



Chatbots for Multilingual Conversations

Mahesh Vanjani^{1,*}, Milam Aiken² and Mina Park³

¹JHJ School of Business, Texas Southern University, Houston, TX 77004; ²School of Business Administration, University of Mississippi, University, MS 38677; ³School of Business, Southern Connecticut State University, New Haven, CT 06515

*Email: Mahesh.Vanjani@tsu.edu

Received on 05/21/2019; revised on 06/10/2019; published on 06/30/2019

Abstract

Chatbots emulate human dialogue to provide a more intuitive user interface to applications or simply provide entertainment. Chatbots rely on technology to function and new and emerging technologies such as NLP (Natural Language Processing) and AI (Artificial Intelligence) can be used to increase the ability of chatbots to emulate a more natural and free flowing conversation. As more and more mobile device users transition to increased use of texts and messaging chatbots can be used to provide consumers with multilingual support and services. While some chatbots have been developed in other languages, currently most converse only in English, and only a few can communicate in multiple languages. If configured correctly multilingual chatbots have the potential of providing a digital communication option that transcends language barriers. For our research we focused on the use of a chatbot that links the English-speaking Tutor Mike system with Google Translate, thus providing conversational capability in 103 languages, which is more than any other artificial multilingual agent is currently capable of. Two humans communicated with the system using German, Spanish, and Korean, and a group of undergraduate students reviewed the English translations of the chatbot's replies. Results show that the responses from German and Spanish were cogent and natural, but those from Korean were less understandable. As a caveat, Asian languages lack much of the linguistic nuances of European languages. For example, there may be no plural form or gender in the Asian language. Unlike German where nouns and adjectives constantly change endings depending on what they are doing in a sentence and, unlike Spanish, which have numerous verb conjugations, Asian languages require no such changes. This might impact the translation ability and quality of a multilingual chatbot. Additional research and enhancements can improve chatbots used for European to Asian language translations and vice-versa. Regardless, our research shows promising results in the future use of multilingual chatbots to allow communication across the globe with business potential in the use of such chatbots to provide customer service and online live interaction with customers across the world.

Keywords: Chatbot, Multilingual, Communication, Machine Translation

1 Introduction

A chatbot is a computer program or application to emulate conversations with human users via the internet. The goal of such chatbots is to use artificial intelligence techniques to convincingly simulate a human conversational partner. The chatbots' objective is to emulate human dialogue to provide a more intuitive user interface to applications or, sometimes, to simply provide entertainment. As the use of texting for communication via mobile devices becomes common the use of chatbots seems to be a natural extension of the use of applications to communicate effectively and especially so when the participants use different languages.

People are using chatbots (also known as bots, smartbots, talkbots, chatterbots, interactive agents, conversational agents, or conversational entities) more as a natural way to communicate, both for practical purposes (e.g., gathering information) and for amusement (Lommel, 2018;

Oppenheimer, 2016; Wu, 2017). One recent development is the use of these systems for enhanced learning (De Pietro & Frontera, 2005; Eynon, et al., 2009; Feng, et al., 2006; Veletsianos, et al., 2009; Vieira, et al., 2004), and some are being used as language learning tools (De Gasperis 2012; Ji, 2004; Kreisa, 2018; Tiwari, 2018).

Used for language practice, chatbots provide a partner with which to converse. Not only do the systems provide convenience, allowing practice at anytime, anywhere, but also they are capable of supporting many more languages than could reasonably be accommodated otherwise. That is, it might be possible to arrange for a 'pen pal' in Spanish, German, or French relatively quickly and easily, but it is much more difficult to get someone to communicate with you in Chichewa or Igbo. In addition, people might be less intimidated using a chatbot. People could be shy or awkward chatting with people in a new language, especially when they make mistakes (Fryer, et al., 2006).

This paper describes a new chatbot that provides communication in 103 languages, thus providing a tool for foreign language practice. Next, we discuss other multilingual systems and describe the new conversational agent. Finally, we evaluate the software and provide ideas for further research.

2 Background

To assist with learning a foreign language, a person could choose a chatbot that supports that language. For example, the Website Chatbot.orgs lists the following numbers of bots categorized by language: Arabic (4), Basque (7), Catalan (11), Chinese (4), Czech (2), Danish (11), Dutch (150), English (818), Finnish (1), French (111), Galician (2), Georgian (2), German (81), Greek (4), Hebrew (2), Hungarian (9), Indonesian (6), Italian (35), Japanese (4), Mandarin (10), Norwegian (7), Polish (155), Portuguese (20), Romanian (4), Russian (15), Slovak (1), Slovene (3), Spanish (95), Swahili (2), Swedish (11), and Turkish (16). However, most are not freely available on the Web and are not meant for conversation but rather for a specific task.

In addition, several multilingual chatbots have been developed including (Mohammad, 2018): Mondly - supports 30 languages, Memrise - 20, Watson - 21, and, Eggbun - 3 languages. Again, these systems are not free. To overcome this problem, people can use no-cost, publicly available chatbots that use a single language (typically, English) together with online translation software such as Google Translate (Figure 1) to chat in a large variety of languages. That is, they can copy and paste text from the conversational agent to the translation system and vice versa.

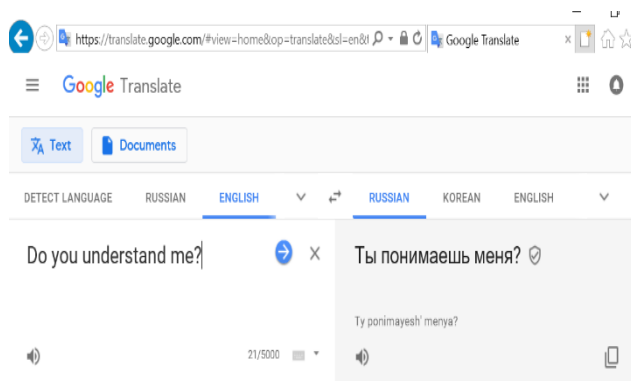


Figure 1: Use of Google Translate (English to Russian)
(URL - <https://translate.google.com/>)

Several conversational agents are available on the Web including Mitsuku, Elbot, jabberwacky, Amy Rose, and cleverbot. Another example, the chatbot Tutor Mike (Figure 2), was specifically designed to help people learn English, and the software informs them if they make mistakes.

However, this copy-and-paste solution is very tedious and time-consuming, taking possibly 5 or 6 seconds for each translation, and most people do not want to take the additional time and effort (Aiken, 2009). One study of computer system delay (Miller, 1968) showed that interactions lose their conversational nature after a maximum delay of 2 seconds, and a delay of about 0.5 seconds is the value resulting in highest conversational flow. Time delays affect satisfaction with a Web site, intentions to return, and number of tasks performed (Galletta, et al., 2004; Galletta, et al., 2006), and if the software is not easy to use, satisfaction is affected, and

people might not use the system again (Belanche, et al., 2012; Davis, 1989).

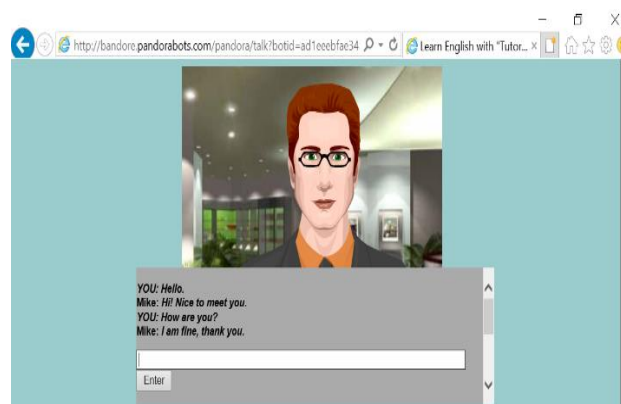


Figure 2: Tutor Mike Chatbot
(URL:

<http://bandore.pandorabots.com/pandora/talk?botid=ad1eeebfac345abc>)

A fully automated system can provide translation support to a chatbot in less than a second, thus increasing system satisfaction and enhancing intentions to return. In addition, it allows people to converse faster and generate more comments if the delays are shorter.

3 Software Description

We developed a program using Microsoft Visual Studio that linked Google Translate with the Tutor Mike chatbot, thus allowing users to chat in any of 103 different languages: Afrikaans, Albanian, Amharic, Arabic, Armenian, Azerbaijani, Basque, Belarusian, Bengali, Bosnian, Bulgarian, Catalan, Cebuano, Chichewa, Chinese, Corsican, Croatian, Czech, Danish, Dutch, English, Esperanto, Estonian, Filipino, Finnish, French, Frisian, Galician, Georgian, German, Greek, Gujarati, Haitian Creole, Hausa, Hawaiian, Hebrew, Hindi, Hmong, Hungarian, Icelandic, Igbo, Indonesian, Irish, Italian, Japanese, Javanese, Kannada, Kazakh, Khmer, Korean, Kurdish (Kurmanji), Kyrgyz, Lao, Latin, Latvian, Lithuanian, Luxembourgish, Macedonian, Malagasy, Malay, Malayalam, Maltese, Maori, Marathi, Mongolian, Myanmar (Burmese), Nepali, Norwegian, Pashto, Persian, Polish, Portuguese, Punjabi, Romanian, Russian, Samoan, Scots Gaelic, Serbian, Sesotho, Shona, Sindhi, Sinhala, Slovak, Slovenian, Somali, Spanish, Sundanese, Swahili, Swedish, Tajik, Tamil, Telugu, Thai, Turkish, Ukrainian, Urdu, Uzbek, Vietnamese, Welsh, Xhosa, Yiddish, Yoruba, and Zulu. In addition, the source language does not need to match the target language. That is, a user could type a comment in Spanish and receive a reply from the chatbot in Irish, for example, giving 10,506 combinations of languages.

Tutor Mike was chosen because it is a free, online service, and it performs well. It won second place in the 2018 Loebner Prize, fifth place in 2017, second place in 2016, fifth place in 2014, first place in 2013, and first place in 2011 (Loebner, 2018). In this contest, judges evaluate the ability of chatbots to emulate human conversation in an attempt to pass the Turing Test (Mauldin, 1994; Pinar Saygin, et al., 2000).

Figure 3 shows a sample screen using the software with Lao and Hindi selected. We obtained a phrase in Lao for "Do you speak English?" from www.omniglot.com and sent it to the system. It translated the Lao phrase as "Can you speak English?" and this was sent automatically to Tutor Mike. The reply from the chatbot was "English is my native language."

and this was translated to Hindi. The bottom of the window shows the chatbot operating in its natural form of English.

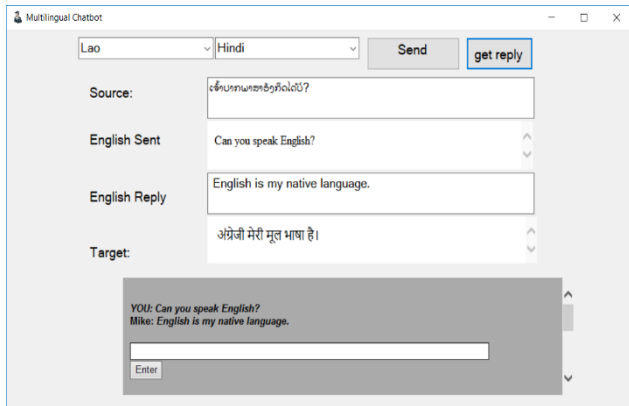


Figure 3: Sample Screen using the Multilingual Chatbot

4 Software Evaluation

4.1 Purpose

The purpose of this study was to evaluate the chatbot to ascertain whether it will function appropriately in an actual conversation with several languages.

4.2 Subjects and Task Description

We used a sample of 29 students (62% male) from a university in the northeast region of the United States to evaluate the responses from the system shown in **Tables 1 to 4**. All of the students spoke English fluently, but few knew any other language.

We entered 12 identical comments into the system in English, German, Spanish, and Korean, and the students scored the responses returned by Mike in English. Despite all of the input being the same (or equivalent), some input translations were different. For example, German and Spanish speakers would not say “What is up?” but rather “What is happening?” In addition, there is a certain amount of randomness in responses from the chatbot. That is, it does not always reply exactly the same way with identical input, especially for common questions or statements such as “How are you?”

We advised the students to give a score of 1 if the system changed the subject, 5 if it responded with something noncommittal, and 10 if it answered a question correctly or had a good response.

Finally, a German and Spanish speaker and a Korean speaker evaluated the responses translated into those languages.

Table 1. Chatbot discussion using English

	Input	Response from Mike	Mean Score
1	My name is Susan, what's yours?	Hi, nice to meet you, Susan! My name is Mike. How are you doing?	9.82

2	What's up?	Not much.	6.97
3	What's my name?	You are Susan, right?	9.62
4	How now brown cow?	Sorry, I don't know how.	5.07*
5	What color is bread?	Bread is food made of flour, water, and yeast, mixed together and baked.	2.34
6	Are you a chatbot?	I'm a human talking to you over the computer.	6.59
7	How much wood could a woodchuck chuck if a woodchuck could chuck wood?	Not much.	5.93*
8	Do you have a voice?	Yes. I have a beautiful voice.	9.17
9	Would you like a cup of tea?	Sure.	7.86
10	Do you dream of electric sheep?	Yes, I do. The Electric Sheep is a computing project that offers a free screensaver to participants.	6.72*
11	Could you repeat that please?	You said, " Do you dream of electric sheep?" I replied, " Yes, I do. The Electric Sheep is a computing project that offers a free screensaver to participants.".	9.34
12	Where are you right now?	We're in the same room. I'm right in front of you.	8.90

* Not significantly different from the neutral measure of 5 at $\alpha = 0.01$

Table 2. Chatbot discussion using German

(Text variations from English conversation are shown in bold font.)

	Input	Response from Mike	Mean Score
1	My name is Susan, what is your name?	Hello, Susan. My name is Mike. It's nice to meet you.	9.83
2	What's happening?	Not much.	7.00
3	What is my name?	You are Susan, right?	9.62
4	How is brown cow now?	Sorry, I don't know how.	5.66*
5	What color is bread?	Bread is food made of flour, water, and yeast, mixed together and baked.	2.28
6	Are you a chatbot?	I'm a human talking to you over the computer.	6.45*

7	How much wood could a woodchuck feed if a woodchuck could chop wood?	Not much.	5.41*
8	Do you have a voice?	Yes. I have a beautiful voice.	9.28
9	Do you want a cup of tea?	Sure.	7.93
10	Do you dream of electric sheep?	Yes, I do. The Electric Sheep is a computing project that offers a free screensaver to participants.	6.69*
11	Could you please repeat that?	You said, "Do you dream of electric sheep?" I replied, " Yes, I do. The Electric Sheep is a computing project that offers a free screensaver to participants.".	9.21
12	Where are you now?	We're in the same room. I'm right in front of you.	8.97

* Not significantly different from the neutral measure of 5 at $\alpha = 0.025$

Table 3. Chatbot discussion using Spanish

(Text variations from English conversation are shown in bold font.)

	Input	Response from Mike	Mean Score
1	My name is Susan, what is yours?	Hi, nice to meet you, Susan! Mike. How's it going?	8.76 ^
2	What happens?	I don't have an answer to that.	4.97*^
3	What is my name?	You are Susan, right?	9.72
4	How now brown cow?	Hmmmm! I'm not really quite sure how.	6.17*
5	What color is the bread?	Bread is food made of flour, water, and yeast, mixed together and baked.	2.62
6	Are you a chatbot?	I'm a human talking to you over the computer.	7.17
7	How much wood can a woodchuck throw if a woodchuck can throw wood?	I can't tell you exactly.	7.76^
8	Do you have a voice?	Yes. I have a beautiful voice.	9.31
9	Do you like a cup of tea?	Yes, I like tea.	8.03
10	Do you dream of electric sheep?	Yes, I do. The Electric Sheep is a computing project that offers a free screensaver to participants.	8.03

11	Could you repeat that please?	You said, " Do you dream of electric sheep?" I replied, " Yes, I do. The Electric Sheep is a computing project that offers a free screensaver to participants.".	6.76
12	Where are you right now?	We're in the same room. I'm right in front of you.	9.31

* Not significantly different from the neutral measure of 5 at $\alpha = 0.025$

^ Significantly different from English conversation rating for this comment at $\alpha = 0.025$

Table 4. Chatbot discussion using Korean

(Text variations from English conversation are shown in bold font.)

	Input	Response from Mike	Mean Score
1	My name is Susan, what is your name?	It's a pleasure meeting you, Susan. I'm Mike.	9.69
2	How is it going?	Things are fine. How about you?	9.76^
3	What is my name?	You are Susan, right?	9.62
4	Brown cow, what do you think?	I like the brown color.	6.59
5	What color is bread?	Bread is food made of flour, water, and yeast, mixed together and baked.	2.51
6	Are you a chat robot?	I'm just an ordinary human being.	6.59
7	How many trees can a guinea pig throw if a guinea pig throws a tree?	Not much.	4.27*
8	Do you have a voice?	Yes. I have a beautiful voice.	9.17
9	Would you like a cup of tea?	I'm not so sure.	5.45*^
10	Do you dream of electric sheep?	Yes, I do. The Electric Sheep is a computing project that offers a free screensaver to participants.	6.93
11	Could you repeat it?	You said, " Do you dream of electric sheep?" I replied, " Yes, I do. The Electric Sheep is a computing project that offers a free screensaver to participants.".	9.48
12	Where are you now?	We're in the same room. I'm right in front of you.	8.97

* Not significantly different from the neutral measure of 5 at $\alpha = 0.025$

[^] Significantly different from English conversation rating for this comment at $\alpha = 0.025$

4.3 Survey Analysis

Mean evaluations for each question or comment are shown in **Tables 1 to 4**. There was no significant difference between male and female evaluations ($p=0.54$), and there were no significant differences in the ratings (overall) between English and Spanish ($p=0.83$), English and German ($p=1.00$), and English and Korean ($p=0.95$). The mean ratings were 7.36 for English, 7.36 for German, 7.17 for Spanish, and 7.42 for Korean, all significantly above the mean rating of 5, indicating that the students thought that the system gave good answers, in general, for each language.

Three questions were designed to elicit vague or noncommittal answers (#2 What's up?, #4 How now brown cow?, and #7 How much wood could a woodchuck chuck if a woodchuck could chuck wood?). In the English conversation, however, only replies to #4 and #7 were not significantly different from the neutral rating. The reply to "#10 - Do you dream of electric sheep?" was also noncommittal, but it should have resulted in a more direct answer.

In German, replies to #4, #7, and #10 were also not significantly different from the noncommittal rating, but so was the reply to #6, "Are you a chatbot?", even though the reply was exactly the same as that in the English conversation.

In Spanish, the replies to #2 "What happens?" and #5 "What color is the bread?" were rated significantly lower than the equivalents in English because of the poor translations of the questions. However, the rating for #7 about woodchucks was rated significantly higher than the equivalent for English.

In Korean, replies to #2 and #4 were more definite than those in English, but for some reason, the reply was much vaguer to #9, "Would you like a cup of tea?"

4.4 Text Analysis

In the tables, input and output text variations from English are shown in bold. In many cases, the differences are negligible and don't affect the meaning. However, other differences sometimes had awkward wording. Only the Spanish input repeated "How now brown cow?" exactly, but this question can have different interpretations. However, the Korean "Brown cow, what do you think?" is probably the most different from the original. The Spanish input "What happens?" is very odd, and the Korean input "How many trees can a guinea pig throw if a guinea pig throws a tree?" was wrong because a woodchuck is not the same as a guinea pig.

In all three translations (German, Spanish, and Korean), the equivalent meanings to the English replies from Tutor Mike were very good, if not perfect. However, the words "Electric Sheep" were repeated sometimes and translated at other times. It might make sense to repeat literally the name of a project, a proper noun, rather than translate.

Although not a major problem, there were three responses in the German translation where the informal/personal "du" form was used, but just one instance of the formal "Sie". We believe the same level of formality should have been used throughout, but since the translations were made individually, the system had no record of the previous use.

In the Spanish reply of "Hi, nice to meet you, Susan! Mike. How's it going?", the system translated "Mike" as "Micro" instead of Miguel or just Mike. The meaning of "Micro" is very unclear.

Finally, in response to the question "Do you dream of Electric Sheep?", the English sentence "Yes, I do." was translated to Korean as "Yes I am like that too." which did not make sense. In addition, another reply ("Bread is mixed and then baked out of food that is made of flour, water, and yeast."), the added words "food that is made of" caused it to be a little awkward.

Thus, even though most of the conversations were natural, translations to German and Spanish had a few problems, and the Korean transcript was considered the worst. This might be expected because of the many similarities among English, German, and Spanish, but Korean is a radically different language.

4 Conclusion

In this study, students evaluated responses from a multilingual chatbot to determine its potential effectiveness in actual conversation using different languages. Results showed that the performance was good using English, German, Spanish, and Korean, but two translations to Korean were a little awkward.

One limitation of the study is that students evaluated only 12 responses from the system in four languages. Other comments in other languages might generate different results. For example, one study (Aiken & Balan, 2011) showed that Google Translate achieves greater accuracy using European languages than with Asian languages.

While this study does validate the potential of using multilingual chatbots for communication further studies should use a broader selection of inputs using additional languages. In addition, actual foreign language speakers should evaluate the system by typing in their own text and reading the responses rather than English speakers rating the transcripts.

References

- Aiken, M. (2009). Translating multilingual electronic meetings. *International Journal of Management and Information Systems*, 13(1), 35-46.
- Aiken, M. and Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal*, 16(2) April.
- Belanche, D., Casaló, L., and Guinalíu, M. (2012). Website usability, consumer satisfaction and the intention to use a website: The moderating effect of perceived risk. *Journal of Retailing and Consumer Services*, 19(1), 124-132.
- Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- 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*, 152. Springer, Berlin, Heidelberg.
- De Pietro, O. and Frontera, G. (2005): TutorBot: An application AIML-based for Web-learning. In: *Advanced Technology for Learning*.
- Eynon, R., Davies, C., and Wilks, Y. (2009): The learning companion: An embodied conversational agent for learning. In: *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*.

- Fryer, L. and Carpenter, R. (2006): Bots as language learning tools. *Language Learning and Technology*, 10(3), 8-14.
- Galletta, D., Henry, R., McCoy, S., and Polak, P. (2004). Web site delays: How tolerant are users? *Journal of the Association for Information Systems*, 5(1), 1-28.
- Galletta, D., Henry, R., McCoy, S., and Polak, P. (2006). When the wait isn't so bad: The interacting effects of website delay, familiarity, and breadth. *Information Systems Research*, 17(1), 1-104.
- Ji, J. (2004). The study of the application of a web-based chatbot system on the teaching of foreign languages. In: *Proceedings of the 15th Annual Conference of the Society for Information Technology and Teacher Education*, 1201-1207.
- Kreisa, M. (2018). Five resources for chatbots to be your language learning BFFs. <https://www.fluentu.com/blog/language-learning-chatbot/>
- Loebner, (2018). Loebner Prize. <https://www.aisb.org.uk/events/loebner-prize>
- Lommel, A. (2018). Plan ahead to build successful multilingual chatbots. Common Sense Advisory Blogs. <http://www.commonsenseadvisory.com/Blogs>
- Mauldin, M. (1994). Chatterbots, tinymuds, and the Turing Test: Entering the Loebner Prize Competition. *AAAI-94 Proceedings*, 16-21.
- Miller, R. (1968). Response time in man-computer conversational transactions. *Proceedings of the Fall Joint Computer Conference*, 267-277.
- Mohammad, Z. (2018). Build multilingual chatbots with Watson Language Translator & Watson Assistant. <https://medium.com/ibm-watson/>
- Oppenheimer, G. (2016). How to build a multilingual chatbot for billions of users. <https://venturebeat.com/>
- Pinar Saygin, A., Cicekli, I., and Akman, V. (2000). *Turing Test: 50 years later. Minds and Machines*, 10(4), 463-518.
- Tiwari, M. (2018). How multilingual chatbots will change the voice of business. *Forbes CommunityVoice*, August 13.
- Veletsianos, G., Heller, R., Overmyer, S., and Procter, M. (2010): Conversational agents in virtual worlds: Bridging disciplines. *British Journal of Educational Technology*.
- Vieira, A.C., Teixeira, L., Timteo, A., Tedesco, P., and Barros, F. (2004): Analyzing online collaborative dialogues: The OXEnTCHL-Chat. In: *7th International Conference on Proceedings of the Intelligent Tutoring Systems*.
- Wu, K. (2017). Chatbots are getting unsettlingly good at conversations. Inverse <https://www.inverse.com/article/37615-best-chatbot>