Assessing chatbots for EFL learner use

Luke Fryer
Kyushu Sangyo University
Kaori Nakao
Seinan Gakuin

For many students in Japan, opportunities to use English are rare. For Japanese tertiary students, studying English once a week, their "eigokaiwa" class may be their only chance to use English. Chatbots have been proposed as a means of providing Foreign Language Learners (FLLs) with English communication practice. Little or no analysis of chatbots' practicality for such practice has been carried out to this point. 36 first year mixed major students were asked to communicate textually with a chatbot on weekly topics. Transcripts of the communication were obtained and qualitatively analyzed to assess the utility of a chatbots as language practice tools for FLLs. The chatbot's common inability to "stay-on-topic" was found to be a salient factor confounding communication (chatbot to student). Students' lexio-grammatical errors were observed to be a salient cause of miscommunication (student to chatbot). Suggestions for the design of future chatbots, and a specifically "FLL chatbot", are outlined.

多くの日本の大学生にとって外国語としての英語を使用する機会は少ないといえる。日本の大学生にとって週1回の"英語会話"の授業は彼らの言語学習の唯一の機会である。Chatbotは、外国語学習者（以下、FLLs）の言語学習の練習手段として提案された。Chatbotの効果についてはこれまで研究がなされてこなかった。不特定の学部1年生36名を対象に、毎週決められたトピックについてChatbotと一緒にテクストによる会話をするよう課題を与えた。FLLsにとって言語練習のツールとしてChatbotが実用的であるかどうか調べるためテクストの写しを得て質的分析を行った。その結果、Chatbotは一つのトピックについて会話することが困難であり、大きな課題となった（Chatbotから学生に対して）。学生の文法と語彙の間違いもコミュニケーションの大きな問題となっていた（学生からChatbotに対して）。未来のChatbotのデザインとFLL Chatbotのデザインの概要を述べる。

Foreign Language Learning contexts, by their nature, are often lacking in opportunities for students to use their language of study. Language use is fundamental for language learner development (Chamot, 2005). In Japanese tertiary compulsory English learning courses, where classes are often only once a week, the need for practice and thus concretization of studied language is acute. Although opportunities for textual (readers) and aural input (DVDs and CDs) are easily obtained, due to Japan's
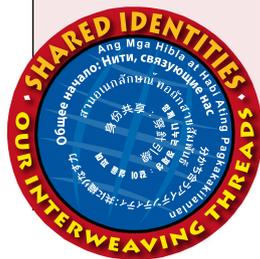
demographics (more than 99% ethnically Japanese) (World-Factbook, September 2008), communication "practice" may not be commonly available to many students.

Fryer and Carpenter (2006) identified "chatterbots" as a potential tool for providing students in Japanese tertiary English learning contexts with "language practice". Chatterbots, or chatbots, are, "a type of conversational agent, a computer program designed to simulate an intelligent conversation with one or more human users, via auditory or textual methods." (Wikipedia, 2008). Chatbots are a modern day instantiation of Turing's (1950) timeless question "Can machines think?". The programming of chatbots began with Wizenbaum's program "Eliza", which was designed around the concept of a common type of psychoanalysis used in the 1960s (Weizenbaum, 1966) . The program was simple and used a text interface to ask a user about whatever information they provided about themselves. Despite its basic nature, it still proved to fool a few users into believing that it was not software, but instead a real person. The internet, increasingly sophisticated hardware and new computing languages have meant that in the four decades since Eliza's creation, chatbots have become common on the internet. In a case study with Japanese tertiary studying English, Fryer and Carpenter (2006) found students generally perceived interacting with the chatbot to be of low risk to their self-confidence and often interesting. Research discussing the use of chatbot in the EFL classroom, has suggested that chatbots designed for native speakers, may not be ideally suited to language learning students needs (Coniam, 2008a, 2008b; Fryer & Carpenter, 2006). Unfortunately, unchanged from the time of Fryer and Carpenter's 2006 publication, a

chatbot designed to meet foreign language learners' (FLLs) needs, has yet to be constructed. The logical first step before beginning such a construction, however, would be to assess in what ways present chatbot technology might be insufficient for FLLs. As hundreds of different chatbots are presently available online (see Coniam, 2008a for a sample of well-known chatbots and a brief analysis of their abilities) , it was determined that rather than trying a number of different chatbots and comparing them, it would instead be more meaningful to analyze the results of FLL students' interaction with one Chatbot on a variety of topics. The need to obtain the transcripts of the student-chatbot interaction, limited the potential number of chatbots that could be used. Finally, the most recent winner of the Loebner prize annual medal (Anonymous, 2008), which also makes transcripts available, was chosen as the chatbot to be used: Jabberwacky. The Loebner prize is an award given out annually to the chatbot within the Lobner competition, which is perceived to be the most intelligent and human-like (Anonymous, 2008); this is assessed by an adapted format of the Turning Test (Turing, 1950). Jabberwacky, in addition to meeting the outlined needs of this research project, has also recently been judged to have the most potential for ESL learner use (Coniam, 2008b).

The research questions for this paper are: 1) What aspects of the chatbot (Jabberwacky) are presenting obstacles to its efficacious use as an English language communication review for FLLs? 2) What obstacles to human-chatbot communication are low-level Japanese tertiary students causing? 3) What strategies might these learners use to ensure they maximize the quality of their interaction with

chatbots presently available online.

Methods

During semester two of a year-long Eigokaiwa course at a large private university in Kyushu, students in two first year mixed major English communication classes were asked to textually converse with the chatbot (Jabberwacky) once a week. Students (n=36) within the classes sampled had a TOEIC score of between 360 and 400. Each week, the students were all given the same topic—a communication topic covered in class during the first semester of the course—to communicate with the chatbot upon. The students were asked to complete a minimum of twenty "A-B" couplets of interaction with the chatbot, then "cut/paste" and email the transcript of the conversation to the researcher. The transcripts were moved to one excel file and initially cross-referenced by student and topic. The texts were then coded into successful and unsuccessful communication categories. These two categories then underwent further coding, to better understand the salient reasons for communication failure, within student and topic. This analysis involved looking for regularly occurring types of miscommunications (chatbot to student, and student to chatbot separately) within the text, then developing appropriate categories and finally coding. In addition, the chatbot's and the students' text was analyzed for vocabulary, using Lextutor's (Cobb, 2009) vocabulary profile software to assess the difficulty of vocabulary in use.

The combination of both assessing the vocabulary in

use by students and the chatbot, as well as the nature of the miscommunications commonly occurring, may yield suggestions for how chatbots might be developed in the future. In addition, this analysis may clarify whether and to what extent chatbots are practically useful for FLL students now.

Results

Vocabulary

Table 1. Chatbot vocabulary profile

K1 Words (1-1000)*	88.35%
K2 Words (1001-2000)*	2.42%
1k+2k	90.77%
AWL Words (academic)	1.21%
Off-List Words	8.03%

*General Service Wordlist (West, 1953)

The chatbot's vocabulary contained a large percentage of K1 (most frequently used 1000 words of English) words (88.35 %), suggesting that at least in terms of vocabulary, the chatbot was easy to comprehend. The off-list words (8.03%), rather than being made of low-frequency vocabulary, mostly consisted of city and country names (.64%), and spelling mistakes (6.01%).

Table 2. Student vocabulary profile

K1 Words (1-1000)*	86.41%
K2 Words (1001-2000)*	1.70%
1k+2k	88.11%
AWL Words (academic)	1.04%
Off-List Words	10.85%

*General Service Wordlist (West, 1953)

The students' vocabulary also contained a large percentage of K1 words (86.41%), which was consistent with their generally low language proficiency. The students' off-list words (10.85%) were almost entirely spelling mistakes (9.11%) and a few names (celebrities, places).

Chatbot-student communication results

Analysis of student-chatbot conversations shows that the chatbot has a short memory

> Chatbot: Who's your favorite actor?
>
> Student A: Johny Dep.
>
> Chatbot: Johny?
>
> Student A: Yes. His masterpiece is Pirate Of The Cribbean.
>
> Chatbot: I don't know.
>
> Student A: He is very greate actor.

> Chatbot: Ok, do you like Oscar?
>
> Student AA: Who is Oscar?
>
> Chatbot: I don't know who is he?

*Problems in italics

The above transcript displays a problem students commonly have when communicating with chatbots: chatbots have very short memories. In the example above, the chatbot asks if the student knows "Oscar"; however, when the student asks who "Oscar" is—in the next line—the chatbot doesn't know who Oscar is.

Analysis of student-chatbot conversations shows that the topic of conversation is too difficult for the chatbot

> student B: Let's talk about character!
>
> Chatbot: No, let's talk about politics.
>
> Student B: *I think that I have a character hating to lose.*
>
> Chatbot: *Everyone has a bad period from time to time.*
>
> Student B: *And I am very careless. Are you?*
>
> Chatbot: *No.*
>
> Student B: Good. *There are many careless people in my friend.*
>
> Chatbot: *There are many problems too.*
>
> Student B: *But, everybody gets along well very much.*

Chatbot: Hmm. *So I gather that you cannot identify your actual age. Yeah?*

Student B: *I think so too. Therefore I work hard to get over this character now.*

Chatbot: *No way... I would not have make humans nearly as picky and superficial as they are.*

*Problem in italics

Two problems seem to be causing miscommunication in this transcript: the unusual and often incoherent phrasing on the part of the student is playing a large role and the fact that the chatbot is having difficulty with the topic's content. Similar to the previous transcript, the chatbot's inability to grasp anything beyond the most recent question or answer is preventing it from communicating with the student effectively. In the transcript above, it is hypothesized that the ethereal nature of the topic "one's character" is playing a role in the difficulties observed.

Spelling mistakes were expected to be a large barrier to communication between student and chatbot, due to the number of words they both misspelled.

Analysis of student-chatbot conversations shows that the chatbot can overcome student spelling errors

Student C: I like avrillavigne.

Chatbot: I know her. My favorite song is ' I'm with you'.

Student C: My favorite song is 'Innocence'.

This short transcript shows however, that the chatbot can overcome small misspellings on the part of the student. Also,

the vocabulary analysis suggested that both students and the chatbot generally used relatively simple vocabulary that they can generally spell. Spelling errors were more often typing mistakes such as the one in the previous transcript, rather than actual misspellings.

It may simply have been as a result of students' low language proficiency; however, the majority of the communication problems were found to be due to student lexio-grammatical errors.

Analysis of student-chatbot conversations shows that student lexio-grammatical errors can cause miscommuniction

Student D: Because I live in Japan, I do not occur to the thing called much this.

Chatbot: I knew this. Where exactly?

Student D: Do you think that Japanese is kind?

Chatbot: No, because I have a huge one, and it hasn't gotten me anywhere has it?

Persistent students, who despite miscommunications kept asking questions, e.g.:

Analysis of student-chatbot conversations shows that student persistence plays an important role in successful communication

Student D: Do you think that Japan is safe?

Chatbot: I don't know. It was your idea so you tell me.

were eventually rewarded with communicative interaction with the chatbot. In some cases, even student questions that weren't well formed received coherent responses, as the

chatbot displayed it could overcome simple miss-phrasings:

Analysis of student-chatbot conversations shows that chatbot can overcome students' miss-phrasing

> Student D: Are you interested about any Japan?
>
> Chatbot: Very, I think the culture is unique and fascinating and it is measurably the most rational/ secular.

Finally, although miscommunications were common and often frustrating for many of the students participating, the transcripts also contained some occurrences of clear and connected student-chatbot communication:

Analysis of student-chatbot conversations shows that occasionally clear, connected student-chatbot communication occurs

> Student E: It's difficult words for me. I don't like person who often tells lie.
>
> Chatbot: I don't like you too.
>
> Student E: Really? Do you have girlfriend?
>
> Chatbot: No, do you?
>
> Student E: Yes. My boyfriend is very gentleman!
>
> Chatbot: Your boyfriend?? You have a boyfriend!?
>
> Student E: Yes, my boy friend. Are you surprise?
>
> Chatbot: Is he handsome??
>
> Student E: So so. But his character is very good!

> Chatbot: That didn't answer my question.

The chatbot seems to maintain a sense of the conversation and even manages to push the student to accurately answer its question.

Future directions

Although some directions for future chatbot construction have been discussed, it is important for these to be reaffirmed here, in the hopes that interested programmers find them helpful for future chatbot design.

Present chatbots' usefulness to FLLs could be significantly improved by:

1. Programming them to persistently remain on one topic of conversation. Strategies for making this possible might be: a) seeking help for the user when the chatbot does not understand, rather than changing the topic as a defense; b) suggesting topics to the user that the chatbot can coherently talk about; c) having the chatbot take the role of "questioner" on topics that it is less familiar with; d) having the chatbot ask to change the topic, rather then suddenly shifting.

2. By including a spell-check and grammar check function for the user's input. The spell-check would largely function to catch student typing errors. The grammar check would be useful, if only to point out to the student that a grammatical error might have been made and encourage rephrasing.

3. Programming them to promote conversation that is more complicated than simple A-B; A-B. This

suggestion is connected to suggestion number one but goes further than simply "staying on topic", it means that chatbots "memories'" need to be enhanced and their general knowledge of how language is connected would need to be increased.

If the programming necessary to achieve these goals is not possible at present, it may be possible to create a chatbot interface that provides a means of limiting the student-chatbot conversations to topics and language that they are both capable of competently understanding and producing.

If a chatbot were to be created specifically for FLLs, then its programming should include an ability to help students with vocabulary and grammar. Such chatbot support could be either direct, through grammatical and vocabulary instruction when miscommunications arose, or indirect, through scaffolding questions, such as "don't you mean (vocabulary or phrase)" or simply "sorry, I didn't understand that, could you try that again."

The potential for chatbots for FLLs appear deep and as yet largely ignored by both the programmers building them and most of the language learning profession. Chatbots are bound to become useful, perhaps indispensible tools to language learners, it is just a matter of when.

Conclusion

From the perspective of the students' communication, grammar was a common limitation to the quality of communication found; less commonly, misspellings cause some communication problems. Both the chatbot and the students were found to be using high frequency vocabulary, and therefore discrete word understanding was rarely a problem.

During in-class group discussions about using the chatbot (weekly 10 minute group discussion, with the teacher moving amongst them, answering questions and hearing what they had to say—conversations recorded but not yet transcribed for analysis), students discussed using: roll-over online dictionaries, Microsoft WORD's spell and grammar check function and then cut/pasting phrases into the webpage. Students also suggested that easy, concrete topics, which had been recently covered in class, decreased the incidence of student errors. Many students also cited just being persistent and asking lots of questions, as an easy way to keep the conversation going.

The bots inability to stay on topic, and occasionally nonexistent memory of previous comments and questions, was found to be particularly discouraging of good chatbot-student communication. Presently, as the students suggested, having the students stick to one topic and playing the role of questioner seemed to be the best strategy for overcoming this impediment. Future chatbots, especially those interacting with speakers in their second language, need to be able to remember what they have just said, so they can explain what they meant and develop conversations beyond A-B, A-B, B-A. If it is too difficult to create chatbots that can stay on topic, it may be possible to simply have a "drag-down" menu of conversation topics for the student to choose from.

Presently, chatbots provide a potential low-risk and virtually shame-free communication environment for FLL students of English. In support of Fryer and Carpenter's (2006) results, a six item open ended survey (completed

in Japanese) at the end of the study (n=34 received and translated) suggested that all of the students found the communication with the chatbot unintimidating, if occasionally frustrating. From the perspective of the teacher, it gave the students an opportunity to independently review communication-based learning in a way that textbooks, CDs, or any reading/assignments could not achieve. So, it was found to be useful for risk-free, externally-regulated, class communication-task review. Does this make it a worthwhile tool for student independent study, considering the communication problems students encountered? The answer depends on the teacher, the students and the learning goals of the course. In their present form, it is safe to say that chatbots, even with all of their limitations, fill a language learning tool niche, which nothing else presently can.

Pedagogical Implications

At this stage, the usefulness of Chatbots are limited by their weak language skills and often primitive interface (Coniam, 2008a). However, the limited results presented in this research report suggest that for students encouraged to be persistent in their questioning and possessing tolerance for ambiguity, they are a useful means for practicing new language in FLL contexts.

Luke Fryer is a lecturer at Kyushu Sangyo University. <fryer@ip.kyusan-u.ac.jp>

Kaori Nakao is a graduate student as Seinan Gakuen.<kaorinewyork@hotmail.com>

References

Anonymous (2008). Home page of the Loebner prize in artificial intelligence. [Online] Available: http://www.loebner.net/Prizef/loebner-prize.html

Chamot, A. U. (2005). Language learning strategy instruction: current issues and research. *Annual Review of Applied Linguistics, 25*(-1), 112-130.

Cobb, T. (2009). Lextutor. [Online] Available: http://www.lextutor.ca/vp/

Coniam, D. (2008a). An evaluation of chatbots as software aids to learning English as a second language. *The Eurocall Review, 13*. Retrieved from http://www.eurocall-language.org.

Coniam, D. (2008b). Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL, 20*(01), 98-116.

Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language Learning and Technology, 10*(3), 8-14.

Turing, A. M. (1950). Computer machinery and intelligence. Mind, 59, 433-460. [Online] Available: http://loebner.net/Prizef/TuringArticle.html

Unknown (2008). Home page of the Loebner prize in artificial intelligence. [Online] Available: h ttp://www.loebner.net/Prizef/loebner-prize.html

Weizenbaum, J. (1966). ELIZA--A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1). [Online] Available: http://i5.nyu.edu/~mm64/ x52.9265/january1966.html

West, M. (1953). A general service list of English words.
  London: Longman.

World-Factbook (September 2008). Japan—People. [Online]
  Available: https://www.cia.gov/libr ary/publications/the-
  world-factbook/geos/ja.html#People