

USING A MULTILINGUAL CHATBOT FOR FOREIGN LANGUAGE PRACTICE

Mahesh Vanjani, Texas Southern University, mahesh.vanjani@tsu.edu

Milam Aiken, University of Mississippi, maiken@bus.olemiss.edu

Mina Park, Southern Connecticut State University, parkm2@southernct.edu

ABSTRACT

Chatbots, or conversational agents, are computer programs that emulate the way people talk to each other, providing a more natural user interface. Students learning a foreign language might benefit from using the systems to practice their skills. In this study, we asked 49 students to evaluate transcripts of conversations with a multilingual chatbot to assess the accuracy of translations and appropriateness of replies. Results show that translations and responses were generally very good. Students gave poor ratings for only four of 24 responses, and the vast majority thought the system would help with acquiring a new language.

Keywords: Chatbot, Foreign Language, Conversational Agent, Multilingual

INTRODUCTION

As educators we aspire for our students to be active and productive participants in a multilingual world. As such, dual-language or multilingual skills, and, multicultural education is crucial. Contrary to this, in January 2019, the Modern Language Association noted that from 2013 to 2016, colleges across the United States cut 651 foreign language programs (Johnson, 2019). French was the hardest hit, losing 129 programs, followed by Spanish with 118, German with 86 and Italian with 56. Once these programs close, they are very hard to reopen. Given tightening budgets in Higher Education such trends continue. However, students and, for that matter, anyone interested in learning a foreign language can utilize free online options available. One such option is to learn a foreign language using an online multilingual chatbot.

Twenty-four polyglot experts were asked to provide their two most useful tips to learn a new language. The common theme of their suggestions was to start using the language as soon as possible. Do not get bogged down by trying to learn and understand everything and be able to speak with perfect pronunciation and enunciation. Instead use the language as often as possible and create as many opportunities as possible to speak the language. Daily practice is best. However, with a paucity of proximate foreign language speakers in the foreign language that you are interested in this can be difficult. That is where multilingual chatbots come in and serve as a great digital tool to practice and learn a language regularly and without obligation or embarrassment.

When learning a new language, students should practice what they have learned in a conversation. However, they might be shy or embarrassed about making mistakes when talking to another human (Fryer, et al., 2006). Instead, they might use technology-enhanced learning. For example, some Web sites such as Duolingo (<http://bots.duolingo.com/>) allow students to converse in several different languages including French, Spanish, and German using a chatbot, a computer program that emulates human conversation (Coniam, 2008; De Gasperis & Florio, 2012). These virtual humans are available at any time, and learners might be more motivated to practice with them (Jia & Chen, 2008). General purpose conversational agents have been used for tutoring and learning companions (Eynon, et al., 2009; Feng, et al., 2006; Pietro & Frontera, 2004; Veletsianos, et al., 2010; Vieira, et al., 2004), but relatively little research has been devoted to how the systems can improve language instruction.

Here, we describe a new multilingual chatbot incorporating 103 languages, far more than any other. The following section provides a background on these systems. We then describe the new system and test it with transcript evaluations.

BACKGROUND

Researchers have recognized that multilingual dialogue systems can assist with learning and teaching (Kerly, et al., 2007; Kerry, et al. 2009; Seneff, 2006). For example, Kerfoot, et al. (2006) used chatbots as a teaching tool for the training of medical students, and results showed that learning efficiency increased three-fold.

Chatbots can also be used for foreign language instruction, offering an alternative to traditional classroom-based conversation practice (Abu Shawar & Atwell, 2007). For example, the chatbot *Tutor Mike* (<http://bandore.pandorabots.com/pandora/talk?botid=ad1eeebfae345abc>) was specifically designed to help students practice their skills with English. Created by Professor Ron Lee, the system placed second in the 2018 Loebner Prize Contest (Ferrara, 2016), a version of the Turing Test for determining the presence of artificial intelligence. In addition to being able to emulate human conversation very well, it is very knowledgeable about languages, cultures, government, history, geography, and general information (Park, et al., 2019). Additionally, it might be friendlier than other systems, perhaps motivating people to use it more (Wang, et al., 2019).

Although it is possible to use English-based chatbots such as *Tutor Mike* to practice with other languages the experience can be very tedious and time consuming. The process involves manually copying-and-pasting a machine translation (e.g., French to English), getting an answer in English, and then translating with the service back to French manually. In one study (Park & Aiken, 2019), students took nearly 1.5 times longer with this manual approach than what they stated was their upper, acceptable time limit. In contrast, completely automated translation is at least 30 times faster than the manual process, which significantly increases satisfaction with the system. Few students would be willing to copy-and-paste translations when using a chatbot to achieve multilingual functionality.

As a consequence, researchers have developed a few chatbots that support multiple languages such as *Eggbun*, *Memrise*, *Mondly*, and *Watson* (Mohammad, 2018). However, these support relatively few languages and no comprehensive studies have been conducted on how effective they are.

TUTOR MIKE MULTILINGUAL CHATBOT

We have developed a new chatbot linked with *Google Translate* (<https://translate.google.com/>) and the *Tutor Mike* chatbot that provides human conversation in any of 103 languages in 10,506 combinations. *Google Translate* was chosen because it is more accurate than other online translators overall, although it might not be best for some specific language pairs (Aiken, 2019). In addition, Google provides support for far more languages than competitors do.

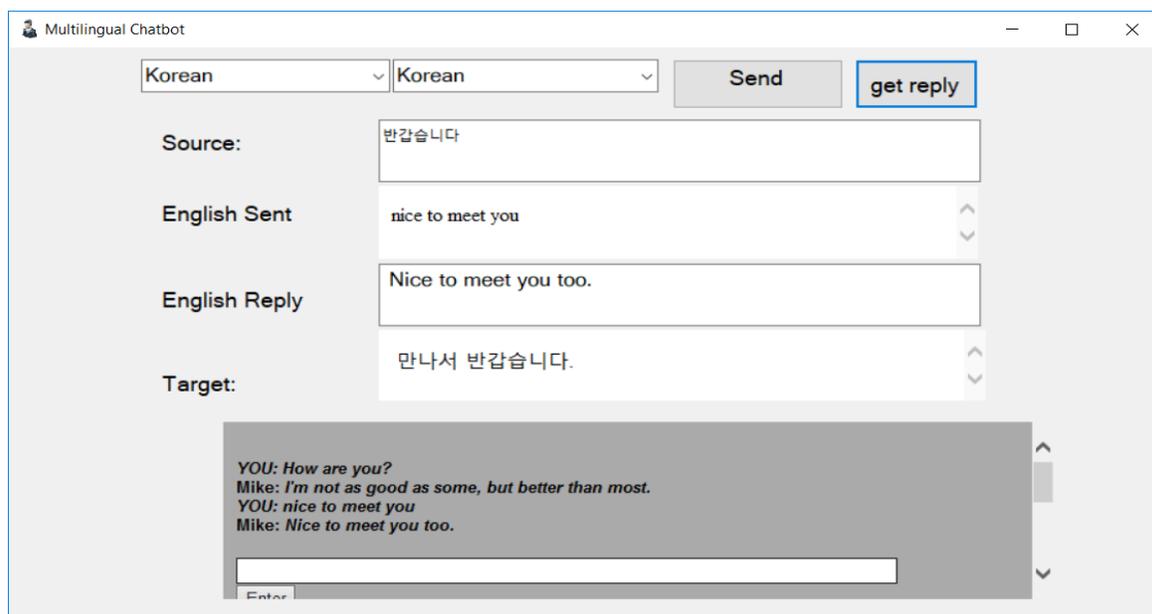


Figure 1: Multilingual chatbot showing Korean input and output

Figure 1 illustrates the use of this multilingual chatbot with Korean input and output, as an example. The user types something in the Source textbox, presses the Send button, and in less than a second gets the translation into English and the response from *Tutor Mike*. Finally, the user presses the Get Reply button to see the response in the source language. A web browser at the bottom of the application shows the ongoing conversation in English.

STUDY DESCRIPTION

In a prior study using the system (Vanjani, et al., 2019a), 12 comments were entered into the multilingual system using English and English translated from German, Spanish, and Korean: “What’s my name?”, “My name is Susan, what’s yours?”, “What’s up?”, “How now brown cow?”, “What color is bread?”, “Are you a chatbot?”, “How much wood could a woodchuck chuck if a woodchuck could chuck wood?”, “Do you have a voice?”, “Would you like a cup of tea?”, “Do you dream of electric sheep?”, “Could you repeat that please?”, and “Where are you right now?”. Students evaluated the English responses with scores ranging from a low of 1 if the system changed the subject to a high score of 10 if it answered a question correctly or had a good response. Although several of the questions were unusual, the results showed that the performance of the system was good using English, German, Spanish, and Korean, but two translations to Korean were somewhat awkward. In this study, 29 students evaluated all three language transcripts and their evaluations might have been biased by other responses. That is, they could see from the English dialog what was actually meant, e.g., “How now brown cow?” and thus the translation from Korean “Brown cow, what do you think?” for example, might have suffered in comparison.

To address the limitations in the prior study of the odd dialog and evaluation of all transcripts, we asked 49 students from a university in the northeastern region of the United States to evaluate a different set of transcripts from the multilingual chatbot. A total of 16 students analyzed responses in English given input in German, 17 analyzed from Spanish input, and 16 from Korean, giving ratings ranging from a low of 1 (poor) to a high of 7 (good) for understanding and grammatical accuracy of the source text (variables not in the previous study) and the appropriateness of the response from the chatbot. Finally, the survey asked students whether or not they believed the system would be good for foreign language practice.

We attempted to keep the input equivalent for all three scenarios, but due to differences in languages, the translations to English were not always the same. For example, the translation from German to English was ‘*Hi there*’ but the

translation from Spanish was simply ‘Hi’. Given the differences with the input, it is natural that the responses from the chatbot would also be different. For example, the response to the German input was ‘Hi! What’s up?’ and the response to the Spanish input was ‘Hi! What can I do for you today?’ In addition, even given identical input in English, the replies from the agent are not always equal, as some randomness is involved to keep the conversation interesting. Finally, some input was changed to fit the dialog. For example, the German input mentioned visiting Germany, while the Spanish input mentioned travelling to Spain. Therefore, when evaluating the three scenarios, the limitation of non-identical responses must be kept in mind.

STUDY RESULTS

The students’ response was generally positive and they thought the system would be good to practice conversing in a foreign language, giving an average rating of 6.08 (German=6.44, Spanish=5.76, Korean=6.06). The ratings were significantly greater than the neutral value of 4 on the 1 to 7 scale ($t=13.24$, $p < 0.001$).

Tables 1 to 3 show summaries of the questionnaire variables for each language. Table 1 summarizes responses from German input, Table 2 summarizes responses from Spanish input, and, Table 3 summarizes responses from Korean input.

Table 1: Responses from German input

Input	Understood	Accurate	Response	Appropriate
Hi there.	6.94	6.69	Hi! What’s up.	7.0
I am pleased to meet you.	6.81	6.94	Nice meeting you, too.	5.75
Do you understand German?	6.94	6.75	Just a little. Danke.	5.38
What’s your name?	7.00	6.88	My name is Mike. How are you doing?	6.13
Do you have a topic for your conversation?	6.94	6.88	Yes. How often do you clean your house?	5.50
The bathroom is cleaned twice a week, and the living room and bedroom are cleaned once a week.	7.00	6.56	All right.	5.13
Are you going to visit Germany?	7.00	7.00	I’m not certain.	6.69
Would you like to learn German to visit Germany?	5.88	5.25	Yes. I’d like to learn it. It’s one of my favorite things.	5.00

Table 2: Responses from Spanish input

Input	Understood	Accurate	Response	Appropriate
Hi.	7.00	7.00	Hi! What can I do for you today?	6.82
Pleasure.	5.24	6.41	What do you think of Trump?	2.00
You understand Spanish?	6.35	5.06	Just a little. Muchas gracias.	5.18
What is your name?	7.00	7.00	My name is Mike. How are you doing?	6.76
Do you have a topic for your conversation?	6.53	6.35	Yes. How often do you clean your house?	5.59
The bathroom is cleaned twice a week, and the living room and bedroom are cleaned once a week.	6.71	6.65	All right.	4.29
Will you see Spain?	5.29	4.76	Yes, I can see that.	2.88
Would you like to learn Spanish to visit Spain?	6.41	5.94	Yes. I'd like to learn it. It's one of my favorite things.	5.18

Table 3: Responses from Korean input

Input	Understood	Accurate	Response	Appropriate
Hello?	6.88	5.56	Hi! How are you today?	6.8
Nice to meet you.	6.56	6.88	Nice to meet you too.	6.38
Do you understand Korean?	6.88	6.81	Honestly, I don't know.	3.00
What's your name?	7.00	6.50	My name is Mike. How are you doing?	4.63
Do you have a topic for your conversation?	6.81	6.00	Yes. How often do you clean your house?	5.94
The bathroom is cleaned twice a week, and the living room and bedroom are cleaned once a week.	6.94	6.13	You're right.	2.00
Are you going to visit Korea?	7.00	6.69	No, I'm not.	6.88
Do you want to learn Korean for visiting Korea?	5.88	3.94	Yes. I can learn very quickly.	4.94

Students thought the translations from German were very good (understood mean=6.81, grammatical accuracy mean=5.25) and the responses in English from the system were good (appropriate mean=5.65). Evaluations of the Korean system were rated a little lower (understood=6.74, accurate=6.06, appropriate=5.07), and the Spanish system was rated worst (understood=6.32, accurate=6.15, appropriate=4.84), although the evaluations were still high. **Poor responses are indicated in bold font.**

Some students also added comments on the questionnaires, describing the errors. For example, one wrote: “good translations from users input, but some responses may be slightly out of context or short.” Another wrote about the “Danke” in the third German response: “This may be confusing if you do not know what that word means in German.” One student wrote: “But some of the responses don't make sense. If improved can be very good.” and another added: “It is definitely a good way to learn foreign language, at least at a conversational level.”

Some translations were somewhat colloquial. For example, the Spanish equivalent for ‘*Nice to meet you*’ was translated to English as ‘*Pleasure*’, i.e., ‘*It is a pleasure to meet you.*’ This might be a little confusing for the evaluator and might not be completely understood, but is normal in a human conversation. But, if the response is too unusual, it could detract from the quality of the discussion (Vanjani, et al., 2019b).

As with most chatbots, if the system does not understand the input or does not know a good response, it might change the topic completely. But, this can also occur during a human conversation. The chatbot could also change the topic randomly. For example, the Spanish system responded to ‘*Pleasure*’ with ‘*What do you think of Trump?*’ and this was rated as poor.

ACCURACY OF TRANSLATIONS

A native Korean speaker conducted a more thorough analysis of the Korean transcript. For example, entering “Hi, how are you?” in Korean using the more formal way and the less formal way, the system provided different translations (“Hello” for a formal way and “Good morning?” for a less formal way). However, the other sentences (#2 through #7) received the same English translation whether a formal or less formal Korean sentence was provided. All of the seven more formal Korean sentences were perfectly translated to English, and all of the chatbot’s responses were perfectly translated to Korean.

All of the less formal Korean sentences were perfectly translated to English, while one of the chatbot’s responses was partially translated. For the sentence “Sorry, I don’t know,” only the word “Sorry” was translated. For both the more formal Korean sentences and the less formal Korean sentences, all of the responses were of good quality and appropriate in conversations.

When “Do you understand Korean?” was asked in a formal Korean sentence, the chatbot avoided providing a clear answer by saying “Hmmm. A good question.” When the same question was asked in a less formal Korean sentence, the answer from the chatbot was “Sorry, I don’t know.” Neither of the responses were perfect or ideal; however, those responses would be good enough for typical conversations.

When “Do you have any topic that you would like to talk about?” was asked in a formal Korean sentence, the chatbot answered “Yes” and it changed the topic. When the answer for the changed topic (about house cleaning) was entered in a formal Korean sentence, the chatbot provided a short answer by saying “You’re right!” It is not considered an inappropriate response; instead, it was a rather short answer that did not sound natural in a conversation. When the same question was asked in a less formal Korean sentence, the chatbot wanted to talk about travel and asked about travel experiences. When the answer was provided in a less formal Korean sentence, the chatbot responded in a normal way by saying “Good for you.”

German and Spanish also have these formal and informal (familiar) aspects of language that are determined by the relative status of the people communicating. The chatbot often responds in the familiar style (e.g., “du” in German or “tu” in Spanish), indicating you are communicating with a friend which could make the user more comfortable with the system.

CONCLUSION

In this study, students reviewed transcripts generated with a multilingual chatbot using German, Spanish, and Korean as input. Like an earlier review of the system, results were good. Translations were generally understood and grammatically accurate, and responses in English were considered appropriate. There were only four poor replies out of 24 translations.

The study suffers from several limitations. As noted, all of the inputs were not identical due to language and culture variations and some randomness in *Google Translate*’s performance. That is, there can be several different, acceptable

translations of a given phrase. And with these different inputs, the response can naturally vary. Thus, a fair comparison among the three language transcripts is difficult.

Another limitation is the relatively short transcript, only eight sentences. A more thorough analysis with more text in other languages should be conducted. For example, *Google Translate* does well with most Western European languages, but does poorly with some less used languages such as Hebrew and Zulu (Aiken & Balan, 2011).

Finally, people make spelling and grammatical errors in their conversations with chatbots (Park, et al., 2018). Although the source spelling and grammar were correct in this study, future research should investigate the limits of errors on understanding (Vanjani, et al., 2015).

REFERENCES

- Abu Shawar, B. and Atwell, E. (2007). Fostering language learner autonomy via adaptive conversation. *Proceedings of Corpus Linguistics*.
- Aiken, M. (2019). An updated evaluation of Google Translate accuracy. *Studies in Linguistics and Literature*, 3(3), 253-260.
- Aiken, M. and Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*, 16(2).
- Coniam, D. (2008). Evaluating the language resources of chatbots for their potential in English as a second language. *Recall*, 20(1), 98-116.
- De Gasperis, G. and Florio N. (2012). Learning to read/type a second language in a chatbot enhanced environment. In: Vittorini P., Gennari R., Marenzi I., de la Prieta F., Rodríguez J. (eds). *International Workshop on Evidence-Based Technology Enhanced Learning. Advances in Intelligent and Soft Computing*, Vol. 152. Springer, Berlin, Heidelberg.
- Engle, J. (2019). How Important Is Knowing a Foreign Language. *New York Times*. Retrieved from <https://www.nytimes.com/2019/03/29/learning/how-important-is-knowing-a-foreign-language.html>
- Eynon, R., Davies, C., and Wilks, Y. (2009). The learning companion: An embodied conversational agent for learning. *Proceedings of the WebSci 2009: Society On-Line*.
- Feng, D., Shaw, E., Kim, J., and Hovy, E. (2006). An intelligent discussion-bot for answering student queries in threaded discussions. In: *Proceeding of the International Conference on Intelligent User Interfaces IUI*.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59 (7), 96-104.
- Fryer, L. and Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), 8-14.
- Jia, J. and Chen W. (2008). Motivate the learners to practice English through playing with chatbot CSIEC. In: Pan Z., Zhang X., El Rhalibi A., Woo W., Li Y. (eds) *Technologies for E-Learning and Digital Entertainment. Edutainment 2008. Lecture Notes in Computer Science*, Vol. 5093. Springer, Berlin, Heidelberg.
- Johnson, S. (2019). Colleges Lose a 'Stunning' 651 Foreign-Language Programs in 3 Years. *The Chronicle of Higher Education*, January 22, 2019.
- Kerfoot, P., Baker, H., Jackson, T., Hulbert, W. Federman, D. Oates, R., and DeWolf, W. (2006). A multi-institutional randomized controlled trial of Web-based teaching to medical students. *Academic Medicine*, 81(3), 224–230.

- Kerly, A., Hall, P., and Bull, S. (2007). Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems*, 20(2), 177–185.
- Kerry A., Ellis R., and Bull S. (2009). Conversational agents in E-Learning. In: Allen T., Ellis R., Petridis M. (eds). *Applications and Innovations in Intelligent Systems XVI*. SGAI 2008. Springer, London.
- Middlebury Interactive Languages (2019). Twenty-Four Polyglot Experts Reveal Two Most Useful Tips To Learn A New Language. *Middlebury Interactive Languages Summer Academy Blog*. Retrieved April 24, 2020, from <https://www.summerlanguageacademy.com/blog/language-learning>
- Mohammad, Z. (2018). Build multilingual chatbots with Watson Language Translator & Watson Assistant. *IBM Watson - AI Platform for the Enterprise* <https://medium.com/ibm-watson/>
- Park, M. and Aiken, M. (2019). Automated translation preference with a bilingual chatbot. *International Journal of Computers and Technology*, 19, 7502-7509.
- Park, M., Aiken, M., and Salvador, L. (2018). How do humans interact with chatbots?: An analysis of transcripts. *International Journal of Management and Information Technology*, 14, 3338-3350.
- Park, M., Aiken, M., and Vanjani, M. (2019). Evaluating the knowledge of conversational agents. *Southwestern Business Administration Journal*, 17(1).
- Pietro, O. and Frontera, G. (2004). Tutorbot: An application AIML based for web-learning. *Advanced Technology for Learning*, 2, 284-290.
- Seneff, S. (2006). Interactive computer aids for acquiring proficiency in Mandarin. In: Huo, Q., Ma, B., Chng, E.-S., Li, H. (eds.) *ISCSLP 2006. LNCS (LNAI)*, vol. 4274, pp. 1–12. Springer, Heidelberg.
- Vanjani, M., Aiken, M., and Park, M. (2019a). Chatbots for multilingual conversations. *Journal of Management Science and Business Intelligence*, 4(1), 19-24.
- Vanjani, M., Aiken, M., and Posey, J. (2019b). An evaluation of a multilingual chatbot. *Issues in Information Systems*, 20(1), 134-143.
- Vanjani, M., Aiken, M., and Park, M. (2015). A study of factors influencing machine translation comprehension. *Quarterly Review of Business Disciplines*, 2(2), 121-132.
- Veletsianos, G., Heller, R., Overmyer, S., and Procter, M. (2010). Conversational agents in virtual worlds: Bridging disciplines. *British Journal of Educational Technology*, 41 (1), 123-140.
- Vieira, A.C., Teixeira, L., Timteo, A., Tedesco, P., and Barros, F. (2004). Analyzing online collaborative dialogues: The OXEnCHL-Chat. *7th International Conference on Proceedings of the Intelligent Tutoring Systems*.
- Wang, J., Gu, L., and Aiken, M. (2019). A comparison of three conversational agents. *Asian Journal of Science and Technology*, 10(9), 10051-10055, September.