Chatbots

Jang Ho Lee, Hyejin Yang, Dongkwang Shin, and Heyoung Kim

In this series, we explore technology-related themes and topics. This series aims to discuss and demystify what may be new areas for some readers and to consider their relevance for English language teachers.

Introduction

Chatbot technology has gained a considerable amount of attention and popularity in a wide range of fields, including ELT. Chatbot research in ELT literature has discussed how chatbots could make learners feel more comfortable in English conversation than human interlocutors and also provide appropriate linguistic input (Fryer and Carpenter 2006), stimulate their interest in English learning (Fryer, Ainley, Thompson, Gibson, and Sherlock 2017; Gallacher, Thompson, and Howarth 2018), and contribute to their linguistic growth (Jia and Ruan 2008; Kim 2019). Despite growing interest among the research community, chatbots still appear to be a relatively rare technology among ELT professionals, including teachers and educators. Thus, the present article aims at providing an accessible introduction to this powerful resource for language teachers. It will introduce different types of chatbots, propose criteria for evaluating their appropriateness for ELT purposes, and present concrete examples of chatbot-based activities for language teachers. It concludes by calling for more empirical research on chatbots from the ELT community and by encouraging language teachers to take an interest in this technology.

Backgrounds

The fascination with computer systems that make human–computer conversational interactions possible dates back to the 1970s, when 'dialogue systems' were developed by the speech-processing computer science community. About 20 years later 'chatbots' emerged, which were initially developed by computer programmers for text-based conversation. The emergence of chatbots happened to coincide with the rise in demand for such conversational systems among language teachers (Atwell 1999), who believed that these systems could supplement their teaching by offering learners the opportunity to practise English.

From the 1970s to the present day, these conversational systems have been studied in the language-learning research community under different terms (or strands), including intelligent tutoring systems (see Swartz 1992), spoken dialogue systems and conversational agents (see Morton and Jack 2005), and chatbots (see Fryer and Nakao 2009). Experts have recently tended to synthesize these different strands of research under the overarching definition of 'dialogue-based computer-assisted language learning' (CALL) (Bibauw, François, and Desmet 2019).

Nowadays a clear distinction between dialogue systems and chatbots no longer exists, and this article will mainly use the latter term.

Different types of chatbots

Chatbots can be classified based on their purpose (i.e. assistant or conversation) and mode of communication (i.e. text or voice). Regarding the former, assistant chatbots are programmed to assist users in their daily activities, such as searching for information on the internet, making telephone calls, scheduling an appointment, and so on. They are integrated in mobile devices such as tablet computers and smartphones (e.g. Siri by Apple, Cortana by Microsoft), or in stand-alone devices such as smart artificial intelligence (AI) speakers. With most interactions, these chatbots provide short, structured, and simple responses to the users.

Conversation chatbots are built to engage in human-like conversations by offering more extended responses to users' input. Mitsuku (see Appendix for websites related to chatbots and chatbot builders introduced in the present article), originally developed by Steve Worswick in 2005, is a well-known conversation chatbot. It adopts the AI Markup Language system, which selects its responses from a knowledge base that identifies known keywords. As another conversation chatbot, Cleverbot has also drawn attention from chatbot research (e.g. Gallacher, Thompson, and Howarth 2018; Shin 2019) because of its impressive ability to hold human-like conversations through its system wherein it responds to users' inquiries by searching for other users' responses in its database. The following excerpt illustrates a conversation with Cleverbot:

Cleverbot: How are you?

User: I am good. How about you?

Cleverbot: I'm fine. Thanks.

User: What are you doing now?

Cleverbot: I'm listening to music.

User: What kind of music are you listening to?

Cleverbot: Panic at the disco.

User: I like Panic at the disco!

Cleverbot: Me too.

According to the mode of communication, chatbots can be classified as text or voice chatbots. The early versions of chatbots, such as Eliza and ALICE, and the previously mentioned conversation chatbots (i.e. Mitsuku and Cleverbot) are examples of text chatbots.

Voice chatbots interact with users via auditory means. As well as the assistant chatbots mentioned above, namely Siri and Cortana, voice chatbots are also implemented in a range of smart AI speakers and 'internet of things' (IoT) systems, with well-known examples including Google Assistant in Google Speakers and Alexa in Amazon Echo. Some of the latest chatbots combine both modes of communication (text and voice), and function in mobile applications such as XiaoIce and Lyra Virtual Assistant. Among these, XiaoIce, developed by Microsoft STCA in 2014, shows remarkable features, such as initiating new conversation topics and identifying users' emotions.

More recently, the field has seen an increasing number of chatbots built for ELT purposes. For example, Mondly is a language-learning application that operates in text and voice modes. For low-proficiency students, Mondly is able to suggest several possible responses to inquiries so that students can simply select the most appropriate answers while communicating with the chatbot. Andy English Bot is another example of this kind. Compared with normal conversation chatbots, these chatbots tend to initiate conversations by raising different questions, which could be particularly effective for low-proficiency learners.

Criteria for evaluating the appropriateness of chatbots for ELT purposes

Most of the chatbots introduced in the previous section have not been built for ELT purposes and thus their appropriateness needs to be evaluated carefully if they are to be used for such purposes. In this section, we provide a selective review of ELT research and offer some evidence-based suggestions for such evaluation.

Coniam's (2014) research is one of the earliest studies in this regard, wherein responses generated from five conversation chatbots were examined in terms of vocabulary, grammar, and meaning fit. In vocabulary, 69–78 per cent of the words generated by the five chatbots were included in the most frequent 2,000 words. The grammatical accuracy of the chatbots was very high, with a range of 77–93 per cent across the five target chatbots. In contrast, the meaning fit was much lower, ranging from 47 to 74 per cent.

Shin (2019) examined the vocabulary included in the utterances of two widely used conversation chatbots, Mitsuku and Cleverbot. To this end, conversations between the chatbots and Korean college students were analysed, focusing on the level of the first 4,000 words in the BNC-COCA 25,000 vocabulary list (Nation 2012). The results showed that, in the case of Mitsuku, over 90 per cent of the vocabulary in her utterances was within the 3,000-word level, and in the case of Cleverbot, about 92 per cent was within the 3,000-word level. In view of this finding, the chatbots' language was judged to be fine-tuned to the level of these EFL learners.

With a similar goal, Kim, Shin, Yang, and Lee (2019) examined vocabulary and readability of the utterances of two voice chatbots, Amazon's Alexa and Google Assistant. The results showed that 97 per cent of the vocabulary of the chatbots was included in the 3,000 words of the Basic English Word List recommended by the Ministry of Education in South Korea. In addition, the readability analysis showed that both chatbots'

utterances were at a difficulty level that is appropriate for children aged between 11 and 13 years (in terms of native speakers of English).

Another approach to evaluating the appropriateness of chatbot language has been suggested by Shum, He, and Li (2018), who counted the number of successful conversation-turns in human–chatbot conversations. In doing so, they employed a measurement called conversation-turns per session (CPS), with a high CPS score indicating a lengthy maintenance of a specific topic in a conversation. By analysing the CPS essentially required for each type of chatbot, Shum et al. proposed that the expected CPS for an assistant chatbot was 1–3, and 10+ for a social (conversation) chatbot.

To summarize, the appropriateness of chatbots for language learning has been examined in terms of vocabulary level, grammatical accuracy, and CPS, with the vocabulary level receiving the most attention. Language teachers may refer to the findings of the previously mentioned studies when selecting a chatbot for their purposes, and evaluate the appropriateness of that chatbot in view of their learners' English proficiency and target activities (see the next section for examples). Nevertheless, the meaning fit of chatbot utterances deserves more attention in future research through in-depth conversation analysis of chatbot–user dialogues.

Applying chatbots

In this section, we will discuss how chatbots can be employed for ELT purposes. In accordance with Chapelle (2001), who first pointed to the need for carefully designed tasks in CALL-based instruction, Kim (2018) proposed three categories of chatbot-based language tasks, which can be performed with different types of chatbot.

First, chatbots can be effectively used for practising speaking and listening skills in the classroom and at home. However, most English learners may not, by themselves, be able to engage in extended conversation with chatbots. Thus, teachers need to provide concrete language tasks with clear directions or worksheets. For example, as small talk activities, learners can be asked to carry out the following tasks: 'Find out at least five things that Kuki (Mistuku) likes (for example, food, movies, sports, etc.) by communicating with her'; 'Talk to Alexa and gather her personal information such as age, address, family, friends, or hobbies'; and 'Express your current feeling or concern and see how Google soothes you.' Talking to these chatbots would also offer excellent listening practice, as some chatbots (e.g. Google) produce a reasonably short input at a normal speed, and others (e.g. Alexa) provide challenging listening materials in lengthier forms (Kim, Shin, Yang, and Lee 2019). We advise that teachers should experiment with the conversational capabilities of a target chatbot, because this would allow relevant dialogue patterns to be identified and thus facilitate the design of speaking and listening tasks for their lessons.

Second, assistant chatbots could effectively lend themselves to reading and writing activities, particularly group-based activities. During reading, students can ask a chatbot for the background information related to the target passage and the meaning of unfamiliar words therein (e.g. 'What is the meaning of V-O-L-C-A-N-O?', 'Where can volunteers work in Nepal?', and 'Who is Georgia O'Keefe?'). In such processes, a chatbot could

serve as an accessible scaffolding tool for intensive reading and schema building. For writing, the chatbot can be employed in 'Listen (read) and summarize' activities, providing a variation on dictogloss (Wajnryb 1990). Students, for example, could be given a task, such as 'Find out former US president Barack Obama's favourite foods and describe what they are in 30 words', for which group members must phrase a relevant question to query the chatbot, ask that question to the chatbot, listen to the chatbot's responses, and produce a written output based on the information obtained from the chatbot.

Third, problem-solving tasks can also be devised for practising integrated language skills. These tasks vary greatly, ranging from finding a place (e.g. 'Recommend a Chinese restaurant near Central Park by investigating with Alexa and explain three reasons'), to matching (e.g. 'Match pictures of pasta shapes below with the corresponding names' by asking Google) and mini-research (e.g. 'Talk with Mitsuku for ten minutes and write down what you have learnt about Queen Elizabeth'), with most of these being examples of content-based instruction (Brinton, Snow, and Wesche 1989). Such tasks should first be designed and piloted by language teachers, with clear target learning objectives in mind. It is expected that the amount of interaction with the chatbot and the required levels of cognitive thinking and strategic planning will correlate with the level of complexity of the target problem-solving tasks.

In this section we introduced some possible ways to integrate chatbots into the ELT curriculum. Although the list is not exhaustive, it points towards some important directions that language teachers need to consider when designing their own chatbot-based language tasks.

Conclusions

This article aimed at familiarizing language teachers with a range of available chatbots and proposing criteria for evaluating the appropriateness of chatbots along with possible ways of applying chatbots for ELT purposes. Before concluding this article, we would like to point out that chatbot research has not kept pace with the fast-growing chatbot-related AI technology, which calls for a greater number of chatbot studies in the ELT research community. These studies may range from ones evaluating the linguistic appropriateness of chatbots for English learners, those proposing how to improve meaning and grammar fits of chatbot utterances, to others comparing the relative effects of different chatbots on language learning, which would elucidate more effective ways of applying chatbots in our English teaching.

Finally, we would like to encourage language teachers to take an interest in open-source chatbot builders, such as Dialogflow and Botsify. Equipped with a little knowledge of programming languages such as Python, PHP, or C++, teachers could design customized chatbots tailored to their students' characteristics and learning objectives. Although some level of studying and training would be required to build such chatbots, we expect that customized chatbots, if constructed properly, could cater to the pedagogical needs of teachers seeking to create optimized chatbot-based language tasks.

Acknowledgement

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A5A2A03037255).

References

Atwell, E. 1999. *The Language Machine: The Impact of Speech and Language Technologies on English Language Teaching*. London: British Council.

Bibauw, S., **T. François**, and **P. Desmet**. 2019. 'Discussing with a computer to practice a foreign language: research synthesis and conceptual framework of dialogue-based CALL'. *Computer Assisted Language Learning* 32/8: 827–77.

Brinton, D. M., **M. A. Snow**, and **M. B. Wesche**. 1989. *Content-Based Second Language Instruction*. Boston, MA: Heinle & Heinle Publishers.

Chapelle, C. A. 2001. *Computer Applications in Second Language Acquisition*. Cambridge: Cambridge University Press.

Coniam, D. 2014. 'The linguistic accuracy of chatbots: usability from an ESL perspective'. *Text & Talk* 34/5: 545–67.

Fryer, L. and **K. Nakao**. 2009. 'Assessing chatbots for EFL learner use' in A. Stoke (ed.). JALT2008 Conference Proceedings. Tokyo: JALT.

Fryer, L. and **R. Carpenter**. 2006. 'Bots as language learning tools'. *Language Learning & Technology* 10/3: 8–14.

Fryer, L., **M. Ainley**, **A. Thompson**, **A. Gibson**, and **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: 461–68.

Gallacher, A., **A. Thompson**, and **M. Howarth**. 2018. '"My robot is an idiot!" –students' perceptions of AI in the L2 classroom' in P. Taalas, J. Jalkanen, L. Bradley and S. Thouësny (eds.). *Future-Proof CALL: Language Learning as Exploration and Encounters—Short Papers from EUROCALL 2018*. Research-publishing.net.

Jia, J. and **M. Ruan**. 2008. 'Use chatbots CSIEC to facilitate the individual learning in English instruction: a case study' in P. Woolf, E. Aïmeur, R. Nkambou and S. Lajoie (eds.). *ITS 2008: Intelligent Tutoring Systems*. Berlin: Springer Verlag.

Kim, H. 2018. 'Designing L2 interactive tasks with an artificial intelligence robot'. Paper presented at the Asia TEFL International Conference, Macau, China.

Kim, H., **D. Shin**, **H. Yang**, and **J. H. Lee**. 2019. 'A study of AI chatbot as an assistant tool for school English curriculum'. *Journal of Learner-Centered Curriculum and Instruction* 19/1: 89–110.

Kim, N.-Y. 2019. 'A study on the use of artificial intelligence chatbots for improving English grammar skills'. *Journal of Digital Convergence* 17/8: 37–46.

Morton, H. and **M. A. Jack**. 2005. 'Scenario-based spoken interaction with virtual agents'. *Computer Assisted Language Learning* 18/3: 171–91.

Nation, I. S. P. 2012. *The BNC-COCA word family lists*. Available at http://www.victoria.ac.nz/lals/about/staff/publications/BNC_COCA_25000.zip (accessed on 17 January 2020).

Shin, D. 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.

Shum, H.-Y., **X. He**, and **D. Li**. 2018. 'From Eliza to XiaoIce: challenges and opportunities with social chatbots'. *Frontiers of Information Technology & Electronic Engineering* 19/1: 10–26.

Swartz, M. L. 1992. 'Issues for tutoring knowledge in foreign language intelligent tutoring systems' in M. L. Swartz and M. Yazdani (eds.). *Intelligent Tutoring Systems for Foreign Language Learning*. Berlin: Springer Verlag.

Wajnryb, R. 1990. *Grammar Dictation*. Oxford: Oxford University Press.

The authors

Lee Jang Ho received his DPhil in education from the University of Oxford, and is presently an associate professor at Chung-Ang University, South Korea. His areas of interest are chatbot-based language learning, the effects of study abroad on language learning, and vocabulary acquisition.

Yang Hyejin received her PhD in applied linguistics and technology from Iowa State University and is currently a full-time researcher at Chung-Ang University, South Korea. Her research interests include language testing and CALL.

Shin Dongkwang received his PhD in applied linguistics from Victoria University of Wellington and is currently an associate professor at Gwangju National University of Education, South Korea. His research interests include corpus linguistics, CALL, and chatbot-based language learning.

Kim Heyoung received her PhD in second and foreign language education from the State University of New York at Buffalo and is currently a professor at Chung-Ang University, South Korea. Her research interests include CALL, digital literacy, and curriculum design. All correspondence regarding this publication should be addressed to her.

Email: englishnet@cau.ac.kr

Appendix

List of the websites related to chatbots and chatbot builders introduced in this article.

- Botsify (developed by Botsify) https://botsify.com/chatbot-for-education
- Cleverbot (developed by R. Carpenter) https://www.cleverbot.com/
- Dialogflow (developed by Google) https://dialogflow.com/
- Mitsuku (developed by S. Worswick) https://www.pandorabots.com/mitsuku/