The Effects of Chatbot on Grammar Competence for Korean EFL College Students

Soojin Ahn
Visiting Professor, Liberal Art Institute, Incheon National University

한국 대학생 영어학습자들의 문법 습득에 있어 챗봇의 효과

안수진
인천대학교 기초교육원 객원교수

Abstract The purpose of this study was to test whether or not the AI chatbot is effective in acquiring target grammar for Korean EFL college students: prepositions and articles. A quasi-experiment was conducted with 46 first-year students taking part in a required English course. They were randomly divided into two groups: the experimental and control groups (23 students for each, respectively). The experimental group was engaged in six chat sessions with a chatbot over 6 weeks. A pretest and a posttest were used to examine the effectiveness of the chatbot by comparing any changes made in error frequencies of the target grammar in participants' English compositions. The results show that after a conversation with the chatbot, the experimental group significantly reduced the mean of omission errors in both prepositions and articles. To have a great effect in other error categories, chatbot feedback needs to be improved to reduce short responses or inaccurate utterances of students and induce them to actively participate in the conversation.

Key Words : chatbot, target grammar, prepositions, articles, Korean EFL college students

요 약 본 연구의 목적은 인공지능 챗봇이 한국 대학생 영어학습자들에게 전치사와 관사 두 목표 문법을 습득하는데 효과적인지 실험하기 위한 것이었다. 필수 영어 과목을 수강한 46명의 1학년 학생들을 대상으로 하였다. 참여자들은 실험군과 통제군. 두 그룹으로 무작위로 선별되었다 (각 그룹 23명). 실험 그룹은 6주 동안 챗봇과 6번의 채팅 세션을 가졌다. 사전·사후 실험을 통해 영어 작문에서 목표 문법의 오류 빈도의 변화를 비교함으로써 효과를 검토했다. 실험 결과는 챗봇과의 대화 후에 실험군이 전치사와 관사 모두에서 누락 오류의 평균을 유의하게 줄였다는 것을 보여줬다. 다른 오류 범주에서 큰 효과를 거두기 위해서는 학생들의 단답형 또는 부정확한 답변을 줄이고 챗봇과의 대화에 적극적으로 참여하도록 유도하는 챗봇 피드백이 개선되어야 한다.

주제어 : 챗봇, 목표 문법, 전치사, 관사, 한국 대학생 영어학습자

1. Introduction

One of the most difficult grammar for English learners to master is article and preposition, and it is known that they account for 20% to 50% of English learners' English writing errors[1]. In general, the use of incorrect prepositions or articles is observed in English composition by English as a Foreign Language (EFL) learners, which hinders the understanding of sentences.

Meanwhile, over the past decades, there has been an interest in various technology/computer-assisted language learning, including websites, virtual environments, and online chatting, to provide help to language learners. In the era of the 4th Industrial Revolution, there is a growing number of studies seeking if AI chatbot is effective in teaching and learning English. Chatbot, which is a kind of computer application, allows learners to communicate via text or voice[2]. Particularly, chatbots are proposed as efficient English education tools in that they offer EFL learners with fewer opportunities to communicate in English an authentic environment to speak English continuously without time and space constraints, as well as increase confidence, motivation, and interest in English[3,4].

Creating an AI chatbot needs IT and coding skills, which is why not many language teachers use it for language teaching methods[5]. Fortunately, these days, chatbot platforms such as Google's Dialogflow have emerged allowing users to program according to their needs without professional programming knowledge. Most of the research on AI chatbots as an English learning medium is currently focused on communication between chatbots and users[6,7]. Yang et al.[6] and Xu et al.[7] proved that the chatbot-guided conversation using Dialogflow was turned out to help children improve their English fluency. The number of utterances was increased while children interact with the chatbot[6], and children improve reading comprehension while conversing with the chatbot[7]. The studies claimed that the participating children were generally positive and active in conversation with the experimental English chatbot.

On the other hand, research on whether chatbots can provide effective feedback regarding grammar errors of English learners is very rare. For example, Kwon et al.[8] examined the possibility if a chatbot called Genie Tutor gave corrective feedback on grammatical errors after having conversations with English learners about a given topic, The result showed that grammar error correction had a success rate of 79.2%. However, they claimed that GenieTutor had a fixed dialogue flow given a topic, so it did not generate natural conversations for the learner. Kim[9] also examined whether chatting activities with chatbot Replica were effective in improving the English grammar skills of Korean college students. For a total of 16 weeks, the experimental group engaged in chatting with the chatbot. T-test results indicated that the chatbot group showed statistically significant improvement and demonstrated a positive role of a chatbot in improving learners' grammar ability. Haristiani et al. [10] aimed to develop an application for language learning, namely *Gengobot* integrated with mobile instant messaging service LINE. *Gengobot* was a chatbot-based dictionary of Japanese grammar, with explanations in Japanese, Indonesian, and English. The test results showed that all its features successfully functioned as expected. They suggested that the application was user-friendly and suitable to encourage independent learning by allowing users to adjust their own learning pace. Despite the growing number of studies on chatbots in English education, limited attention has been devoted to the effects of chatbots on L2 grammar learning.

Particularly, Fryer and Carpenter[11] mentioned there is yet no chatbot designed from the

'bottom up' to meet foreign language learners' needs. It is necessary to design a task-based chatbot to perform specific learning activities, not the existing conversational chatbot. Research on the effects of chatbots as a language learning medium or the development of chatbot programs suitable for language learning is still in its infancy[12]. In addition, a chatbot has not been widely applied in foreign language teaching at the tertiary level. College students are mature enough to be autonomous in learning a language and use IT technologies such as mobile and social networks. Therefore, this study aims to develop and apply a chatbot for teaching a foreign language at the tertiary level through a specific grammar point of English: preposition and article. Prepositions and articles are prominently difficult for Korean EFL learners because of the polysemous nature of prepositions and articles in front of a countable and uncountable noun that is not present in the Korean language. The following research question is thus addressed:

RQ: What are the effects of chatbot on grammar competence for Korean EFL college students?

2. Method

2.1 Participants

The participants in the present study are sixty-three Korean university freshmen enrolled in the required English course in the spring semester of 2021 taught by the researcher. They were all first-year students from various majors who had not lived in English-speaking countries and they have taken at least 6 years of public English education in Korea. Participants in the study were randomly divided into two groups: an experimental group and a control group. Participants' English skills may differ due to differences in learning or personal experience, so it was confirmed that there was no difference between the two groups by examining the frequency of target grammar errors in the participants' English composition written before the experiment. As seen in Table 1, an independent t-test showed that there was no statistically significant difference between the groups before the experiment. In the case of the experimental group, only students who have completed a total of six chat sessions with the chatbot were included in the experiment. Thus, a total of 46 were recruited for this study, 23 students for each, respectively.

2.2 Instruments

To create an experimental English chatbot, first, you sign in to Dialogflow and create an agent according to the topic of conversation. For example, the agent name 'wrongpreposition' creates various intents necessary to proceed with the conversation. For each intent, you create training phrases (phrases for what the user might say) and text response (phrases for what the chatbot says). Take an example of this study, you input 'what is the missing preposition in the following sentences?' in text response and then train your chatbot to recognize various user responses such as 'for,' 'I think of,' 'missing preposition is about' and so on (about 10-20 possible answers) in training phrases. The advantage of Dialogflow is that it is possible to make conversation flow according to the developer's intents without coding input. As the conversation flow expands in this linear flowchart, it is necessary to check whether input context and output context are correctly set in context. When constructing continuous conversations, dozens of intents are connected to each other, so when the context is not properly connected or the necessary intents are omitted, the conversation success rate decreases.

The target grammatical errors in the participants' English compositions were analyzed by using Grammarly, an online English grammar check software based on artificial intelligence. The frequency of errors was classified into three categories: wrong error, omission error, and additional error using the classification method of Corder[13], which was mainly used in previous research[14].

2.3 Procedure and Analysis

As the first week's assignment, all participants wrote an English composition with a length of 200 words under the topic of 'Do you agree to have a part-time job for college students?' Using Grammarly, it was confirmed that the most frequent grammar errors committed by the participants were prepositions and articles. A total of six chatbot dialogues were created based on preposition and article errors extracted from the students' English compositions ('wrong prepositions,' 'omission prepositions,' 'additional prepositions,' 'wrong article,' 'omission article,' and 'additional article'). This study used the chatbot platform Dialogflow provided by Google, which allows users to program according to their purpose without professional programming knowledge.

All participants received feedback on preposition and article errors extracted from English composition written in the first week. The control group excluded from the conversation with the chatbot was only provided with the teacher's explanation, and the experimental group was provided with both the explanation and the chatbot conversation. The researcher created Web Demo using the integration of the completed Dialogflow, and then gave the students the Web Demo address of the completed dialog and had them talk to Google Assistant on their smartphones, laptops, or PCs. The conversation between the user and the chatbot is converted into text on the screen, which the

participants were asked to submit as homework. Thus, the experimental group participated in a total of six chat sessions with the Dialogflow chatbot developed by the researcher for six weeks.

To find out the difference in error frequency before and after the conversation with the chatbot, after the chatbot conversation sessions for six weeks, the participants were asked to write English composition once again with the same topic as in the first week. A paired t-test was performed to examine the difference in the number of target grammar errors in the participants' English composition before and after the conversation with the chatbot. Before performing the paired t-test, the normality for the 12 independent variables was satisfied by performing a normality test.

3. Results

As seen in Table 1, it was confirmed that there was no difference between the two groups by examining the frequency of target grammar errors classified into three categories in the participants' English composition written before the experiment.

Table 1. Error Frequency before Intervention

	Wrong Preposition	Omission Preposition	Additional Preposition	Wrong Article	Omission Article	Additional Article
Experimental	15 (13%)	11(9%)	11(9%)	10(8%)	57 (48%)	14(12%)
Control	15 (13%)	10(9%)	11 (10%)	10(9%)	55 (48%)	14(12%)

Experimental (n=23)		Control (n=23)			
M	SD	M	SD	t	p
5.13	3.95	5.00	3.75	.110	.912

Table 1 shows that the frequency of errors between the groups before the experiment is almost the same. In other words, Korean EFL learners account for 48% of the total errors in

the omission articles, followed by wrong prepositions, additional articles, and the rest showing similar frequencies.

(1) a. (The) First reason is students can resolve the finance problems.
   b. I think that having a part time job is good *to (for) college students.
   c. As *a college student, we develop our abilities for our future career.

Example (1a) is a missing article error in which a definite article is omitted in front of the ordinal number, (1b) is a wrong preposition used for the preposition 'for,' and (1c) is an additional article error in which they should have used 'as college students' instead of 'as a college student' in association with 'we.'

<Table 2> and <Table 3> are the results of the paired t-test comparing the frequency of prepositions and article errors in the participants' English composition before and after the conversation with the chatbot.

Table 2. The Effect of Chatbot on a Preposition

Group	Wrong Preposition		Omission Preposition		Additional Preposition	
	Pre	Post	Pre	Post	Pre	Post
	M(SD)	M(SD)	M(SD)	M(SD)	M(SD)	M(SD)
Experimental (n=23)	1.50 (.70)	1.40 (.54)	2.0 (1.0)	1.22 (.44)	1.22 (.44)	1.16 (.40)
$t(p)$	.276 (.787)		1.957 (.079)*		.246 (.810)	
Control (n=23)	1.11 (.33)	1.22 (.44)	1.0 (.19)	1.12 (.35)	1.33 (.57)	1.0 (.17)
$t(p)$	−.603 (.555)		−.777 (.453)		1.000 (.374)	

*$p \leq .10$

For prepositions, the control group showed no difference in the mean number of preposition errors before and after the experiment. However, there was a difference in the means of the experimental group at the .10 level ($t = 1.957$, $p \leq .10$), albeit not that big a difference. The experimental group shows that in all preposition errors, the mean decreased after the experiment than before. However, the statistically significant

is only the omission preposition ($M = 2.0$ vs. $M = 1.22$). It can be inferred that the conversation with the chatbot helped reduce the omission of English prepositions for Korean college students. On the other hand, there seem to be likely explanations for why the conversation with the chatbot did not significantly affect other error categories.

First, even though input and output contexts were correctly set, the chatbot's flow was sometimes configured to the wrong intents, which results in conversation breakdown. The finding is aligned with the observation of previous research where the chatbot sometimes gets mismatched among different intents that have similar training phrases[15].

Another reason is that students tended to answer the chatbot's questions incorrectly as seen in Example 2.

(2) C: Okay, here is another question. Number one is I can help my mother blank my university registration fee. Number two is We can see this blank small. They have different prepositions.
   S: the blank is in.

Even though the chatbot stated that the answers were different prepositions, the student did not answer as induced by presenting just one preposition as the answer. This indicates that some students seemed not to perceive the chatbot as a conversational partner and were not seriously engaged in the tasks. This reminds of Fryer et al.'s observation[16] that language learners significantly decreased their interest in interactions with chatbots over time. It is necessary to guide them to positively and actively participate in the conversations with the chatbot.

(3) C: Here is last question. What are the missing prepositions in the following sentences? Each sentence has different

missing prepositions. Number one is we have to do it especially blank private school. Number two is blank work we can make a friend too. Number three is they study hard blank opposed to doing Alba. Number four is As soon as graduating blank high school.

S: from.

As seen in Example 3, students also tended to lose concentration when the chatbot's utterance becomes too long.

(4) C: Alright, here is another question. Is there any preposition to delete in the following sentence? Most of students don't realize to value of money.

S: to.

C: Sorry, could you say that again?

S: Most of students don't realize value of money.

C: Sorry, what was that?

S: Yes, to.

C: Say that one more time?

In addition, Example 4 shows that the conversation often discontinued when students uttered the untrained answers. The training phrase in the chatbot was 'of' but the student answered the unexpected phrase 'to' by which the conversation was not carried on.

Table 3. The Effect of Chatbot on an Article

Group	Wrong Article		Omission Article		Additional Article	
	Pre	Post	Pre	Post	Pre	Post
	M(SD)	M(SD)	M(SD)	M(SD)	M(SD)	M(SD)
Experimental (n=23)	1.66 (.81)	1.0 (.06)	3.16 (1.97)	1.66 (.72)	2.0 (1.0)	1.28 (.48)
$t(p)$	1.366 (.214)		2.780 (.009)**		1.698 (.115)	
Control (n=23)	1.50 (.57)	1.33 (.57)	2.93 (1.38)	2.81 (1.19)	1.53 (.87)	1.0 (.07)
$t(p)$	.378 (.721)		.182 (.857)		1.347 (.197)	

**$p \leq$ .01

For articles, the experimental group showed a statistically significant difference in the mean number of omission article errors before and after the experiment at the. 01 level ($t = 2.780$, $p \leq .01$). Both experimental and control groups show that in all article errors, the mean decreased after the experiment than before but there was no statistically significant difference in the control group. It can be inferred that the conversation with the chatbot particularly helped reduce the omission of English articles for Korean college students. Consideration must also be given as to why the conversation with the chatbot did not significantly affect other error categories.

(5) C: Alright. Here we go. Which one is correct? Number one? or Number two? Number one is Consequently, a part-time job becomes the social problem and number two is Consequently, a part-time job becomes a social problem.

S: number two.

C: Alright, the right answer is number two. Consequently, a part-time job becomes *a* social problem. It says general social problem, not specific one.

Example 5 shows that the students seemed to lose their concentration as they simply had to answer in a short response format like 'number two' or 'correct.' The finding is in accord with Hill, Ford, and Farreras' study[17] that compared to human conversation partners, shorter and more limited words were used in conversations with chatbots.

Example 6 also shows that students tended to answer the chatbot's questions incorrectly as seen in English prepositions.

(6) C: Alright, here is another question. What's wrong with the following sentences? Number one is we usually do a part-time

job from the evening to the late at night. Number two is We are no telling to quit the part-time.

S: <u>number one</u>.

C: I missed that, say that again?

S: <u>delete the.</u>

C: Okay, the answer is you have to delete THE. we usually do part-time job from the evening to late at night. We are no telling to quit part-time. You cannot put an article THE in front of adjective without noun.

The student incorrectly responded like 'number one' as seen in the first underlined student's utterance, but the follow-up conversation continued successfully when he/she uttered again the correct answer 'delete the.'

4. Discussion and Conclusion

Grammar is a fundamental element to ensure the reader understands what the author is trying to say, but there have been few attempts to investigate the effects of chatbots on improving English grammar competence. Thus, this research is of much importance because it attempts to fill the gap by examining whether the preposition and article errors of Korean EFL college students in English composition were reduced after the conversation with the chatbot. To recap the findings, for both prepositions and articles, the mean number of all error types in the experimental group decreased after the conversation with the chatbot. The statistically significant mean difference was the omission error type, which was much larger in the article than in the preposition. On the other hand, the control group showed no statistically significant difference before and after the experiment. It can be said that the conversation with the chatbot was effective in reducing the omission

errors of English prepositions and articles for Korean college students.

The current study sheds light on some implications for the future design of Dialogflow for Korean college students. First, the reason why the mean difference in other error types was not statistically significant after the conversation with the chatbot could be that too many sentences were put in the chatbot's utterances to reduce the interest of students. In future studies, it is necessary to devise a method of accessing grammar content through free conversation rather than focusing too much on an explanation of grammar. In addition, some findings in this study are not aligned with the observations of other research in which the interaction with chatbots lowered learners' anxiety and increased their interest. Thus, future studies are needed as to the way to reduce short responses and inaccurate utterances of students with low interest in conversations with chatbots and encourage them to actively participate in conversations. Finally, since there have been times when the intents of the conversation are mismatched, technical chatbot development education will be needed for English teachers to develop their chatbot models and use them as more effective learning tools.

This study clearly has some limitations: First, a small number of Korean EFL university students were sampled. Therefore, the findings should be confirmed with a wide range of individual difference variables related to interests, proficiency levels, school years, and so on. Second, this study focused on specific target grammar so the effects of chatbots on different grammar forms should be taken into consideration. Finally, future research needs to extend this line of research by adopting qualitative methods such as interview and survey responses to give a more in-depth analysis on the effectiveness of Dialogflow in English education.

Notwithstanding this limitation, this study will

be meaningful in that it provided the possibility that the students at the tertiary level benefit from a new learning experience by using an AI chatbot in teaching a specific point of a foreign language. Fewer studies have attempted to explore the effects of chatbots on EFL grammar competence. Therefore, this study implies the possibility of learning based on the interaction between chatbot and learners, not one－way learning and teaching. In English education, research on developing chatbots and verifying their effects, especially for college students, is still in its infancy, and continuous research is needed to estimate if chatbots using Dialogflow have the effect of improving EFL college students' English skills.

REFERENCES

[1] C. Leacock, M. Chodorow, M. Gamon & J. Tetreault. (2014). A*utomated Grammatical Error Detection for Language Learners.* Morgan & Claypool publishers.

[2] A. Kerly, P. Hall & S. Bull. (2006). Bring Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models. *Knowledge－based Systems, 20,* 177－185.

[3] A. Shevat. (2017). *Designing Bots: Creating Conversational Experiences.* Sebastopol, California, United States of America: O'Reilly Media, Inc.

[4] D. K. Shin. (2019). Exploring the Feasibility of AI Chatbots as a Tool for Improving Learners' Writing Competence of English. *Korean Journal of Teacher Education, 35(1),* 41－55. DOI:https://doi.org/10.14333/KJTE.2019.35.1.41

[5] T. Nghi, T. H. Phuc & N. T. Thang. (2019). Applying AI Chatbot for Teaching a Foreign Language: An Empirical Research. *International Journal of Scientific & Technology Research, 8(12),* 897－902. http://www.ijstr.org/final－print/dec2019/Applying－Ai－Chatbot－For－Teaching－A－Foreign－Language－An－Empirical－Research－.pdf

[6] H. Yang, H. Kim, D. K. Shin & J. H. Lee. (2019). A Study on Adopting AI－Based Chatbot in Elementary English－Speaking Classes. *Multimedia－Assisted Language Learning 22,* 184－205.

[7] Y. Xu, D. Wang, P. Collins, H. Lee & M. Warschauer. (2021). Same Benefits, Different Communication Patterns: Comparing Children's Reading with a Conversational Agent vs. a Human Partner. *Computers & Education, 161,* 104059.

[8] O. W. Kwon, K. S. Lee., Y. K. Kim & Y. Lee. (2015). GenieTutor: A Computer Assisted Second Language Learning System Based on Semantic and Grammar Correctness Evaluations, Helm, F., Bradley, L., & Thouësny, S. (Eds) *Proceedings of the EUROCALL* (pp. 330－335), Dublin: Research－publishing.net. Retrieved from https://files.eric.ed.gov/fulltext/ED564216.pdf

[9] N. Y. Kim. (2019). A Study on the Use of Artificial Intelligence Chatbots for Improving English Grammar Skills. *Journal of Digital Convergence, 17(8),* 37－46. DOI: https://doi.org/10.14400/JDC.2019.17.8.037

[10] N. Haristiani, A. A. Danuwijaya, M. M. Rifa'i & H. Sarila. (2019). Gengobot: A Chatbot－Based Grammar Application on Mobile Instant Messaging as Language Learning Medium. *Journal of Engineering Science and Technology, 14(6),* 3158 – 3173. Retrieved from https://jestec.taylors.edu.my/Vol%2014%20issue%206%20December%202019/14_6_7.pdf

[11] L. Fryer & R. Carpenter. (2006). Bots as Language Learning Tools. Language Learning and Technology. *Language Learning & Technology, 10(3),* 8－14. https://www.researchgate.net/publication/233816040_Bots_as_Language_Learning_Tools

[12] Y. Zhao & C. Lai. (2008). Technology and Second Language Learning: Promises and Problems, Parker, L. (Ed.) *Technology－mediated Learning Environments for Young English Learners* (pp. 167－205), London, UK: Lawrence Erlbaum Associates.

[13] S. Corder. (1973). *Introducing Applied Linguistics.* Harmondsworth: Penguin.

[14] A. A. Khansir. (2012). Error Analysis and Second Language Acquisition. *Theory and Practice in Language Studies, 2(5),* 1027－1032.

[15] E. Ruane, R. Young & A. Ventresque. (2020). Training a Chatbot with Microsoft LUIS: Effect of Intent Imbalance on Prediction Accuracy, *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion* (pp. 63-64). Retrieved from https://dl.acm.org/doi/10.1145/3379336.3381494

[16] L. K. Fryer, M. Ainley, A. Thompson, A. Gibson & Z. Sherlock. (2017). Stimulating and Sustaining Interest in a Language Course: An Experimental Comparison of Chatbot and Human Task Partners. *Computers in Human Behavior, 75,* 461－468.

[17] J. Hill, W. R. Ford & I. G. Farreras. (2015). Real Conversations with Artificial Intelligence: A Comparison Between Human-Human Online Conversations and Human-Chatbot Conversations. *Computers in Human Behavior, 49(3),* 240－250.

안 수 진(Soojin Ahn)   [정회원]

· 2007년   12월   :   Texas   A&M University 영어교육(문학박사)
· 2019년 9월 ~ 현재 : 인천대학교 기초 교육원 객원교수
· 관심분야 : 영어교육, 코퍼스
· E－Mail : sjahn@inu.ac.k