

Applying Ai Chatbot For Teaching A Foreign Language: An Empirical Research

Tran Tin Nghi, Tran Huu Phuc, Nguyen Tat Thang

Abstract: Chatbots or artificially intelligent conversational tools are the automatically new tools designed to interact humans and computers. The tool of chatbot system is very effective in marketing and launching new products. Using chatbot as a tool of learning with logical sequences of cognition has attracted a lot of attention from many foreign language centers, such as VUS, ILA, etc. This research is conducted to apply AI chatbot for helping students to learn a specific knowledge of a foreign language. The research also discusses students' interests and engagements, and performances in two ways of learning: with and without the help of AI chatbot via the case of teaching some English prepositions. 200 students were selected and divided into experimental and control groups (100 students for each respectively). The purpose of this empirical experiment is to test whether or not the AI chatbot is effective and useful for enhancing students' performance and engagement in learning a specific point of a foreign language. With the preliminary results, the students benefit a lot from a new learning experience with the use of AI chatbot in teaching. Most of them perceived AI chatbot tools as an essential part of their learning process. The AI chatbot also generates excitement and fun for their learning. The research may open up a field for language teachers to explore and apply for their teaching in the digital era.

Index Terms: AI chatbot, intelligent conversational tools, students' performance, ELT, Facebook app, CALL, mobile application

1. INTRODUCTION

IN the language class, language teachers normally spend four stages in teaching a foreign language lesson: warm-up, pre-practice, control practice and post-practice. This procedure always repeats from lesson to lesson, so students feel bored with it because the knowledge during some certain stages they might have already learned or known. It is time that language teachers should revise the procedures for teaching based on the applications of IT in the field of language teaching. In fact, there are a lot of pieces of evidence in the education of ITC applications that are applied to the field of language teaching and learning [6]. Various IT technologies have been used in ELT educational applications such as interactive webs, ELT software, synchronous chat, etc. to enhance the language learning for language learning and teaching. Chatbot is a program which is 'distinguished with the other computer applications through simulating an intelligent conversation with human users via auditory or textual methods' [4: 124]. In fact, AI Chatbot can help students to learn with the flexibility in giving the lessons due to the individual competence of the language of the participants. The magnitude and complexity of creating AI chatbot require a lot of skills and IT knowledge. That is why not many language teachers choose it for applying it in language teaching. Besides it, the learners who are willing to learn with an AI chatbot must be patient enough to respond to a series of automated messages which are sent from the chatbot programs. The idea of applying it for teaching a foreign language at tertiary level seems suitable because students are mature enough to force themselves for autonomous learning. It is high time that the language teachers applied IT technologies in their teaching to avoid making students bored in their learning and improve the quality of the learning process, especially with the mobile and social networks. This article presents the summary report of an empirical research for applying an AI chatbot for teaching a foreign language at tertiary level through a specific grammar point of English: preposition uses.

2. LITERATURE REVIEW

The applications of CALL to the field of language learning and

teaching is blossoming in a variety of forms, but most of them are focused on personalizing learning, promote interest in the subject of the field [6]. It should be noted that social networks are strongly affecting to every corner of human life, and education is not an exception. Although EFL websites for learning a foreign are blossoming, searching the content that meets with the EFL learners' needs is a time-consuming job [2]. EFL learners are overwhelmed with the variety of information that the EFL website provided. A well-built chatbot can help them end up wasting time and energy for surfing useless webpages with a very simple key-term search. Chatbot recently has been appearing to possess slightly better performance over former technologies such as cassette tapes, VCRs, CD/DVDs, MP3/4s, websites, etc. The field of chatbots has been intensively studied and new methods have been developed to show how students best learn and practice a new language in the target language. However, chatbots have not yet been fully explored and realized for the language learning community [5], although chatbots are all used as an effective tool for communicating with business customers. This matter is challenging enough to for language teachers do careful research to make it clearer. In our empirical experiment, the AI Chatbot is built with the purpose of motivation in learning English prepositions by testing them and letting them learn by what they are not good at through the entry test and train them for their better performance with logical deep learning arithmetic sequences.

3. METHODOLOGY

3.1. Participants

The participants were recruited from undergraduate students from Ho Chi Minh City University of Food Industry, Ho Chi Minh City, Vietnam on a voluntary basis. There are two groups of students: experimental and control groups. The initial participant number of each group is the same with 100 students. The first group is control in the teaching course for 15 periods of using English Prepositions without the help of Facebook chatbot. The second group does the same with 10

periods of class meeting and with the help of Facebook chatbot. All students are enrolled and taught by same teacher, also one of the researchers. The students are randomly selected ranked from the low to intermediate based on the TOEIC Scores. Students are divided into two groups with the equivalent levels.

3.2 Materials

The materials used in the research are 10 units in Part II of the book entitled English Pronouns and Prepositions which was written by Ed Swick. Prepositions are chosen for the research because they are sometimes underestimated. The topics in the materials present on prepositional forms (viz., location, movement and directions, time and mixed uses) and usages (viz., compound prepositions, prepositions with the other words, participial prepositions, postpositive prepositions, words requiring a specific preposition, prepositions and phrasal verbs). Students are eagerly willing to do all their assignments during the experiment. A chatbot for helping students to learn is also built for experimental groups with main contents attached from the mentioned course book English Pronouns and Prepositions. The chatbot is developed by Chatfuel, which is was established in 2015 with the goal of building bots on Facebook Messenger for marketing. The list of Chatfuel's clients vary from small to large businesses like Adidas, T-Mobile, LEGO, Golden State Warriors, etc. The chatbot is linked to the Facebook page at <https://www.facebook.com/englishprepositionuasages/>.

3.3 Procedures

3.3.1 The control group

All students are required to attend 3 class meetings with 5 periods for each time of the class meeting. They were taught 10 lessons from unit 13 to unit 22 in the course books titled English Pronouns and Prepositions [13]. Students are supplied with the paper and teachers lectured the lesson strictly on the course procedures. Then, they were tested what they have learned with the same with the experimental group.

3.3.1 The experimental group

Step 1: All students are introduced the chatbot which was built in Section 3.2 and lets them do the placement test.

Step 2: Students are asked to write the kind of prepositions they have got incorrect answers during the tests and the suggestions to join which Units in the course books.

Step 3: Students in the experimental group can choose the class to join with the given topics in the course book based on the suggestions chatbot of English Preposition Usages.

Step 4: Students were tested and surveyed with the students in the control group.

3.4 The research design

The research design presents an actual structure in which the research is set up and organized to compare and evaluate the effects of using chatbot for teaching a foreign language. The quantitative research design is subsumed into two types: longitudinal, cross-sectional and panel design structure and

experimental and quasi-experimental [11]. In this research paper, we combine these four kinds of quantitative research design to solve the questions we set out. Each type of research design is applied to each section of the research in the investigation.

3.4.1 Variable and measurement

This research study presents the methodological issues and approaches for identifying and measuring the effects of chatbot in EFL. In order to achieve this purpose, the researchers employ the descriptive method for describing all of the factors, both linguistic and nonlinguistic, the correlation between the variables under investigation and interview with the participants. Rasinger concludes that 'a variable or an outcome of the feature can vary considerably between cases and each case can only show one value for a particular feature' [11: 18]. In this research, the variable "Improving my language skills app", for example, have 5 possible values or outcomes, ranking from 1 to 5 points. The respondents can only choose one of the options for the notion "I intentionally use my mobile app for Improving my language skills" variable and they cannot choose more than one at the same time: option 1 for less than 1 hour, 2 for 2-3 hours, 3 for 3-4 hours, 4 for 4-5 hours, and 5 for larger than 5. For unit test results, we use 10-point scale for grading students' performance. Basing on the students' test result, we transfer the result into the variables: from 1 to 10 [8], [12].

3.4.2 The p-value in the statistics report

Every experimental research will have an effect or the difference between groups or not. The p-value that is less than 0.05 traditionally indicates the results are statistically significant, also including two values of 0.051 and 0.049. P is always italicized and capitalized. Normally, we do not use the number "0" before the decimal point (.) For p-value in statistical results, e, g: $P < .005$ because it cannot equal 1. According to Larson Hall, each statistical test has a certain mathematical symbol associated with it [9: 46]. In this research, these statistics are mainly calculated and resulted in numbers in T-test and an ANOVA. T- test is a test for group differences, even when we test two populations that we might think they are totally the same, there exists a variation in somewhat may be significant.

4. FINDINGS AND DISCUSSION

4.1. Students' distribution and frequency of leisure time

The survey shows the details of statistics on students' leisure time for the use of mobile applications in their daily life is not extensively for the purpose of studying with a mobile app. The Table 1 shows that 95% of students spent less than 1 hour on improving their language skills while they save most of their time for other activities such as Playing online games with 30.5% and Listening to music, watching YouTube favorite channels, Surfing the Facebook and other social network with 64.0%.

Table 1. Improving my language skills

	Frequency	Percent	Valid Percent	Cumulative Percent
< 1 hour	181	90.5	90.5	90.5
1 – 2 hours	14	7.0	7.0	97.5
2 – 3 hours	4	2.0	2.0	99.5
3 - 4 hours	1	.5	.5	100.0
Total	200	100.0	100.0	

The survey also investigates the other activities like: Reading news, stories, etc. Or chatting with friends or Checking email, schedules, and lesson assignments. The table 2 shows that the average of spending time for these activities is 1.46 hours and 1.16 hours. These figures are too small in comparison with the other activities in their leisure time. That is why we conduct a chatbot for their learning English as a foreign language because students are fond up with spending time with their mobiles.

Table 2. Statistics on Students' spending time for their mobile apps

	Improving my language skills	Playing online games	Reading news, stories, etc. or chatting with friends	Checking email, schedules, and lesson assignments	Listening to music, watching YouTube favourite channels, Surfing the Facebook and other social network
N	Valid 200 Missing 0	200	200	200	200
Mean	1.13	2.26	1.46	1.16	4.07
Std. Error of Mean	.030	.078	.053	.029	.067
Std. Deviation	.425	1.104	.756	.415	.948
Variance	.180	1.218	.571	.172	.900
Sum	225	452	291	231	814
Percentiles	25 1.00 50 1.00 75 1.00	1.00 1.00 2.00 2.00 3.00 3.00	1.00 1.00 1.00 1.00 2.00 2.00	1.00 1.00 1.00 1.00 1.00 1.00	3.00 3.00 4.00 4.00 5.00 5.00

This survey aims to investigate into the students' interests based on the four-phase model of interest development [3], [7] towards using mobile apps in their daily life. The learners' results of this investigation are carried out to rebust the theory of four-phase model of interest: triggered situational interest, maintained situational interest, emerging individual interest, and well-developed individual one for the later stage of this research.

4.2 Evaluation of chatbots for students' performance

As we have mentioned in the Section 3.2 and Section 3.3, the evaluation of chatbot towards students' performance for each unit in the book is weighed and compared to give the

differences in the control and experimental group. The Sig. $P = .000$ in the ANOVA test indicates that there is a difference in two observed groups: the control and experimental.

Table 3. ANOVA test

Average	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	21.780	1	21.780	24.475	.000
Within Groups	176.200	198	.890		
Total	197.980	199			

The Table 4 presents that there are differences in the students' results between control and experimental groups - the group with the application of chatbot, for each of the units in the coursebook. For the three unit topics Participial Prepositions, Postpositive Prepositions and Prepositions that indicate the location, the experimental group does better jobs than the control one with the mean differences ranking from 0.45 to 0.50, based on the grading scale of 10. There also exists a significant difference between the group control group and experimental in the unit entitled Prepositions that Indicate Movement or Direction (with Mean difference of 1.21), Prepositions that indicate time (with Mean difference of 1.05), Prepositions and Phrasal verbs (with Mean difference of 0.89), and A variety of Prepositions Uses (with Mean difference of 0.90). From these statistics, the chatbot in the experimental is so useful that it could help us to predict which units can be ignored and which unit should be taught carefully in the real teaching. These Mean differences only appear in the condition that all the variances are assumed to be equal and 95% Confidence Interval of the Difference is applied to. (See the Table 5). These findings contribute to the conclusion that chatbot is not only used for 'assessing the academic progress, predicting future performance and spotting out the potential issues' [4].

Table 4. Group Statistics

	define	N	Mean	Std. Deviation	Std. Error Mean
Prepositions That Indicate location	Control	100	6.21	1.465	.147
	Experimental	100	6.71	1.719	.172
Prepositions That Indicate Movement or Direction	Control	100	6.32	1.681	.168
	Experimental	100	7.53	1.201	.120
Prepositions that indicate time	Control	100	6.18	1.336	.134
	Experimental	100	7.23	1.427	.143
Participial Prepositions	Control	100	6.21	1.387	.139
	Experimental	100	6.66	1.616	.162
Postpositive Prepositions	Control	100	6.33	1.223	.122
	Experimental	100	6.85	1.566	.157
Prepositions and Phrasal verbs	Control	100	5.94	1.469	.147
	Experimental	100	6.83	1.724	.172
A variety of Prepositions Uses	Control	100	6.05	1.690	.169
	Experimental	100	6.95	1.617	.162

Table 5. Independent Samples Test

Levene's Test for Equality of Variances	T-test for Equality of Means

	F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Prepositions That Indicate location	2.540	.113	2.213	198	.028	-.500	.226	-.945	-.055
Prepositions That Indicate Movement or Direction	13.642	.000	5.856	198	.000	-1.210	.207	1.617	-.803
Prepositions that indicate time	.070	.791	5.370	198	.000	-1.050	.196	1.436	-.664
Participial Prepositions	2.363	.126	2.113	198	.036	-.450	.213	-.870	-.030
Postpositive Prepositions	6.474	.012	2.617	198	.010	-.520	.199	-.912	-.128
Words that require a specific Preposition	.020	.887	2.406	198	.017	-.560	.233	1.019	-.101

on	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed
Prepositions and Phrasal verbs	2.681	.103	3.930	198	.000	-.890	.226	1.337	-.443
A variety of Prepositions Uses	.473	.493	3.848	198	.000	-.900	.234	1.361	-.439

The findings also open up to the hypothesis that human-to-human dialogues are easier lost than human-to-chatbot dialogues. The comparison is early mentioned in the earlier research of ALICE chatbot 'which illustrates the strength or weakness as a human simulation towards linguistic features such as lexical, part of speech, and semantic differences' [1]. In general, the combination of the class meeting and the help of chatbot boost up the students' performance in most of units of the course book. However, students' performance maintains the same in the two units Compound Prepositions and Prepositions that combine with other words (See the Table 6). The difference may appear when the Significance of difference is set up at 10%. The lowest rank of the Mean difference for the Unit Test 4- Compound Prepositions between two groups is 1.80 and its highest Mean difference reaches to 0.70, and the figures for Unit test 5 - Prepositions that combine with other words are in turn 0.61. and 0.78 (See the Table 7).

Table 6. Group Statistics

	Define	N	Mean	Std. Deviation	Std. Error Mean
Compound Prepositions	Control	100	6.27	1.536	.154
	Experimental	100	6.53	1.617	.162
Prepositions that Combine with other words	Control	100	6.24	1.379	.138
	Experimental	100	6.60	1.633	.163

Table 7. Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Compound Prepositions	.045	.832	-1.166	198	.245	-.260	.223	-.700	.180
			-1.166	197.483	.245	-.260	.223	-.700	.180
Prepositions that Combine with other words	4.524	.035	-1.684	198	.094	-.360	.214	-.781	.061
			-1.684	192.595	.094	-.360	.214	-.782	.062

Improving my language skills	1.379	.459	-.207	3.004	.03	-.213	-.210	.203	.969	1.032
Playing online games	-.275	.133	-.185	2.064	.040	.043	-.146	.140	.571	1.751
Reading news, stories, etc. or chatting with friends	-.106	.198	.048	.536	.592	-.109	.038	.036	.584	1.712
Listening to music, watching YouTube favorite channels, Surfing the Facebook and other social network	.601	.176	.342	3.419	.001	.219	.238	.232	.457	2.187

a. Dependent Variable: Words that require a specific Preposition

When we further examine the correlations between the distribution of frequency of using mobile apps and results of students' performance, the results shows that Improving language skills and Listening to music, watching YouTube favorite channels, Surfing the Facebook and other social network increase the students' performance in mastering English preposition, except playing online games. The Table 8 presents the correlations between students' leisure activities and results for unit test 8 - Words that require a specific Preposition. The significant figures are below 0.5 and consistent with the meanings of affecting to the results.

Table 8. Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Correlations			Collinearity Statistics	
	B	Std. Error				Beta	Zero-order	Partial	Part	Tolerance
1 (Constant)	5.983	.960		6.231	.000					

5. CONCLUSION

Although there were a lot of limitations in the current research, the initial findings enhance our understanding about applying chatbots in ELT. The authors have combined Facebook auto messenger with conversational activities to learn a new knowledge of English grammar, which allowed learners to interact automatically at anytime and anywhere. Chatbots have proved themselves as a very useful tool to help students engage in the lessons, especially in practice activities with mobile-based application through social networks. Students are eager to share their understandings and their performance with their peers. Supported by data in the current research, a grammar point is best understood with students' autonomy in learning with their good moods in a relaxed way. It was concluded that the research was consistent with previous research in applying ITC in language teaching and learning. Then the students had to manage to complete the duties assigned by ITC applications to them and soon they loved the new way of learning, especially with a chatbot in the experimental. The students' performance is improved a lot when they are learning with the combination of class meeting and chatbot for learning a foreign language.

REFERENCES

[1] Abu, B., & Atwell, E. (2008). Evaluation of Chatbot Information System. In Eighth Maghrebian Conference on Software Engineering and Artificial Intelligence (pp.

- 1–12). Retrieved from <https://pdfs.semanticscholar.org/7f05/22051e7de60f8166563cc381a12142db030e.pdf>
- [2] Angheliescu, P., & Nicolaescu, S. V. (2018). Chatbot Application using Search Engines and Teaching Methods. In 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1–6). IEEE. <https://doi.org/10.1109/ECAI.2018.8678948>
- [3] Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: putting a Lexical Approach to the test. *Language Teaching Research*, 10(3), 245–261. <https://doi.org/10.1191/1362168806lr195oa>
- [4] Fei, Y., & Petrina, S. (2013). Using Learning Analytics to Understand the Design of an Intelligent Language Tutor – Chatbot Lucy. *International Journal of Advanced Computer Science and Applications*. <https://doi.org/10.14569/ijacsa.2013.041117>
- [5] Fryer, L., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior* (Vol. 93). <https://doi.org/10.1016/j.chb.2018.12.023>
- [6] Gastelú, C. A. T., Kiss, G., & Domínguez, A. L. (2015). ITC Integration Process in Basic Education: The Case of Veracruz City in Mexico. *Procedia - Social and Behavioral Sciences*. <https://doi.org/10.1016/j.sbspro.2015.01.545>
- [7] Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4
- [8] Howell, D. C. (1997). *Statistical methods for psychology*. Duxbury Press. Retrieved from https://books.google.com.vn/books/about/Statistical_Methods_for_Psychology.html?id=ktkkAQAIAAJ&redir_esc=y
- [9] Larson-Hall, J. (2010). *A Guide to Doing Statistics in Second Language Research Using SPSS and R* (Second Edi). New York and London: Routledge. Retrieved from <https://www.routledge.com/A-Guide-to-Doing-Statistics-in-Second-Language-Research-Using-SPSS-and/Larson-Hall/p/book/9781138024571>
- [10] Macaro, E., & Masterman, L. (2006). Does intensive explicit grammar instruction make all the difference? *Language Teaching Research*, 10(3), 297–327. <https://doi.org/10.1191/1362168806lr197oa>
- [11] Rasinger, S. M. (2013). Quantitative research in linguistics: an introduction. Retrieved from <https://www.bloomsbury.com/uk/quantitative-research-in-linguistics-9781472566966/>
- [12] Sharma, D. (2016). Language transfer and discourse universals in Indian English article use. Retrieved from <https://www.semanticscholar.org/paper/Language-transfer-and-discourse-universals-in-use-Sharma/b07afce5d5a93ef88e6ff26a697709d4d5bf0f70>
- [13] Swick, E. (2011). *English pronouns and prepositions*. McGraw-Hill.