results of English speakers were compared with those of Chinese speakers. We verified that the system can support English speakers to learn Japanese role words. We also discuss improvements of our system by referring comments from English speakers. They said that the system should enable the users to learn the role words in enjoying comic stories.

Keywords: language learning support, comic engineering, role word

1. INTRODUCTION

With the rapid progress of globalization, many people are learning foreign languages that are different from their native languages. Japanese language was learned by 2,979,820 learners as one of the examples of foreign language in 2006 [1]. The number in 2015 was 3,655,024 [2] that is 1.2 times bigger than that in 2006. Other languages like Chinese and English also have been learned by many foreign learners. Though the machine translation technique is generally used in daily lives, the needs of learning foreign languages are increasing in the world.

The learners are practicing their skills of foreign language. Their skill levels can be evaluated by examinations like TOEIC and TOEFL for English learners. The examinations evaluate their language levels from the four points: reading, writing, listening, and speaking. The learners practice their skills by exercises. For reading skills, they may read many sentences. As writing many sentences, they may improve their own writing skills. Their listening skills would be improved by listening to varied and many speeches. And, they may talk about something with others for their speaking skills. The three kinds of the skill (reading, writing, and listening) of those can be practiced by many learning materials. There are many documents such as news, books, and weblogs that can be used for practicing reading and writing skills. And, a lot of music, news programs, and TV dramas are very useful for practicing listening skill. A plenty of contents in the real world and the Web are available for learners of foreign language.

In contrast, there are few learning materials for practicing speaking skill; the practicing of the speaking skill has been less supported. Speaking skill can not be improved by just speaking sentences one-sided, and that should be improved by talking with other people. That should be the reason why the number of learning materials for the speaking skill is less. Moreover, speaking manner is different from writing manner. If words generally used in speaking are used in writing, the written sentence gives a strange impression to the readers. If words generally used in writing are used in speaking, the speech also might give too formal impression. People need to use proper words according to the situation and their characters: a role word is one of the typical examples of such words.

1.1 Role words for speaking in foreign language

The role word is one of the expressions for letting
people imagine one’s character [3]. The characters are one’s age, gender, occupation, class, era, looks, appearance, personality, and so on. When a young girl talks with others, she uses proper words for her character. In Japanese language, the young girl may use “ATASHI” (that equals to “I” in English) as a word concerning herself. Japanese people can imagine the character of the speaker from the phrase and the voice of “ATASHI.” In another case, a little boy uses usual words that may be different from his mother, and a little girl uses also different words in her talking. Though the number of role words is different among foreign languages, people tend to use proper words and expressions according to their characters. People unconsciously have obtained the knowledge of the relationships between role words and characters in their native languages from their daily talking. However, it is hard to obtain the knowledge in foreign languages because the knowledge has not been clearly described in language textbooks and learning materials.

We insist that learning materials for practicing the speaking skill must include the knowledge of the relationships between role words and characters. The learning materials should give learners the pairs of characters and sentences with role words. Through the survey on comic data, we found that lines of the comic scene are good examples of the pairs. In a comic, many characters appear and act their roles. A story of comic basically flows by conversations of the characters. The comic characters use their proper words for speech in the conversations. Especially, comic contains living words and phrases in the current real world. The existing study has confirmed that the lines and images of anime are useful learning materials for living words and phrases [4]. There are several characters in general comics. Each character uses proper role words to let readers know one’s feature. That means characters merely use the same role words. Readers of comic often use lines and character’s faces to identify each character. If the readers have trouble in the identification, they may not enjoy reading the comic because the story becomes unclear. Learners might obtain the knowledge of a role word from a few comic characters in a comic. In learning role words, the learners should check many examples of such scene and must find common features of characters using the corresponding role word. Then, it is necessary to use several comics for learning to find common features. Moreover, in order to learn a role word from several comic scenes, the scenes with lines have to be classified into several groups according to the role words. The learners should do these tasks by themselves; it costs a lot of time.

1.2 Learning support system of role words

We have already proposed a support system for learning the role words [5]. The outline is explained in this section though the details of the system are described in Section 2. The system extracts role words in lines of comics, and classifies the scenes into some groups according to the role words. As far as we know, there are few systems to support learning the role words for practicing speaking skills. The system provides the learners with a set of classified comic scenes corresponding to the role word; the learners can learn how to use the role words. This idea can be categorized into the application of comics in comic computing [6].

We did an experiment of the system with Chinese participants. They learned how to use the role words by using the system and took quizzes to answer proper role words. Their scores of quiz were compared with scores by other participants who learned how to use the role words by reading a textbook of role word. The scores by the participants using the system were higher. We verified the efficiency of the system from the results.

Chinese language culture is similar with Japanese language culture; both of them use Kanji in writing sentences. The good results might be caused by the cultural similarity. We need to conduct an experiment of the system with other participants of different language culture.

1.2 Objective of this paper

In this paper, we conducted an experiment of the system with English speakers and report experimental results. The results are compared with the results obtained from Chinese speakers in the previous experiment [5].

2. LEARNING SUPPORT SYSTEM OF ROLE WORDS

Figure 1 shows the outline of the system. The system consists of two parts: a data processing module and an interface.

In the data processing module, the system takes comic data as the input. The comic data is processed through optical character recognition machine to obtain text data of lines of comics. The system extracts role words from the obtained lines. The system classifies scenes into groups according to the extracted role word. The classified scenes are sent to the interface.

On the interface of the system, a list of the role words and classified comic scenes are visualized. The overview
of the interface is shown in Figure 2. The interface works on a Web browser. The interface has two parts: a list of role words on the left side, and comic scenes including a role word selected by the system user on the right side. As the system user selects a role word, the system shows comic scenes including a selected role word on the interface. The system user checks the images and reads the lines in the scenes to learn how to use the role word. The following sections will explain the procedures of the system from input to output.

2.1 List of Japanese role words
The system uses a list of role words defined in a previous study. Many academic papers and books related to the role words have been published [3,7]. In Japanese language, most of the expressions for role words are personal pronouns and sentence-end expressions.

2.2 Texts of line of comic scene
The system obtains texts of line from comic image data. As the experimental dataset of comics, this study uses a dataset of comics named Manga109 [8]. Manga109 dataset has been compiled by Aizawa and Yamasaki Laboratory of the University of Tokyo. In Manga109, 109 comics made by professional manga artists between 1970s and 2010s are collected.

The system obtains text lines of comic from Manga109 through an optical character recognition machine (OCR machine). The system uses Google Cloud Vision API (https://cloud.google.com/vision/?hl=ja) as the OCR machine which has a function to extract texts in an image. Many languages can be detected by the Vision API, and the language type is also detected automatically. As an image is sent to the API, a recognition result is back to the console as a json format. The proposed system extracts role words from the json files.

2.3 Japanese role word detection and scene classification
Some of the obtained lines include role words. The system extracts the role words from the lines by using a parser (morphological analyzer). In the system, MeCab (http://taku910.github.io/mecab/) and Neologd (https://github.com/neologd/mecab-ipadic-neologd) are used as the parser and the dictionary for parsing, respectively. The parser returns the parsed words and their parts of speech.

After parsing, the system detects role words by collating the list of role words. If a role word is detected in a line text, the system gives the role word as a label for classification to the scene including the line text. Some of the scenes includes several role words. Then, multiple labels are given to such scenes.

The labeled scenes are classified according to the labels of role word. If a scene has multiple labels of role word, the scene is classified into multiple classes.

3. EXPERIMENT
We conducted an experiment for English speakers as participants with our support system. The results were compared with the results of Chinese participants obtained in the previous experiment. We obtained comments from the English participants after the experiment. The comments were considered to improve the system.

The experimental procedures were the same as those in the previous experiment. We set evaluation items in Section 3.3 to evaluate the effect concretely. An experimenter (the second author) coordinated the experiment in the following procedures.
1) The experimenter gathered participants (called English participants).
2) The participants learned how to use the role words by using the system.
3) After learning, the participants took quizzes to
4) The experimenter compared quiz scores between English participants and Chinese participants. The participants were English people who were learning Japanese language. They lived in Vancouver, Canada. The number of the participants was 12. Most of them have obtained scores of JLPT like TOEIC and TOEFL for English. Five of them have obtained N1 that was highest, three of them have obtained N2, and two of them have obtained N3.

3.1 Used role words in the experiment
The used role words were the same as those in the previous experiment. There are 53 labels of role word to learn in the system. The experimenter selected some role words for the experiment because it might be impossible to learn all of the role words in the experiment period. The experimenter counted up the number of scenes for each role word, then sorted the role words according to the number of scenes. Top, middle, and bottom 10 role words were respectively selected for the experiment.

3.2 Quizzes for evaluating participant’s level
We evaluated levels of participant with quiz scores. The quiz sheet was almost the same as that in the previous experiment. A role word was removed from the sheet. The quiz sheet had two kinds of quiz. One type of quiz (Quiz 1) was to answer appropriate/inappropriate to each role word in a line text spoken by a comic character. An example of the quiz is shown in Figure 3. The another type (Quiz 2) was to choose the most appropriate role word for the shown scene from a list of three role words. An example of the quiz is shown in Figure 4. For the Quiz 1, we selected two or three role words from each top, middle and bottom category, respectively. Eight role words in all were selected. As same as this, we also selected each two or three role words respectively for Quiz 2. There was no common role words between the two sets of role word. Table 1 shows the role words used for the quizzes.

Table 1: Role words used in the quizzes

<table>
<thead>
<tr>
<th>Quiz 1</th>
<th>Quiz 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 10</td>
<td>さ、ぜ、な、や、な、じゃ</td>
</tr>
<tr>
<td>Middle</td>
<td>あら、かしら、ぼく、あたし、あんた、わし</td>
</tr>
<tr>
<td>Bottom</td>
<td>たまえ、拙者、申す、えらい、貴様、わたくし</td>
</tr>
</tbody>
</table>

3.3 Evaluated items
We evaluated the following six items. The items were also the same as those in the previous experiment.
- Average score
- Average score for Quiz 1
- Average score for Quiz 2
- Average scores for top 10 role words
- Average score for middle 10 role words
- Average score for bottom 10 role words

3.4 Experimental results
Table 2 shows the experimental results. The average total score was 13.9. The average score for Quiz 1 was 7.1. The average score for Quiz 2 was 6.8. The average score for top 10 was 3.5. The average score for middle was 5.2. The average score for bottom was 5.3.

They learned the role words about 40 minutes for the three days, respectively. The average of the 1st day was 42.9 minutes, that of the 2nd day was 43.7 minutes, and that of the 3rd day was 40.3 minutes.
4. DISCUSSION

This section discusses the results by comparing the results obtained from Chinese participants in the previous experiment. We will discuss the comments from the English participants.

Table 2 shows that the average correct rate was 87%. The rate of Chinese participants was 80%. As both of the rates was almost the same, we verified that our system can support English people in learning how to use Japanese role words.

We considered three role words of low correct rate. Table 3 shows the three role words on the left column. The right column shows three role words of low correct rate by Chinese participants. English participants marked low correct rates for answering “あら,” “ざ,” and “えらい.” The correct rates were 53%, 67%, and 67%, respectively. Chinese participants marked low correct rates for answering “き,” “び,” and “申し.” The correct rates were 38%, 63%, and 63%, respectively.

Both of the participants marked lower correct rates in answering a quiz of “ざ.” The word is used to express for multiple roles (countryside people, Okinawa people, Man, and Boy). Therefore, it may be difficult to find the relationships between the word and specific roles. We need to improve the system to learn such role words more efficiently.

Table 3 shows that five of six quizzes were to answer appropriate/inappropriate for a buried role word (Quiz 1). The option quiz shows the list of role words. The participants can answer proper role words by referring the other role words. The Quiz 1 does not show any candidates of role word. The low correct rate for Quiz 1 might be caused by the reason.

Parts of speech in Table 3 were different between English participants and Chinese participants. English participants marked lower correct rates in answering quizzes of adverb role words. “あら” and “えらい” are adverbs mainly used in a sentence though “ざ” is a particle mainly used in the end of sentence. In contrast, Chinese participants marked lower correct rates in answering role words mainly used in the end of sentence. “ざ” and “び” are particles, and “申し” is a verb. All of them are used mainly in the end of sentence. We need to consider the reason as the future work.

We obtained comments from English participants after the experiment. Most of them commented the improvements of visualization of the system interface. They advised us that the system should enable the users to learn the role words in enjoying reading comic stories. The system uses comic scenes. However, the system does not use comic stories. We would like to improve the point. The improvements will also give benefits in enjoying comic nuances.

5. CONCLUSION

This paper reports experimental results of English speakers with our support system for learning Japanese role words. The system extracts line texts of comic scene and detects role words included in the lines. Then, the system classifies comic scenes into classes according to the included role words. Each of the classes is given a role word as a label of class. We had already done an evaluation experiment of the system with Chinese speakers. The experimental results were good that might be caused by the similarity of language culture with Japanese. In this paper, we did another experiment of the system with English speakers as the participants and compared the results with those in the previous experiment.

In the experiment, we asked English participants to learn how to use role words with the system. They took quizzes to answer proper role words after learning. We compared the quiz scores of English participants with the scores of Chinese participants. They were almost the same. We verified that the system can support English people to learn how to use role words.

English participants marked lower correct rates in answering quizzes of adverb role words. In contrast, Chinese participants marked lower correct rates in answering particles and a verb mainly used in the end of sentence. We need to consider the reason as the future work.

We obtained comments from English participants after the experiment. They said that the system should enable...
the users to learn the role words in enjoying comic stories. We would like to improve the point and the usability of system interface. We also try to evaluate the difficulty of line texts of comic scene by using our evaluation method [9].

ACKNOWLEDGMENTS
The experiments in this paper were conducted with the great support of Yokoso Japan Association.

REFERENCES