Shih, JL. et al. (Eds.) (2023). Proceedings of the 31st International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Construction of an English Grammar Quiz Recommendation System Using Explanation by a Knowledge Map

Kensuke TAKII[a*], Naomichi TANIMURA[a], Brendan FLANAGAN[b] & Hiroaki OGATA[c]
[a]*Graduate School of Informatics, Kyoto University, Japan*
[b]*Center for Innovative Research and Education in Data Science, Kyoto University, Japan*
[c]*Academic Center for Computing and Media Studies, Kyoto University, Japan*
*kensuke.takii96@gmail.com

Abstract: Many systems to assist in learning English grammar have been developed in the field of learning material recommendation systems (LMRSs). Compared with those based on knowledge models, recommendations based on data tend to cause the cold-start problem, and it is said that explainable LMRSs may be able to enhance learners' motivation. In our research, we propose an explainable English grammar quiz recommendation system using a knowledge map to support students' learning of English grammar with trust in the system and motivation. The learning effect of the explanation of the system was evaluated in an experiment, in which 349 high school students in Japan participated. This experiment showed that there was little learning effect of the explanation, but the system reliability and motivation improved by the explanation. The limitation and our future work regarding the validity and learning effects of the system are also indicated.

Keywords: English grammar learning, e-learning recommender, knowledge map, learning path, explainable recommendation

1. Introduction

There has been increasing need for learning material recommendation systems (LMRSs), because of the difficulty for learners to select appropriate materials due to the increasing options (Liu, 2019). Although existing LMRSs can be classified into systems based on educational data and those based on domain knowledge models, the former is more likely to cause the cold-start problem for new users than the latter (Sun et al., 2019). Additionally, recent studies have developed explainable LMRSs such as Takami et al. (2022), which can explain the rationale of the recommendations. Since feedback from systems improves learners' motivation and achievement (Duffy and Azevedo, 2015), explainable LMRSs based on knowledge models may improve the learning effect and motivation in learning.

In the field of LMRSs, many systems to assist in learning English grammar have been developed (Fang et al., 2018). English is an important second language in many regions of the world, and the grammatical structures present in various languages pose challenges for reading comprehension in foreign languages (Alderson et al., 1984). According to previous research, teaching grammar items in a specific order is necessary to gain maximum educational effect (Izumi and Isahara, 2004). Hence, the order in which grammar items should be learned is vital in learning English as a second language (ESL).

In our research, we propose an explainable English grammar quiz recommendation system using a graph structure called a knowledge map. We developed an English grammar knowledge map as a directed graph structure that can represent English grammar items, their relations, and the order in which they should be learned. Besides, the recommendation using the knowledge map can explain its rationale by tracking learners' learning logs and recommended quizzes. In addition, we implemented the system into an existing web-based learning system and conducted an experiment targeting Japanese high school students who

are studying ESL to verify the effect of the explainable recommendation in learning English grammar. Our research focuses on the explanation of the rationale of the recommendations and the following research question is set: Does the explanation of the rationale of the recommendation for learning English grammar have a learning promotion effect?

2. Proposed Approach

2.1 English Grammar Knowledge Map

We developed an English grammar knowledge map, which represents English grammar items, the relations between them, and the learning order as a graph structure. We referred to CEFR-J Grammar Profile (Ishii and Tono, 2018), as a list of English grammar items. It includes 501 grammar items, and each item is classified into the CEFR (Council of Europe, 2011) level. Each item in the profile has a regular expression, which is used to judge if the item is used in a certain sentence. For example, the sentence “I am a student.” includes the grammar item “Personal pronoun nominative (I)+be: I am”, and the system can automatically determine that the sentence includes the grammar item by using the regular expression.

The knowledge map was generated based on the method by Flanagan et al. (2019): The nodes representing grammar items in CEFR-J Grammar Profile are connected by directed weighted edges, which represent the similarity of the regular expressions of the items and their proper learning order. The order was based on “English Grammar/Syntax Vintage” (Shinoda and Yoneyama, 2010), a workbook used for ESL learning in Japanese high schools. The knowledge map is structured as a directed maximum spanning tree by pruning the edges.

2.2 Explainable Quiz Recommendation with Knowledge Map

To recommend quizzes, our system first judges if each grammar item is learned or misunderstood by learners. The system creates two grammar sets using learners’ learning logs: The first is Incorrect Grammar Set (IGS), including grammar items in quizzes that the user has answered incorrectly, and the second is Unanswered Grammar Set (UGS), including grammar items in quizzes that the user has not yet answered. Then, the system recommends quizzes based on the following three conditions regarding the number of grammar items included in the quizzes. The higher these numbers, the more highly the quizzes are recommended; however, these 3 numbers are used in order of priority from 1 to 3. Besides, if there exists a directed edge from A to B in the knowledge map, we call A “a prerequisite” of B.

1. The number of grammar items that are prerequisites for other grammar items in IGS, and that are also included in IGS
2. The number of grammar items that are prerequisites for other grammar items in IGS, and that are included in UGS
3. The number of grammar items that are not prerequisites for other grammar items in IGS, but which themselves would be included in IGS

In generating an explanation in the proposed system, when the system recommends grammar quizzes, the grammar items are distinguished by the following criteria: whether the learner has incorrectly answered quizzes including these items before, or the learner has not answered these quizzes yet. Furthermore, the system notifies the learner whether a certain grammar item is a prerequisite to be known for another item that the learner has misunderstood, i.e., the learner has incorrectly answered quizzes including the grammar item.

2.3 User Interface of the Platform

As an interface for answering quizzes, we adopted BookRoll (Flanagan and Ogata, 2018), a multifunctional e-book reader system, and registered English grammar quizzes in a workbook being used in Japan for learning English. BookRoll has many functions to realize e-learning environments; one of which is a quiz function, with which learners can answer quizzes

registered to materials. We made e-books for answering the grammar quizzes, by editing the contents of the workbook so that each page has one quiz. Users can solve the quizzes on BookRoll and register whether their answer for each quiz was correct or incorrect.

The user interface of our recommendation system provides learners with information on (1) quizzes recommended by the system, (2) links to open the quizzes on BookRoll, and (3) explanations of why the recommendation was made for the learner. The interface displays 5 recommended quizzes, each of which has at most 3 types of explanations for why the quiz was recommended. These explanations clarify what kind of grammar items the user can learn with the recommended quizzes.

3. Experiment

3.1 Method Overview

We conducted an online A/B test, targeting high school students in Japan who are learning ESL by the curriculum in Japan. The target students were divided into an experimental group and a control group. Students in the experimental group used a recommendation system that displayed the explanation of the recommendation rationale, and students in the control group used a system that did not. Furthermore, we set the following 4 perspectives from which the effect of the explanations according to the RQ shown in the Introduction section:

- The number of recommendation usage: How highly did the students engage in learning?
- Scores of the pre- and post-test: Change of the learners' performance
- Answers for the poll: Willingness to learn with the system and trust in it

3.2 Settings

The experiment targeted 349 first graders in 9 classes at high school (about 40 students per class, the students who did not agree to participate were excluded). The participants were divided into the experimental group (4 classes, 152 students) and the control group (5 classes, 197 students), both of which used the recommendation system while learning for 14 days. This experiment was conducted during a winter vacation, and the students were encouraged in doing the learning activity with the system as a review, after learning English grammar with a paper workbook. This online learning activity was done with BookRoll, and 219 grammar quizzes were registered to it. Table 1 shows an example of grammar quizzes.

Table 1. *Examples of the quizzes registered to BookRoll*

Question	Answer
日本語の意味に合うように ( ) に適語を書きなさい。 (Write an appropriate word to match the Japanese sentence.) あなたがチームにいてくれて私たちはラッキーです。 We are ( ) ( ) have you in our team.	lucky to

3.3 Experimental Procedure

Step 1. The English teacher explains how to use the system to the participants.
Step 2. Before starting the learning activity, the participants take an English grammar pre-test.
Step 3. The participants solve some quizzes with BookRoll.
Step 4. The participants work on the quizzes using the recommendation and BookRoll.
Step 5. After ending the learning activity, the participants take an English grammar post-test.
Step 6. The participants answer the poll.

The contents of the pre and post tests were 5 grammar quizzes, which were randomly selected from the quizzes registered to BookRoll. The contents of both the quizzes were same. The poll was intended to an evaluation of the motivating effects of the system from the

viewpoints of trust and motivation in learning. This poll consists of 11 questions with a 5-Likert scale and 2 descriptive questions. For measuring the trust and motivation in using the system, we adopted Hoffman et al. (2018)'s "Trust Scale Recommended for XAI" and Keller (1987)'s ARCS model. The perspective of the questions is shown in the Table 3 in Result section.

4. Result

4.1 Usage of the System

Table 2. *Numbers of students who used the system*

	Participant	Recommendation	BookRoll	Both
With explanation	152	16 (10.5%)	10 (6.6%)	26 (17.1%)
Without explanation	197	18 (9.1%)	10 (5.1%)	28 (14.2%)
Total	349	34 (9.7%)	20 (5.7%)	54 (15.5%)

Table 2 shows the numbers of students who used both the recommendation system and BookRoll, who accessed the recommendation interface, and who used BookRoll only. This table indicates that about 10% of the participants used the recommendation and that about 6% of them used BookRoll only, without using the recommendation.

4.2 Result of the Poll

Table 3. *Result of Mann-Whitney's U test on the questions in a 5-Likert scale of the poll*

Question	Perspective	Group	# of Users	Mean	Std.	P-value
Q1	Trust	With explanation	14	3.29	0.61	P=0.59
		Without explanation	18	3.06	0.94	
Q2	Trust (predictability)	With explanation	14	3.64	1.01	P=0.58
		Without explanation	18	3.06	0.99	
Q3	Trust (reliability)	With explanation	14	2.93	1.00	P=0.69
		Without explanation	18	2.78	1.00	
Q4	Trust (efficiency)	With explanation	14	2.86	1.17	P=0.64
		Without explanation	18	2.61	1.04	
Q5	Trust (endorsement)	With explanation	14	2.57	1.16	P=0.98
		Without explanation	18	2.56	0.86	
Q6	Motivation (alert)	With explanation	14	2.57	0.94	P=0.79
		Without explanation	18	2.50	0.86	
Q7	Motivation (relation)	With explanation	14	3.14	1.10	P=0.56
		Without explanation	18	2.94	0.87	
Q8	Motivation (confidence)	With explanation	14	4.07	0.73	P=0.61
		Without explanation	18	4.00	0.69	
Q9	Motivation (satisfaction)	With explanation	14	2.79	1.12	P=0.66
		Without explanation	18	2.61	0.98	

Table 3 summarizes the result of the poll. It shows that from both the perspectives of trust and motivation, the impression of the students in the experimental group was more favorable than that of the students in the control group. In terms of Trust, there are differences of more than 0.2 points in the mean scores in the items of trust and efficiency. In terms of Motivations, there are differences of more than 0.2 points in the items of relation. In the free-text items, the answers regarding the efficiency in review and finding weak points were the most common response which was favorable for the recommendation system. On the other hand, as the response that was not favorable, we found the answers that doubted the correctness of the diagnosis by the system and found selecting the quizzes troublesome. In addition, some students found the system interface inefficient.

4.3 Result of the Pre- and Post-test

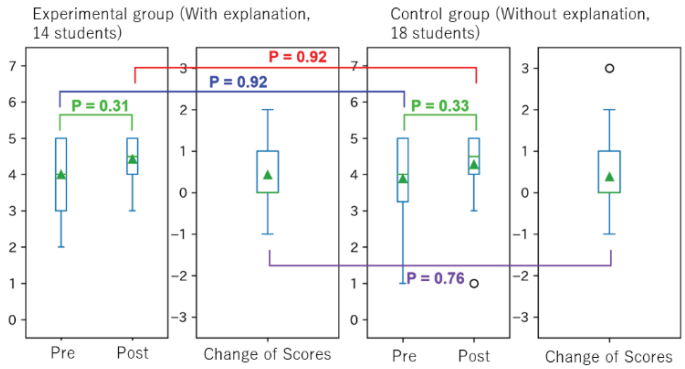

Figure 1. Result of the pre- and post-test and the change of the scores

Figure 1 shows the distribution of the results of the pre- and post-test and the change in the scores. The scores of the post-test were higher than those of the pre-test, but this difference was not statistically significant. Comparing the experimental and control groups, the scores of the pre- and post-test and the change in the scores did not indicate significant differences.

5. Discussion

5.1 Learning Promotion Effect by the Explainable Recommendation

In the post-poll, there were relatively large differences on the items that asked about trust, predictability, reliability, and efficiency in the Trust items. This implies that the explanation of the recommendation rationale improved the reliability of the system as perceived by the users. On the other hand, the item that asked about endorsement showed less difference. Considering that many answers for the free-text questions referred to the difficulty in using the system, the use of the system itself was a burden in learning. Also, the reliability of the explanation would be less influential on the system acceptance than the system's user-friendliness. In the question about Motivation, there were relatively large differences on the items about relation and satisfaction, which implies that the display of explanation of the recommendation rationale made users feel the relationship between their proficiency and learning contents proposed by the system. Furthermore, in the free-text questions, some users in the control group doubted the reliability of the diagnosis by the system, and such answers could not be found in the experimental group. This implies that it is difficult for users to judge the correctness of the recommendation without an explanation of the rationale.

The result of the grammar tests shows no significant differences between the two groups, and the post-test scores were higher than the pre-test scores but not significantly different from them. This implies that the student did not use the system often enough for the difference in the condition to influence the scores. Also, the number of quizzes in pre and post tests may be so small that no significant differences in the grammar tests were found.

5.2 Limitation

Our system was used infrequently regardless of conditions. We argue that the main reasons were lack of usability, and the burden in reading the explanation. When the students solve the quizzes with BookRoll and the proposed system, they must switch their screens many times. Second, if reading the explanation requires long time, it may be the users' burden to propose the explanation. Next, our system does not consider the level of importance of each grammar item in learning. Each quiz includes several grammar items, some of which are necessary to solve it, but others are not. Without considering the importance of each grammar item, the system might recommend quizzes that include items not important for solving them.

6. Conclusion and Future Work

Our experimental result did not show a learning effect from the explainable recommendation but implied that the system reliability and motivation were improved. Therefore, the answer to our research question was that the explainable recommendation improved the users' learning in the aspects of (1) reliability and motivation and (2) the frequency of the system use.

Our future work will focus on the validity and learning effects of the system. First, we should verify the accuracy of the knowledge map and the learning effects by considering learning paths. Second, the sufficient result has not been gained about whether the explanation of the recommendation rationale has a learning effect. The burden of using the system and understanding the explanation should also be investigated.

Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) 20H01722, JSPS Grant-in-Aid for Scientific Research (Exploratory) 21K19824, and NEDO JPNP20006.

References

Alderson, C. J., Alderson, J. C., & Urquhart, A. H. (Eds.). (1984). *Reading in a foreign language*. Longman Publishing Group.

Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, *52*, 338-348.

Fang, L., Tuan, L. A., Hui, S. C., & Wu, L. (2018). Personalized question recommendation for English grammar learning. *Expert Systems*, *35*(2), e12244.

Flanagan, B., Chen, M. A., Lecailliez, L., Majumdar, R., Akçapinar, G., Ocheja, P., & Ogata, H. (2019, December). Automatic vocabulary study map generation by semantic context and learning material analysis. In *Proceedings of the 27th International Conference on Computers in Education (ICCE2019)* (pp. 698-702).

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning*, *10*(4), 469-484.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608*.

Ishii, Y., & Tono, Y. (2018). Investigating Japanese EFL learners' overuse/underuse of English grammar categories and their relevance to CEFR levels. In *Proceedings of the 4th Asia Pacific Corpus Linguistics Conference* (pp. 160-165).

Izumi, E., & Isahara, H. (2004, December). Investigation into language learners' acquisition order based on an error analysis of a learner corpus. In *IWLeL 2004: an interactive workshop on language e-learning* (pp. 63-71).

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of instructional development*, *10*(3), 2.

Liu, T. Y. (2009). A context‐aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning*, *25*(6), 515-527.

Shinoda, S., Yoneyama, T. (2010). *Eibunpo Goho Vintage New Edition*, Iizuna Shoten. (*in Japanese*)

Sun, Z., Guo, Q., Yang, J., Fang, H., Guo, G., Zhang, J., & Burke, R. (2019). Research commentary on recommendations with side information: A survey and research directions. *Electronic Commerce Research and Applications*, *37*, 100879.

Takami, K., Dai, Y., Flanagan, B., & Ogata, H. (2022, March). Educational explainable recommender usage and its effectiveness in high school summer vacation assignment. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 458-464).