Learning Effect of Implicit Learning in Joining-in-type Robot-assisted Language Learning System

AlBara Khalifa(✉), Tsuneo Kato, Seiichi Yamamoto
Doshisha University, Kyoto, Japan
`albara.khalifa@gmail.com`

Abstract—The introduction of robots into language learning systems has been highly useful, especially in motivating learners to engage in the learning process and in letting human learners converse in more realistic conversational situations. This paper describes a novel robot-assisted language learning system that induces the human learner into a triad conversation with two robots through which he or she improves practical communication skills in various conversational situations. The system applies implicit learning as the main learning style for conveying linguistic knowledge, in an indirect way, through conversations on several topics. A series of experiments was conducted using 80 recruited participants to evaluate the effect of implicit learning and the retention effect in a joining-in-type robot-assisted language learning system. The experimental results show positive effects of implicit learning and repetitive learning in general. Based on these experimental results, we propose an improved method, integrating implicit learning and tutoring with corrective feedback in an adaptive way, to increase performance in practical communication skills even for a wide variety of proficiency of L2 learners.

Keywords—Computer Assisted Language Learning (CALL), Robot Assisted Language Learning (RALL), Implicit Learning, Corrective Feedback

1 Introduction

Today's globalization has made communication in a second language (L2) an everyday matter for a large number of people. Communication in L2 takes place everywhere: abroad, online, and even in one's own home country. Learning an L2 has become more important than ever. As a convenient and economical self-learning method, computer-assisted language learning (CALL) systems are gaining attention. CALL systems originated with Computer Assisted Pronunciation Training (CAPT) [7][25]. CAPT systems are good for training in pronunciation but not for training in how to express one's notions with correct lexical choice and grammatical use.

In accordance with the advances in automatic speech recognition (ASR), the functions of CALL systems have evolved to support the learning of various L2 communication skills, such as grammatical correctness, lexical choice, suitability of expression, and proper pronunciation through dialogues with the system. Dialogue-based CALL

systems ask learners to construct expressions themselves. Various dialogue-based CALL systems have been proposed, including SCILL [30], SPELL [24], DEAL [6], POMY [18], and DISCO [8]. However, the speech of L2 learners poses intrinsic difficulty for ASR due to their various pronunciations, incorrect lexical choices, and grammatical errors. Several techniques for enhancing pedagogical effectiveness have been developed, such as hint generation [6][18], grammatical error detection, and corrective feedback [24][8]. A dialogue-based CALL system can not only train L2 learners to construct expressions by themselves but also raise the learners' interest, motivation, and engagement through its interactive nature. As a notable example, POMY raised learners' motivation and engagement with game contents in a 3D immersive environment.

To make such interactive CALL systems more attractive and realistic, and to help learners prepare for actual face-to-face communication, robot-assisted language learning (RALL) systems have been proposed [11][19]. A RALL system has a physical presence that learners must be aware of while responding to it. The physical presence of a robot was reported to be effective in increasing cognitive learning gains [20]. The RALL system introduces nonverbal modalities such as gestures, nodding, and face tracking to the interaction. These modalities raise the level of experience closer to real face-to-face communication, giving the learner a more realistic sensation. An increasing number of studies, most of which were directed toward children, reported that the introduction of robots enhanced the learners' interest, motivation, and engagement. One study reported the effect of retaining vocabulary with and without a robot assisting the teacher in a second-language classroom [1]. They reported that children using the assistive robot retained more vocabulary than children not using it after two weeks. On the other hand, another study reported that the social behavior of a tutoring robot did not affect the efficiency of retention in L2 learning for children [12].

In terms of the learning effect, one-on-one tutoring by a skilled instructor is believed to be the best way to learn an L2 [22][4]. We can apply the capabilities of CALL or RALL systems to model the functionality of a skilled instructor in one-on-one conversation, which is considered the ideal target of automatic tutoring systems. However, recognizing L2 speech is a challenge even for state-of-the-art ASR engines because it contains various levels of pronunciation quality in addition to lexical, syntactic, and semantic errors. Furthermore, giving appropriate corrective feedback to each learner is still more difficult, considering the wide variety of proficiency of L2 learners and the various reasons that cause the learner to produce erroneous utterances. An integration of tutoring and the implicit learning that naturally happens in a classroom is expected to overcome these difficulties of automatic tutoring systems. In such classrooms, the student is given chances to repeat after the teacher, answer questions, receive correction when necessary, and also learn by viewing the interaction between other students and the teacher; however, the teacher may not be able to give each student enough opportunity to communicate with him or her. Occasionally, students are asked to collaborate on a more complex task and present their thoughts or ideas on the task. Such a combination of tutoring by the teacher and implicit learning from peer learners is considered effective in learning various aspects of communication in L2.

In terms of feasibility, effective implicit learning helps not only the learners but also the ASR of L2 speech. To deal with the challenges mentioned above, implicit learning helps regulate the human learner's utterances. The interaction between robots is expected to encourage human learners to use expressions similar to those used by the robots.

Additionally, a method of regulating the learner's utterance by showing answers using appropriate grammatical patterns as a reference is expected to provide a means of easily expanding the variety of topics compared with conventional dialogue-based CALL systems. Such a method can provide learners with opportunities to learn how to use various expressions in real conversational situations and polish up their practical communication skills they have learned but have had few chances of using to construct their own utterances.

The combination of tutoring by a teacher and implicit learning from peer learners can be simulated with robots in a straightforward manner. That is, a learner receives tutoring when a robot asks a question or corrects the learner directly, and the learner learns implicitly when the robot similarly interacts with another robot. By giving some examples, the interaction between robots can provide hints on how to make a response or simply show a model answer. Robots can even entertain or relax the learner to facilitate spontaneous speaking. Furthermore, it is possible to measure the effect of tutoring and implicit learning by automatically evaluating learner responses through interaction with the robots.

On the basis of this concept, a novel RALL system was proposed that uses two humanoid robots to help L2 learners learn English and conducted a series of experiments to evaluate the effectiveness of the proposed system [13][14][15][16]. One robot plays the role of a teacher, and the other plays the role of an advanced peer learner. Because the robots invite the human learner to join in the dialogue, this system is called a "joining-in-type humanoid-robot-assisted language learning system."

Although it is not easy to achieve truly flexible learning, a typical form of implicit learning is to borrow a useful expression from what a peer learner uses. In order to explore the effect of implicit learning facilitated by the RALL system and to obtain data for optimizing the system, we developed a prototype joining-in-type RALL (JI-RALL) system with two robots and implemented scenarios controlled by a Wizard-of-Oz method. The scenarios have seamless flows that immerse the learners into them, and they contain various questions prompting the human learner to respond in the specific syntactic forms under focus. The interaction is designed to be extended further to promote more effective learning, depending on how the human learner responds to the questions.

We conducted experiments with 80 participants under various experimental conditions and created a learner corpus of multimodal conversational data collected with the system. The multimodal learner corpus consists of audio and video recordings from two fixed cameras.

The rest of this paper is structured as follows. Some of the related studies on RALL are described in Section 2. The system's concept and configuration are given in Section 3. Section 4 overviews a series of experiments to explore the learning effect with implicit learning using a prototype JI-RALL system, the methodology of creating a

conversational scenario, and other issues. We report an experiment on the effect of pre-presentation of a reference in Section 5 and an experiment on the retention effect in Section 6. A discussion on our findings is presented in Section 7. Finally, our conclusions and future works are given in Section 8.

2 Related Studies

The first trial of introducing a robot into language learning was Robovie [11]. Robovie was introduced in a classroom, and it communicated with Japanese elementary school pupils for three weeks. Although the interaction gradually decreased due to the limited patterns of speech synthesis and the limited vocabulary of ASR, some students continued to interact with Robovie.

Lee et al. [19] designed a course of English lessons using a robot for elementary school students and measured its cognitive effect on oral skills and affective effects. RALL provided a significant improvement in speaking skill, though not in listening skill. Moreover, RALL raised the students' satisfaction, interest, confidence, and motivation.

The effect of pairing a human teacher with an assistive robot was evaluated in an L2 classroom for children [1]. Children with both a teacher and a robot learned and retained more vocabulary than did children with only the teacher.

Recently, use of social robots for teaching a second language to preschool children has been under development in a project of universities in Europe [3]. This project aims not only at teaching English to native European children, as well as teaching Dutch and German to immigrants, but also at responding to children's actions and engaging with them adaptively while tutoring.

One of their papers reported an experiment on how the social behaviors of the tutor robot affected child second-language learning [12]. Although children showed significant improvement between pre- and post- tests under conditions of both high verbal availability of the robot and low verbal availability, the difference between the two conditions did not reflect any difference in the improvement gained.

As for the effect of repetition in education, another paper in the project measured the effect of repetition in collaborative tasks between children and a peer robot [2]. The results showed positive effects on child performance generally, and they indicated that this was driven by the individuals because the performance improved even in the case of sparse feedback in the peer-peer interactions.

In terms of implicit learning, borrowing another's expression while in a dialogue is associated with interactive alignment [27] or entrainment. Interactive alignment is an unconscious process in which interlocutors tend to re-use lexical, syntactic and other linguistic structures after their introduction. Such alignment was observed in various areas such as lexical choice [5], syntax [29][23], and style [26].

Alignment or entrainment occurs not only in human-human conversation but also in dialogues between a system and a user. Fandrianto and Eskenazi proposed a dialogue strategy to help regulate users' pronunciation as a way to improve the system's ASR performance [9].

3 Joining-in-type robot-assisted language learning system

3.1 System concept and its pros and cons

One-on-one tutoring is considered desirable because expert tutors are assumed to provide better instruction to the student. The standard hypothesis is that the effectiveness of tutoring is due to the dialogue between the tutor and the student, and that it is the interactive nature of that dialogue that accelerates learning. That is, if students simply read or listen to an explanation instead of participating in a dialogue, learning is less effective. However, VanLehn et al. [32] showed that this is not always true: Reading or listening can sometimes be as effective as one-on-one interactive human tutoring.

The joining-in-type robot-assisted language learning (JI-RALL) system, in which one robot takes the role of a language teacher (i.e., a tutor) and the other robot takes the role of a language-learning student (i.e., a peer), has been proposed to integrate tutoring in dialogue and implicit learning by listening to the dialogue between the two robots [13]. This system allows explicit instruction methods to be used by the tutor robot, while implicit learning methods based on listening to the dialogue can be exploited through the role of the peer learner robot.

In this setup, JI-RALL systems can flexibly switch between the two approaches. The first approach is tutor-initiated interactions between the tutor robot and the peer learner robot while the human learner is passive and merely observes and listens to the robots. The second approach is initiated by smoothly switching the dialogue to the human learner and giving him/her corrective feedback if necessary. While the two robots interact with each other, the human learner is expected to obtain knowledge from the conversation between them in implicit peer learning. Because this system invites the human learner to join in the triad conversations, we called it a joining-in-type humanoid-robot-assisted language learning system [13].

JI-RALL systems may have advantages in situations where the human learner finds it stressful to participate in face-to-face interaction. Learners of a second language (L2) are often able to follow a spoken L2 dialogue when listening passively but find it difficult to step in and speak actively. Here, cultural differences need to be taken into account; for example, many Japanese learners of English find it stressful when they are suddenly required to say something in English without having time for thinking and preparation [33].

JI-RALL systems have the above-mentioned excellent functions; however, no method on how to integrate tutoring and implicit learning has been clarified, especially how to detect erroneous utterances of the human learner and give corrective feedback to him/her.

We have developed a prototype joining-in-type RALL system using two NAO robots to explore how to integrate tutoring and implicit learning as a way to provide learners with opportunities to construct utterances with appropriate grammatical patterns in face-to-face conversations. In order to build a successful interaction between the robot and the human learner, the robot should communicate verbally and non-verbally. For non-verbal communication, the robot's embodiment feature was utilized

by having an expressive set of gestures that were automatically chosen from a library of gestures according to keywords detected in the sentence to be uttered. For this purpose, we used the robot's "Al AnimatedSpeech" module, which is a built-in component of the NAO robot's operating system. Figure 1 shows the experimental prototype system in which the two robots are in a standing position on a table in a triangular configuration with the human learner.
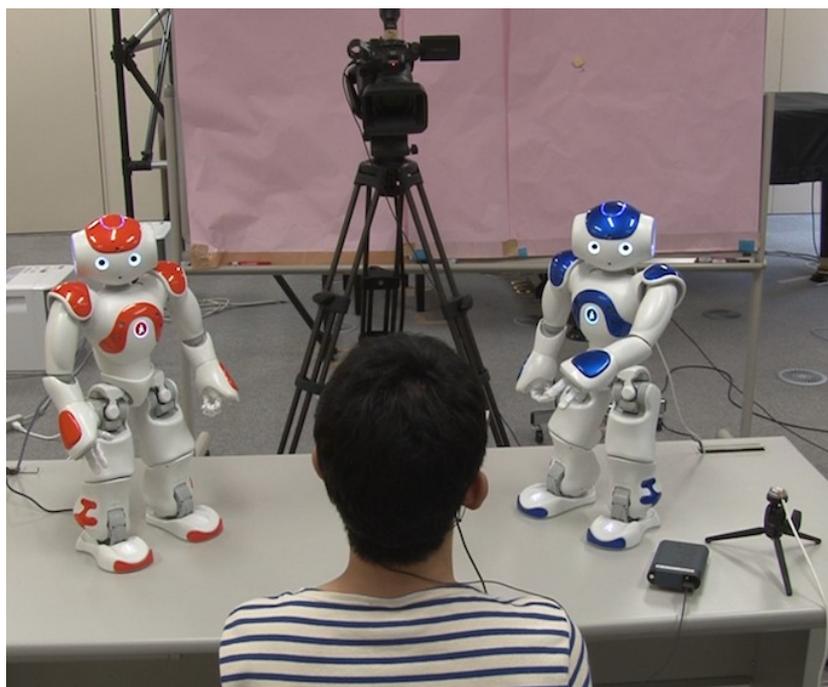

Fig. 1. Back-view snapshot of experiment, showing the two robots looking at a human learner wearing an eye tracker as one of them asks him a question.

4   Experiments

We conducted a series of experiments using the prototype JI-RALL system to quantify the effect of implicit peer learning by focusing on the learning of appropriate grammatical patterns used in dialogue. This was done as an effort to explore how to integrate tutoring and implicit learning.

4.1   Experimental setup and procedure

The triad conversations were conducted with a Wizard-of-Oz method to diminish the effect of misrecognition by the automatic speech recognizer (ASR) and to cover a variety of topics. An experimenter controlled various behaviors of the robots during

the experiment through an Ethernet connection and thus had some flexibility in controlling the flow of the conversation. Specifically, the experimenter could make the tutor robot repeat the question if the human learner could not answer, say a sample answer and repeat the question, or just pass over the current question and continue the conversation in cases where the learner could not answer at all.

To record the experiments, we used two cameras: one in front of the human learner to record his/her activities and facial expressions and the other behind the human learner to focus on the robots' actions toward him/her. Audio was recorded using a headset microphone worn by the human learner as well as the microphones equipped on the video cameras. Information on the learner's point of gaze was captured with a glasses-type eye-tracking system for future analysis of eye-gaze information.

The collected data of the eye-gaze activities were annotated manually by noting the start and end times of every gaze action by the human learner toward each robot. The data also included manual transcription of the utterances of the human learner and of the robots along with their start and end times. In this work, we give no analysis or discussion of eye-gaze activities, although we obtained interesting results on non-verbal information such as the human participants gazing more at robots than at human participants when the listeners joined triad conversations [10].

The participant was instructed to get involved in the conversation with the two robots by answering their questions, to wait for a while, if a question was not clear, for the robot to repeat it or give a hint of the answer, and to speak naturally and clearly. Then, after carrying out the calibration process of the eye-tracking system with the participant, the experimenter started the experiment.

Finally, after the conversation came to an end, the participants were asked to fill out a questionnaire. The questionnaire was used to measure the attitude of the participants toward the experiment. Some questions were about their previous experience of English and about dealing with robots. Other questions were about their impression toward the robots. They were also asked to evaluate their interaction in the conversation and to give their overall impression of the experiment. The scale used in the questionnaire had seven levels. "" lists all questions used in the questionnaire.

4.2 Participants

We recruited a total of 80 Japanese university students between the ages of 18 and 24 as participants in a series of experiments. The participants had acquired Japanese as their L1 and had learned English as their L2. Proficiencies of the recruited students were set to range widely from lower to higher proficiency, that is, we recruited participants of various proficiency levels to evaluate the effectiveness of the JI-RALL system for a realistic range of users. More than half of the participants took the Test of English for International Communication (TOEIC), and their scores ranged from 320 to 980, with an overall average of about 564 (990 being the highest attainable score). We conducted experiments with 57 participants to evaluate the effect of pre-presentation of a reference before answering as well as an experiment with 23 participants to evaluate the retention effect based on the experimental results of the effect of this pre-presentation of the reference before their answering.

4.3 General style of conversational scenarios

The conversational scenarios were designed to mimic daily conversations that begin in greetings and move on to chatting on topics like music, movies, sports, travel, new products and food. These scenarios were designed to draw the human learner into the conversations in a question-and-answer style. The tutor robot R1 asked the same or similar questions to learners and induced the learner to use the same grammatical patterns as those used by the peer learner robot R2. The conversation between two robots consisted of a variety of sentence patterns, such as yes/no questions and 5W1H questions. Some of the questions were expected to be answered in pre-selected grammatical patterns (past or perfect tense, causative verb, passive voice, answer to negative question, etc.). "" shows a shows a sample of the dialogues extracted from a series of experiments.

In implicit peer learning using two robots, the tutor robot R1 pursues dialogue with the peer learner robot R2, asks a question to either the peer learner robot R2 or the human learner first, and then asks a similar question to the other. When the tutor robot R1 asks the question to the peer learner robot R2 first (pre-presentation of a reference) and then asks the same or similar question to the human learner second, we can expect implicit peer learning using the answers of the peer learner robot R2. The human learner can listen to how the peer learner robot R2 responds to the question and refer to or mimic the utterances by the peer learner robot R2 in his/her turn.

5 Experiment on Effect of Pre-Presentation of Reference

5.1 Experimental procedure

The purpose of this experiment is to evaluate to what extent learners refer to or mimic utterances by the peer learner R2 at first glance. Participants are divided into two groups (with and without pre-presentation of reference groups), and each participant is induced to join in the triad conversation along with the two robots. During the conversation, the tutor robot R1 asks the participant five questions that should be answered in a certain grammatical pattern. Three of the questions are spoken to R2 and then to the participants in the group with pre-presentation of a reference, and the same three questions are only spoken to the participants in the group without pre-presentation of a reference. The two remaining questions are spoken to the participant first and then to the peer learner robot R2 second in both groups to evaluate the effect of implicit learning by pre-presentation of a reference under equal conditions. The procedure is summarized in Table 1.

Every participant performed two consecutive conversations of 10 minutes long, taking a rest of five minutes between them. The first and second halves of the consecutive conversations are conducted based on two kinds of similar scenarios in which the tutor robot R1 asks questions that should be answered in the same grammatical pattern but with different vocabulary according to each question.

We designed two kinds of scenarios for the experiment. For the first experiment, we selected a scenario (scenario A) in which the tutor robot R1 asks questions that should be answered in two different grammatical patterns: the past tense and the present perfect tense, which Japanese learners often mistake in speaking English though they study these patterns at the comparatively early stage of grammar learning. As the scenario of the second experiment (scenario B), we selected three different grammatical patterns: causative verb, passive voice, and negative question, which are expected to be difficult for Japanese learners of English. The test procedure is the same as that for the experiment using scenario A. The numbers of participants are 20 for scenario A and 37 for scenario B.

Table 1. Question order with and without pre-presentation of reference

Kind of question (number)	Order of questions from R1 to other chatting members	
	Without pre-presentation of reference	*With pre-presentation of reference*
Training (3)	Questions to learner only	Questions to R2 first and learner second
Evaluation (2)	Questions to learner first and R2 second	Questions to learner first and R2 second

5.2 Experimental results

Figure 2 shows the experimental results as the ratio of using the same grammatical patterns used by the peer learner robot R2 to all answers (abbreviated as SGP in the following). The ratio shows to what extent the participants refer to or mimic the utterances by the peer learner R2. The experimental results are summarized as follows:

- The use ratio of SGP increased by around two times from the first half to the second half in both experiments using scenarios A and B.
- The use ratio of SGP is larger for the participants with pre-presentation of a reference than for those without pre-presentation of a reference.
- The use ratio of SGP in the second half of the experiment of scenario A (use of past and perfect tenses) is higher by about two times than that of scenario B (use of causative verb, passive voice, and negative question).

Result (1) shows that the implicit learning generally promotes human learners to construct utterances with appropriate grammatical patterns. Result (2) shows that the more pre-presentation of reference is given, the more effective can human learners construct utterances with adequate grammatical patterns, and the more repetitive learning is indispensable to making implicit learning effective. Result (3) suggests that the more difficult the used grammatical pattern, the more insufficient the pre-presentation of a reference is for human learners to learn how to construct the utterance appropriately.

Analyses of collected utterances of the participants show that many utterances are answers of a single word or single phrase, or irrelevant answers. The former answers suggest that the participant understands the conversational flow but does not express his or her answer in appropriate grammatical patterns. The latter answers may have

been caused by the participants of low proficiency not understanding the conversational flow. The number of irrelevant answers is larger in the experiment using scenario B than in that using scenario A.
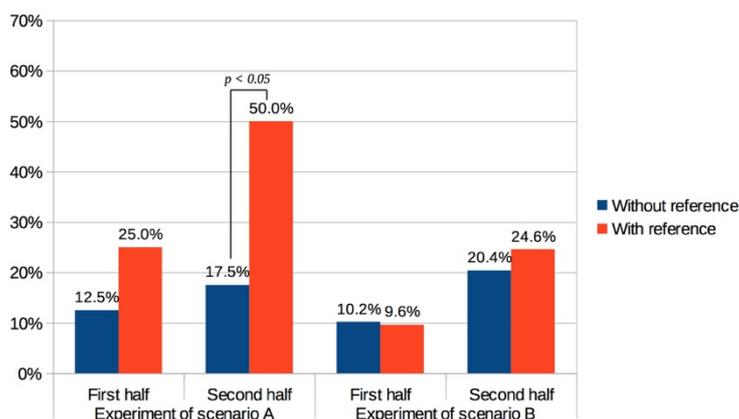

Fig. 2. Average use ratio of SGP (same grammatical patterns) in first and second halves of two consecutive experiments of scenario A (use of past and perfect tenses) and scenario B (use of causative verb, passive voice, and negative question). The difference between the two groups in the second half of Scenario A is significant, using t-test.

6 Experiment on Retention Effect

6.1 Experimental procedure

The previous experimental results show the effect of pre-presentation of a reference in implicit learning, where the more repetition is the more effective the implicit learning is. We created four different scenarios that were used over four consecutive weeks and in a retention test to evaluate the repetitive effect and the retention of implicit learning. As typical grammatical patterns, we selected causative verbs and inanimate subjects to be used in the questions and answers of R2, which are expected to be difficult for Japanese learners of English because such an expressional style is far different from that of their native tongue. In every week, a session of conversations was held, and each participant was asked 10 questions in every turn in every week. Different from the setup for the experiment on pre-presentation of a reference, learners were divided into two groups (with pre- and post-presentation of reference groups), and each participant of both groups had the same number of chances to hear the reference as shown in Table 2. For all 10 questions that were asked to the participant, the tutor robot R1 asked a question to the participant first and then asked the same or similar question to the peer learner robot R2 for the group with post-presentation of a reference. In the case of the pre-presentation of reference group, six questions out of ten were asked first to the peer learner robot R2 and then to the par-

ticipant as shown in Table 2. The scenarios of the first halves of week 1 and week 3 were used in the retention test.

We recruited 23 participants from the same population of Japanese undergraduate or graduate students. Among them, 16 could participate in all sessions.

Table 2. Question order for groups with pre- and post-presentation of reference

Kind of question (number)	Order of questions from R1 to other chatting members	
	With post-presentation of reference	*With pre-presentation of reference*
Training (6)	Questions to learner first and R2 second	Questions to R2 first and learner second
Evaluation (4)	Questions to learner first and R2 second	Questions to learner first and R2 second

6.2 Experimental results

Figure 3 shows the average use scores of SGP in week 1 in the training sessions of three consecutive weeks (intermediate two weeks and last week), and in the retention test. The experimental results are summarized as follows:

- The use ratio of SGP increased from the experiment on the first week to that on the last week.
- The use ratio of SGP is larger for participants with pre-presentation of a reference than for those with post-presentation of a reference.
- The difference in the use ratio of SGP between participants of both groups is lower for the retention test than for the training sessions.
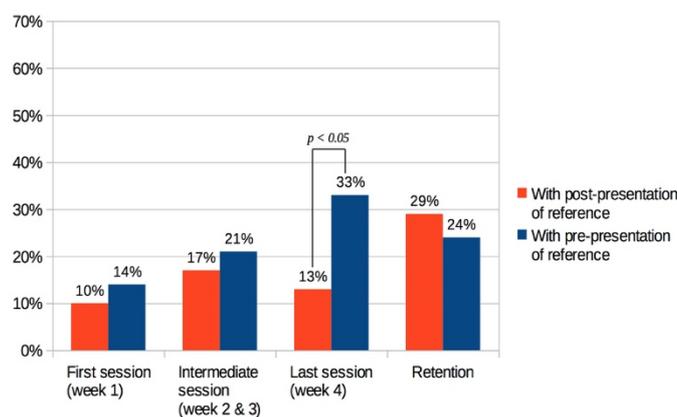- The use ratio of SGP was less than 35%, even for the participants with pre-presentation of a reference.

Fig. 3. Average use ratio of SGP in first session, consecutive training session and retention for experiments (use of causative verb and inanimate subject). The difference between the two groups in the last session is significant, using t-test.

Implicit learning generally promotes human learners to construct utterances with appropriate grammatical patterns; however, the improvement obtained through training of four consecutive weeks does not seem so high. To explore the reason for this, we investigated the improvement of each participant. Figure 4 compares improvement of the use ratio of SGP for participants by dividing them according to their use ratio of SGP into upper half and lower half. The achieved ratio depends on each participant, and its variation is very large as shown in Figure 4. The achieved ratio is largely divided into two classes, and the final achievement is 47% for one class and only 3% for the other class. These results suggest that the scenario may be too difficult for participants of the lower achieved ratio class. The implicit learning is effective for human learners, but mainly for participants of relatively higher proficiency. The teaching material and the conversational scenarios in JI-RALL systems should be designed to fit the proficiency of the human learner, and a method to measure the proficiency of each learner should be explored.
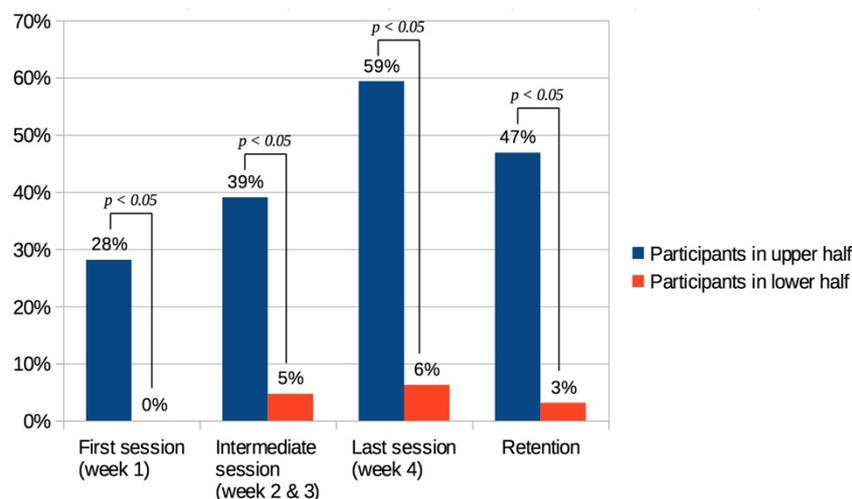

Fig. 4. Average ratio of SGP for participants whose use ratio of SGP is in upper half class and lower half class. The difference between the two classes in all cases is significant, using t-test.

7 Discussion

The experimental results in Section 5 show that the implicit learning provided with the JI-RALL system is promising for giving users the chance to obtain the ability of using appropriate grammatical patterns in real conversational situations. The analysis results of the experiment under scenarios A and B show that the human learner used appropriate grammatical patterns at a high ratio when he/she responded just after the answers of the peer learner robot R2, compared with the responses without the pre-presentation of answers by R2. These results suggest an effect of implicit learning in

the JI-RALL system. The experimental results in Section 6 show that repetitive training of implicit learning increases the ratio of using appropriate grammatical patterns. The ratio is higher for participants with pre-presentation of a reference than for those with post-presentation of a reference in the training session; on the other hand, the difference between the ratios for participants of the two groups decreased in the retention test. These results suggest that post-presentation of a reference performs as implicit learning as well as pre-presentation of a reference, especially for retention, and that post-presentation of the reference may function as corrective feedback to human learners.

The retention test in Figure 3 shows that the results of the participants with pre-presentation of a reference are lower than those of the group with post-presentation of a reference; this is unexpected, especially after gaining a high increase during the training sessions. The reason could be due to the effect of short memory with the references that had been presented by R2 right before the answer. This means that the high increase in their results in the training session could be a combined effect of short memory and implicit learning. The implicit learning effect from the post-presented reference by R2 appeared in the result of the retention test, which could be utilized as a corrective feedback effect.

The analyses of the series of experiments suggest the following structure as an improved prototype of the JI-RALL system. The tutor robot R1 asks questions to the peer learner robot R2 first and then to the human learner in the early stage of implicit learning. The system then moves to the intermediate stage, in which the tutor robot R1 asks questions to the human learner. If the system detects that the human learner could not use an appropriate grammatical pattern in this stage, the tutor robot R1 asks the same question to the peer learner robot R2. The answer of R2 in this case functions as corrective feedback to the human learner.

It is necessary to develop a method of classifying the responses of a human learner into utterances that use appropriate grammatical patterns and those that do not. As mentioned above, speech recognition of L2 speech is still difficult even for state-of-the-art technology; however, classification of utterances is easier in comparison with speech recognition even if the input speech is L2 speech. The classification of utterances has been explored as detection of out-of-domain utterances in the research community of speech recognition [36], and two kinds of language models, a general language model and a model specifically designed for utterances in a domain, are generally used to classify utterances into two categories based on a comparison of acoustic likelihood between outputs from speech recognizers using each language model [17]. In the case of detecting whether speech is constructed based on the appropriate grammatical pattern, a phonetic typewriter model is used as a general language model, and a language model may be designed to consist of a finite state automaton model concatenating words representing the appropriate grammatical pattern sandwiched between two garbage models.

As for the adaptability to human learners of lower proficiency, the system should have a function to count how many utterances the human learner consecutively produces and then change to questions using easier words when the count overcomes a

pre-determined threshold. The behavior of eye gazing by human learners may be used as a measure of evaluating whether they understand the conversational flow [10].

8   Conclusion

This paper described a series of experiments to evaluate the leaning effect with implicit learning in JI-RALL systems. Analyses of the experimental results show that implicit learning in the JI-RALL system is promising for human learners to construct utterances using the appropriate grammatical patterns. However, applying only implicit learning is not sufficient for human learners of a wide variety of proficiencies to obtain the ability to construct utterances of appropriate grammatical patterns, especially if they are considered difficult, such as using an inanimate subject.

To overcome these problems, we proposed an improved version of the JI-RALL system that uses tutoring with corrective feedback in addition to implicit learning. In the tutoring with corrective feedback, the tutor robot R1 asks the same questions to the peer learner robot R2 as those to the human learner. We plan to improve our prototype of JI-RALL so that it has such functions and to conduct an experiment to evaluate the effect of the improved functions on human learners of a wide variety of proficiencies.

9   Acknowledgement

10   References

[1] Alemi, M., Meghdari, A., & Ghazisaedy, M. (2014). Employing humanoid robots for teaching English language in Iranian junior high-schools. International Journal of Humanoid Robotics, 11(3), 4-16. https://doi.org/10.1142/S0219843614500224

[2] Baxter, P., Kennedy, J., Ashurst, E., & Belpaeme, T. (2016). The Effect of Repeating Tasks on Performance Levels in Mediated Child-Robot Interactions. Proc. of Robots 4 Education Workshop 2016.

[3] Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., Kopp, S., Bergmann, K., Leseman, P., Küntay, A. C., Göksun, T., Pandey, A. K., Gelin, R., Koudelkova, P., & Deblieck, T. (2015). L2TOR-Second language tutoring using social robots. Proc. of Int. Workshop on Educational Robots.

[4] Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational researcher, 13(6), 4-16. https://doi.org/10.3102/0013189X013006004

[5] Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(6), 1482. https://doi.org/10.1037/0278-7393.22.6.1482

[6] Brusk, J., Wik, P., & Hjalmarsson, A. (2007). DEAL: A serious game for CALL practicing conversational skills in the trade domain. Proceedings of SlaTE-Workshop on Speech and Language Technology in Education. Pennsylvania, USA, 2007.

[7] Dalby, J. & Kewley-Port, D. (1999). Explicit pronunciation training using automatic speech recognition technology. CALICO journal, 16(3), 425-445.

[8] Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. Computer Assisted Language Learning, 29(4), 833-851. https://doi.org/10.1080/09588221.2016.1167090

[9] Fandrianto, A. & Eskenazi, M. (2012). Prosodic Entrainment in an Information-driven Dialog System. In INTERSPEECH, p. 342-345.

[10] Ijuin, K., Umata, I., Kato, T., & Yamamoto, S. (2018). Difference in eye gaze for floor apportionment in native-and second-language conversations. Journal of nonverbal behavior, 42(1), 113-128. https://doi.org/10.1007/s10919-017-0262-3

[11] Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. Human-computer interaction, 19(1), 61-84. https://doi.org/10.1207/s15327051hci1901&2_4

[12] Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (March 2016). Social robot tutoring for child second-language learning. Proc. of Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference, 231-238.

[13] Khalifa, A., Tanizoe, T., Ishida, M., Katou, T., & Yamamoto, S. (2015). Joining-in-type humanoid robot assisted language learning systems. in Forum on Information Technology (FIT 2015), Matsuyama, Japan.

[14] Khalifa, A., Kato, T., & Yamamoto, S. (2016). Joining-in-type Humanoid Robot Assisted Language Learning System. Proc. of LREC 2016, pp 245-249.

[15] Khalifa, A., Kato, T., & Yamamoto, S. (2017). Measuring Effect of Repetitive Quires and Implicit Learning with Joining-in-Type Robot Assisted Language Learning System. Proc. of SLaTe 2017, pp 13-17. https://doi.org/10.21437/SLaTE.2017-3

[16] Khalifa, A., Kato, T., & Yamamoto, S. (2018). The retention Effect of Learning Grammatical Patterns Implicitly Using Repetitive Joining-in-Type Robot Assisted Language Learning System. Proc. of TSD 2018, (in press).

[17] Kuwa, R., Kato, T., & Yamamoto, S. (2018). Classification of Utterances Based on Multiple BLEU Scores for Translation-Game-Type CALL Systems, IEICE Trans. INF. & SYST., Vol. E101-D, No.3, pp. 750-757, DOI: 10.1587/transinf.2017EDP7151. https://doi.org/10.1587/transinf.2017EDP7151

[18] Lee, K., Kweon, S. O., Sungjin, L. E. E., Hyungjong, N. O. H., & Lee, G. G. (2014). POSTECH immersive English study (POMY): Dialog-based language learning game. IEICE TRANSACTIONS on Information and Systems, E97-D(7), 1830-1841. https://doi.org/10.1587/transinf.E97.D.1830

[19] Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., & Kim, M. (2011). On the effectiveness of robot-assisted language learning. ReCALL, 23(01), 25-58. https://doi.org/10.1017/S0958344010000273

[20] Leyzberg, D., Spaulding, S., Toneva, M., & Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains. Proc. of Cognitive Science 2012, p. 1882–1887.

[21] Lin, C. Y. & Och, F. J. (August 2004). Orange: A method for evaluating automatic evaluation metrics for machine translation. In Proceedings of the 20th international conference on Computational Linguistics, p. 501. Association for Computational Linguistics. https://doi.org/10.3115/1220355.1220427

[22] Long, M. H. (1981). Input, interaction, and second-language acquisition. Annals of the New York academy of sciences, 379(1), 259-278. https://doi.org/10.1111/j.1749-6632.1981.tb42014.x

[23] Mizukami, M., Yoshino, K., Neubig, G., Traum, D., & Nakamura, S. (September 2016). Analyzing the Effect of Entrainment on Dialogue Acts. In 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue SIGDIAL, p. 310-318. https://doi.org/10.18653/v1/W16-3640

[24] Morton, H. & Jack, M. A. (2005). Scenario-based spoken interaction with virtual agents. Computer Assisted Language Learning, 18(3), 171-191. https://doi.org/10.1080/09588220500173344

[25] Neri, A., Cucchiarini, C., & Strik, H. (2001). Effective feedback on L2 pronunciation in ASR-based CALL. Proc. of the workshop on Computer Assisted Language Learning, Artificial Intelligence in Education Conference. pp 40-48.

[26] Niederhoffer, K. G. & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. Journal of Language and Social Psychology, 21(4), 337-360. https://doi.org/10.1177/026192702237953

[27] Pickering, M. J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. Behavioral and brain sciences, 27(2), 169-190. https://doi.org/10.1017/S0140525X04000056

[28] Rayner, E., Bouillon, P., Tsourakis, N., Gerlach, J., Nakao, Y., & Baur, C. (2010). A multilingual CALL game based on speech translation. Proceeding of LREC, Valetta, Malta, 2010. http://archive-ouverte.unige.ch/unige:14926.

[29] Reitter, D. & Moore, J. D. (2007). Predicting success in dialogue. Proc. of Annual Meeting of ACL 2007, p. 808-815.

[30] Seneff, S., Wang, C., & Zhang, J. (2004). Spoken conversational interaction for language learning. Proc. of STIL/ICALL Symposium 2004. Venice, Italy.

[31] Seneff, S. & Wang, C. (October 2007). Web-based dialogue and translation games for spoken language learning. In SLaTE (pp. 9-16).

[32] VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (July 2005). When is reading just as effective as one-on-one interactive human tutoring. In Proceedings of the 27th Annual Meeting of the Cognitive Science Society (pp. 2259-2264).

[33] Wilcock, G. & Yamamoto, S. (2015). Towards computer-assisted language learning with robots, Wikipedia and CogInfoCom. In Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on pp. 115-119. IEEE. https://doi.org/10.1109/CogInfoCom.2015.7390575

[34] Yamamoto, S., Taguchi, K., Ijuin, K., Umata, I., & Nishida, M. (2015). Multimodal corpus of multiparty conversations in L1 and L2 languages and findings obtained from it. Language Resources and Evaluation, pp. 1–26. https://doi.org/10.1007/s10579-015-9299-2

[35] Young, D. J. (1991). Creating a Low-Anxiety Classroom Environment: What Does Language Anxiety Research Suggest? The modern language journal, 75(4), 426-437. https://doi.org/10.1111/j.1540-4781.1991.tb05378.x

[36] Watanabe, T. & Tsukada, S., (1992). Unknown utterance rejection using likelihood normalization based on syllable recognition. IEICE Trans. Inf. & Syst., Vol. E75-D, No. 12, pp. 2002-2009.

11   Appendix 1: Example of conversational scenario

No.	Speaker	Listener	Utterance
1	R1	R2	What do you think your mother will be given by your father for her birthday present?
2	R2	R1	I think my mother will be given a necklace by my father.
3	R1	R2	That's great.
4	R1	Learner	What were you given for a birthday present last year?
5	Learner	R1	I was given a Coffee maker.
6	R1	Learner	Good.
7	R1	Learner	What were you taught by your mother?
8	Learner	R1	I was taught how to cock by my mother
9	R1	Learner	I see.
10	R1	R2	What were you taught by your father?
11	R2	R1	I was taught how to drive by my father
12	R1	R2	Good.
13	R1	R2	While driving, I need a car navigation system.
14	R1	R2	Because my car doesn't have a car navigation system.
15	R1	Learner	Don't you have a car?
16	Learner	R1	Yes, I have car
17	R1	Learner	I see.
18	R1	R2	Don't you have a car?
19	R2	R1	No, I don't have a car.
20	R1	R2	OK.
21	R1	Learner	If you had a car and your car were broken, what would you do?
22	Learner	R1	I would call my father.
23	R1	Learner	Good.

12   Appendix 2: List of questions used in Questionnaire

Category	Questions
Previous experience	
	Have you traveled to an English-speaking country?
	If yes, how long did you stay?
	Have you experienced talking to a foreigner in English before?
	Did you take the TOEIC or TOEFL (or any similar) exam?
	If yes, what was it?
	If yes, what was your score?
	What is your major?
	Did you have any previous experience using robots?
Impression of blue robot (R1)	
	Did you like the robot?
	Did you feel pressure from the robot?

	Was the robot friendly?
	Do you think the robot understood you?
	Did the robot help you to understand the question?
	Did the robot wait a long time for you to answer?
	Did the robot speak quickly?
	How good was the robot's quality of voice?
	Did the robot understand you?
	Do you think the robot speaks English well?
Impression of orange robot (R2)	
	Did you like the robot?
	Did you feel pressure from the robot?
	Was the robot friendly?
	Did the robot understand you?
	Did the robot speak quickly?
	How good was the robot's quality of voice?
	Do you think the robot speaks English well?
Self-Evaluation (overall)	
	How good was your understanding of the conversation?
	How good was your reply?
	How confident were you?
	Did you get nervous when you spoke?
Self-Evaluation (while listening)	
	Did you watch the robot's upper body as a whole while listening?
	Did you watch the robot's face as a whole while listening?
	Did you watch the robot's eyes while listening?
	Did you concentrate when you listened to the robots?
Self-Evaluation (while speaking)	
	Did you watch the robot's upper body as a whole while speaking?
	Did you watch the robot's face as a whole while speaking?
	Did you watch the robot's eyes while speaking?
	Did you concentrate on your utterance?
About the Experiment	
	Did you enjoy the experiment today?
	Did you feel any stress?
	Did the Japanese sentence on the display help you?
	Do you think this is a good way to learn English?
	Please write any comments you may have

13   Authors

AlBara Khalifa is from the Department of Computer Science at Doshisha University, Kyoto, Japan. He has expertise in Human computer Interaction.

Tsuneo Kato is a Doctor of Engineering. He works at the Department of Intelligent Information Engineering and Sciences at Doshisha University, Kyoto, Japan. His expertise lies in Computer Engineering, Phonology and Phonetics.

Seiichi Yamamoto is a Faculty of Science and Engineering at the Department of Information Systems Design at Doshisha University, Kyoto, Japan. He did his Doctor in Engineering from Osaka University. His expertise lies in Psycholinguistics and Neurolinguistics.