Conversational Agents in Language Education: Where They Fit and Their Research Challenges

Rahul R. Divekar, Haley Lepp, Pravin Chopade, Aaron Albin, Daniel Brenner and Vikram Ramanarayanan

Educational Testing Service, USA
{rdivekar, hlepp, pchopade, aalbin, dsbrenner, vramanarayanan}@ets.org

Abstract. Conversational agents (or dialogue agents or Artificial Intelligent AI agents or chatbots) can provide a foreign language learner with otherwise hard-to-find conversational exposure to a new language. Such agents that teach languages differ significantly from their general-purpose counterparts in their goals, their approach, and their users' characteristics, thereby effectively creating a new interaction paradigm for which little literature exists in the HCI of Conversational Interfaces community. The difference from general-purpose agents comes from two themes highlighted in this work: the user is not expected to know the language of interaction, and the purpose of the conversation is language education through task completion rather than task completion itself. This paper highlights the role and the research challenges of interactions with dialogue agents that allow people to learn and practice new languages.

Keywords: Spoken Dialogue Systems · Language Learning · Conversational Agents · Conversational User Interfaces · Immersion · Multimodal Interfaces

1 Introduction and literature

1.1 Introduction

Foreign language learners often lack access to native speaking partners. Dialogue agents within TOEFL MOOC, Duolingo, Mondly, Cleverbot, and Busuu, [11, 27, 31, 32] address this gap by mimicking conversational interactions with a human [35]. However, the HCI community has not given much attention to language-teaching agents which come with unique challenges.

The uniqueness arises from two broad differences between ordinary agents and language teaching agents: user characteristics and conversation goals. Ordinarily, agent designers can assume that the user is proficient in the language of interaction, whereas language learners are seeking to build that proficiency. Secondly, while most agents are designed to assist a user with completing a task in the easiest possible way (e.g., booking a flight), in a language learning application task-completion is mostly a means to the end of language acquisition. The difference in goals of the conversation along with differences in user characteristics create a new paradigm of interaction and consequently new research challenges; as detailed in this paper.

1.2 Literature

Language learning requires repeated conversational practice [16]. Dialogue agents can facilitate this and expose students to natural language, culture, and pragmatics of communication by playing different interaction roles. [24] show a Question-Answer (QA) based teaching agent, [28]'s chatbot can make small talk, converse in a QA format, and give hints to a learner. [34]'s agents role play scenarios and give feedback. [35] use conversational agents to tell a story and ask follow-up questions to the learner; it adjusts questions based on a learner's ability to answer, and focuses on encouragement in dialogue. [17] presented the use of conversational agents in e-learning environments with examples for supporting self-assessment. [7]'s agents in an Extended Reality (XR) street market allow role-play opportunities to practice conversation and colloquial negotiation skills. [2, 27] have done a systematic review. From above, we derive that designers use a combination of text or speech/audio as the interaction medium and situate their agents on platforms ranging from smart speakers to high-fidelity extended reality virtual environments in order to teach various aspects of language such as grammar, vocabulary, culture, pragmatics, interactive dialogue, etc.

2 Research Challenges of Language-Teaching Dialogue Agents

[29] highlight the need to move away from transactional request-response-like interactions to a colloquial conversation with a user that demonstrates pragmatics, personality, turn-taking, and enthusiastic verbal and non-verbal conversational overlap without the need of an invasive wake-up word. This is in line with creating social presence and rapport [3]. The importance of these complex research challenges apply to all agents and cannot be overstated in the case of language education. However, in this text we contextualize broad challenges to language-education agents and articulate new ones.

We do so from four perspectives seen in fig. 1 which map to subsections below. The intersectional perspectives in the figure delineate the bases of situated learning [4], a widely-recognized theory in foreign language acquisition that promotes learning by doing.

2.1 Language learning, teaching, and assessment

Well established learning theories like situated learning and task-based language teaching [20] advocate for learning by doing. For language acquisition, this means not only learning vocabulary and grammar but also applying them to communicate in real-life scenarios. Learners are best able to conceptualize the target language by being visually, culturally, and linguistically immersed in it, i.e. participating in interactions with context and feedback as opposed to only learning translations [20]. Such opportunities are scarce for most language learners due to a lack of fluent conversation partners and appropriate situations in which to practice. Dialogue agents can bring natural conversational opportunities. Combined

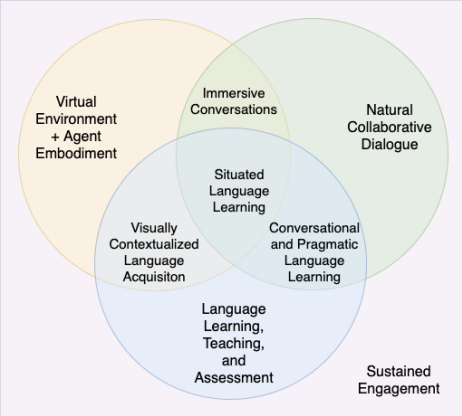

Fig. 1. Illustration of the domain space for an effective language learning solution

with virtual environments and educational elements like targeted feedback, situated language learning can be enabled as seen in fig. 1. Details of their challenges are discussed next.

2.2 Virtual Environments (VEs) with Embodied Agents

Contextual Learning in VEs: Creating a visual context for a educational dialogue agent has many advantages: it sets up the conversational and learning expectations, enhances social and cultural immersion, and increases the chances of incidental learning among others. A wide range of platforms differing in levels of immersiveness and presence (i.e. feeling of "being there") are available in which VEs can be implemented varying in their fidelity, availability, costs, ubiquity, etc. Each medium has different effect on different subsets of populations [15]. [26] have done a systematic review of VEs in language education. We observe that few studies have explored VEs with dialogue agents.

Embodiment: In addition to increasing presence, agents that use human-like gestures are more effective in teaching [23]. An embodied dialogue agent can produce gestures and expressions consistent with the target culture and give students an exposure to appropriate non-verbal communication. Further, they can ease communication by letting students make sense of the conversation through multimodal cues and mouth movements [5].

Creating a holistic cultural-rich visual environment with dialogue agents is important as learning communication without cultural pragmatics sets students up to appear as "fluent fools" [1]. Further, VEs can teach students how to bring visual, contextual information into their conversations. For example, a role-play of ordering food in a virtual restaurant would teach a learner how to use a menu, which is present outside the spoken context and is only present in the visual context. We will explore dialogue agents next.

2.3 Natural Collaborative Dialogue

While VEs can create virtual presence (i.e. feeling of "being there"), dialogue agents can add social presence (i.e. feeling of talking to an "intelligent entity" [3]) thereby completing the picture of an immersive virtual world [6]. This section focuses on role of agents in language education with and without VEs.

Naturalness in dialogue: Dialogue agents can enable realistic interactions in a VE, improving spoken proficiency and willingness to communicate in a foreign language [18]. However, [27] note that some language-education agents' dialogues have a scripted quality. Probabilistic methods can counter this perception but afford less control over teaching materials and responses.

In a speech-enabled dialogue, synthetic voice output can create the impression of artificialness. [23] shows that a human voice is preferable and has a better effect on learning than synthetic voices. Being able to easily synthesize speech that sounds natural is an existing research challenge, amplified in a code-switched language (native and target language) environment. However, the artificialness may lower the stakes of the interaction and lead to more comfortable learning [15, 7].

Role of Dialogue Agents: Agents can play three roles: an interaction medium itself used to navigate through the environment, instructor (e.g., teaching language constructs and giving targeted feedback), or a roleplay partner (e.g., a server in a virtual restaurant).

Conversation exercises can either be focused on a specific construct or general exposure to conversation. E.g., [9]'s game allows learners to play a conversational guessing game where they are shown several avatars, one of which is the "target avatar". The learners can ask yes/no questions to the agent to guess the target avatar. Meanwhile, [7]'s agent provides scenario role-play where a student may learn the language, culture, and pragmatics of shopping at a street market.

The interaction with the agent can range from *exploratory* to *guided*. The former relies on students' initiation of dialogue and provides more autonomy while the latter relies on agent-initiated conversations guiding the student through a conversational experience. Each student may benefit from a different level of autonomy, targeted feedback, and scaffolding in their learning journey.

Speech Recognition: Dialogue agents in language learning tend to be text-based partially because building automatic speech recognition (ASR) systems for language learners is a hard task. Compared to regular speakers, learners make more mistakes, code-switch with their native language, thus requiring more complex modeling with large data sets [30].

Natural Language Understanding (NLU): [25] show that even though ASRs have a high word-error rate with non-native speech, non-native speakers can still successfully communicate with agents; likely because of the NLU's probabilistic intent mapping. Probabilistic NLUs rely on intelligibility rather than accuracy. For example, "I want to go *at* school" is inaccurate but intelligible and would likely map to the same intent as the accurate phrase. The focus of NLUs on intelligibility provides a smooth interaction in ordinary conversations. However, in language learning, such inaccuracies must not be ignored, and errors

must be brought to attention. This calls for a different approach to NLUs that can separate accurate from inaccurate speech yet move the dialogue forward.

Multiple agents and users: Exposure to non-dyadic conversations furthers students' glimpse into the target language and culture. Further, learning with others or cooperative learning is an effective language learning mechanism. An experience where students can learn together can give a sense of accomplishment, encourage students further, and add richness to the interactions. An interesting concept would be a learning space where human teachers, AI agents, and humans cooperatively learn together [7].

Multimodal understanding: Non-verbal cues can help interpret a user's affective state to which the dialogue agent can respond empathetically to nurture a learning progression [14]. Further, it can help communication move forward as it is common for learners to substitute unknown words or elaborate abstract meanings with actions/gestures [8]. Recognizing gestures can allow the dialogue agent to administer Total Physical Response (TPR) exercises to pair new information along with a physical response and improve retention [33].

Repair and Grounding: Foundational research [21] has explored how learners converse with native speakers and learn from them in linguistically immersive settings. Native and non-native speakers modify their language and use a variety of strategies and tactics to communicate with each other such as repair, grounding, and negotiating meaning; all formidable challenges in HCI. Few agents today could cogently respond to a request like "What do you mean?". It remains a research task to identify what kinds of repair and negotiation strategies students use while talking to AI agents while learning a language and how agents can respond. [8] show simple ways like pre-defined translations, explanations, and hints as a start but much of the work in this direction remains.

Personalization in Dialogue: Scaffolding helps learners build new knowledge and skills. A conversational agent thus could identify a learner's language level and provide additional dialogue-based tasks for a student to learn. Such dialogue-based tasks could be a new exercise or the same dialogue but a different parse of the dialogue tree. The agent would have to personalize the amount of user autonomy, generate appropriate hints, etc.

Dialogue-based assessment and feedback: Identifying errors made while speaking at many linguistic levels (e.g., phoneme, syllable, vocabulary, grammar, semantic, dialogue) and then communicating it effectively and encouragingly to the learner in an actionable way while having a dialogue is another hard research challenge from NLP/HCI perspective.

Surprising learning patterns with agents and who-knews await discovery. For example, [8] showed that users in a dialogue with embodied agents found it helpful to see something as simple as live transcriptions of their speech as it ended up being a proxy for pronunciation feedback that signaled errors.

2.4 Sustained engagement

As opposed to ordinary chatbots that help humans with a relatively short task, language learning is a prolonged cognitive effort that spans over months and

years thus needing agents to sustain learners' engagement. Solutions commonly used in Computer-assisted language learning to bring engagement are gamification [10], serious games, immersion, humor, etc [27]. It remains a research question to find how these elements can be executed in dialogue-based language learning and in a longitudinal manner [12].

2.5 Accessibility

Finally, two missing aspects in designing and evaluating agents are accessibility and special needs. [19] discuss how conversational interactions can be designed for the visually impaired, [13] show conversational interactions for deaf and hard of hearing users. In the domain of education, [22] demonstrate using embodied conversational agents to support spoken language development in children with hearing loss by making use of the visual modality. We note that literature in accessibility for conversational interfaces is limited, which provides a further research opportunity to address this gap.

3 Conclusion

The language learning community has rarely seen applications where all the above perspectives have come together. The technical challenges to make it happen are significant and acknowledged but rarely in the context of education which provides a new motivation. We have articulated the motivation for dialogue interfaces that can create unique opportunities for language learners by bringing to them AI-enabled situated learning. We have articulated the progress in the field and highlighted the many inter-disciplinary research challenges of HCI of foreign/second/other language education. New interaction paradigms are waiting to be discovered in the intersection of multiple users, multiple agents, communicating with multiple modalities, and learning together in virtual worlds.

Bibliography

[1] Bennett, M.J.: How not to be a fluent fool: Understanding the cultural dimension of language. New ways in teaching culture pp. 16–21 (1997)

[2] Bibauw, S., François, T., Desmet, P.: Discussing with a computer to practice a foreign language: research synthesis and conceptual framework of dialogue-based call. Computer Assisted Language Learning **32**(8), 827–877 (2019)

[3] Biocca, F., Harms, C.: Defining and measuring social presence: Contribution to the networked minds theory and measure. Proceedings of PRESENCE **2002**, 1–36 (2002)

[4] Brown, J.S., Collins, A., Duguid, P.: Situated cognition and the culture of learning. Educational researcher **18**(1), 32–42 (1989)

[5] Burnham, D., Lau, S.: The integration of auditory and visual speech information with foreign speakers: The role of expectancy. In: AVSP'99-International Conference on Auditory-Visual Speech Processing (1999)
[6] Divekar, R.R.: AI Enabled Foreign Language Immersion: Technology and Method to Acquire Foreign Languages with AI in Immersive Virtual Worlds. Ph.D. thesis, Rensselaer Polytechnic Institute (2020)
[7] Divekar, R.R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., Zhu, H., Hendler, J.A., Braasch, J.: Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. Computer Assisted Language Learning pp. 1–29 (2021)
[8] Divekar, R.R., Drozdal, J., Zhou, Y., Song, Z., Allen, D., Rouhani, R., Zhao, R., Zheng, S., Balagyozyan, L., Su, H.: Interaction challenges in ai equipped environments built to teach foreign languages through dialogue and task-completion. In: Proceedings of the 2018 Designing Interactive Systems Conference, pp. 597–609 (2018)
[9] Evanini, K., Timpe-Laughlin, V., Tsuprun, E., Blood, I., Lee, J., Bruno, J.V., Ramanarayanan, V., Lange, P.L., Suendermann-Oeft, D.: Game-based spoken dialog language learning applications for young students. In: INTERSPEECH, pp. 548–549 (2018)
[10] Flores, J.F.F.: Using gamification to enhance second language learning. Digital Education Review (27), 32–54 (2015)
[11] Fryer, L., Coniam, D., Carpenter, R., Lăpușneanu, D.: Bots for language learning now: Current and future directions (2020)
[12] Fryer, L.K., Ainley, M., Thompson, A., Gibson, A., Sherlock, Z.: Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. Computers in Human Behavior **75**, 461–468 (2017)
[13] Glasser, A., Mande, V., Huenerfauth, M.: Accessibility for deaf and hard of hearing users: Sign language conversational user interfaces. In: Proceedings of the 2nd Conference on Conversational User Interfaces, pp. 1–3 (2020)
[14] Grafsgaard, J., Wiggins, J., Boyer, K.E., Wiebe, E., Lester, J.: Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue. In: Educational Data Mining 2014 (2014)
[15] Hsu, L.: To call or not to call: empirical evidence from neuroscience. Computer Assisted Language Learning pp. 1–24 (2020)
[16] Ismail, J.: Language exposure and second language learning. The English Teacher p. 11 (2017)
[17] Kerry, A., Ellis, R., Bull, S.: Conversational agents in e-learning. In: Allen, T., Ellis, R., Petridis, M. (eds.) Applications and Innovations in Intelligent Systems XVI, pp. 169–182, Springer London, London (2009), ISBN 978-1-84882-215-3
[18] Lee, J.S., Lee, K.: Affective factors, virtual intercultural experiences, and l2 willingness to communicate in in-class, out-of-class, and digital settings. Language Teaching Research **24**(6), 813–833 (2020)
[19] Loddo, I., Martini, D.: The cocktail party effect. an inclusive vision of conversational interactions. The Design Journal **20**(sup1), S4076–S4086 (2017)

[20] Long, M.: Second language acquisition and task-based language teaching. John Wiley & Sons (2014)
[21] Long, M.H.: Native speaker/non-native speaker conversation and the negotiation of comprehensible input1. Applied linguistics **4**(2), 126–141 (1983)
[22] Massaro, D., Liu, Y., Chen, T., Perfetti, C.: A multilingual embodied conversational agent for tutoring speech and language learning. vol. 2 (01 2006)
[23] Mayer, R.E.: Using multimedia for e-learning. Journal of Computer Assisted Learning **33**(5), 403–423 (2017)
[24] Molnár, G., Szüts, Z.: The role of chatbots in formal education. In: 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), pp. 000197–000202, IEEE (2018)
[25] Moussalli, S., Cardoso, W.: Intelligent personal assistants: can they understand and be understood by accented l2 learners? Computer Assisted Language Learning **33**(8), 865–890 (2020)
[26] Peixoto, B., Pinto, R., Melo, M., Cabral, L., Bessa, M.: Immersive virtual reality for foreign language education: A prisma systematic review. IEEE Access pp. 1–1 (2021)
[27] Petrovic, J., Jovanovic, M.: Conversational agents for learning foreign languages–a survey. arXiv preprint arXiv:2011.07901 (2020)
[28] Pham, X.L., Pham, T., Nguyen, Q.M., Nguyen, T.H., Cao, T.T.H.: Chatbot as an intelligent personal assistant for mobile language learning. In: Proceedings of the 2018 2nd International Conference on Education and E-Learning, pp. 16–21 (2018)
[29] Pinhanez, C.S.: Hci research challenges for the next generation of conversational systems. In: Proceedings of the 2nd Conference on Conversational User Interfaces, pp. 1–4 (2020)
[30] Qian, Y., Ubale, R., Lange, P., Evanini, K., Ramanarayanan, V., Soong, F.K.: Spoken language understanding of human-machine conversations for language learning applications. Journal of Signal Processing Systems pp. 1–13 (2019)
[31] Ramanarayanan, V., Pautler, D., Lange, P.L., Tsuprun, E., Ubale, R., Evanini, K., Suendermann-Oeft, D.: Toward scalable dialog technology for conversational language learning: Case study of the toefl® mooc. In: INTERSPEECH, pp. 1960–1961 (2018)
[32] Rosell-Aguilar, F.: Autonomous language learning through a mobile application: a user evaluation of the busuu app. Computer Assisted Language Learning **31**(8), 854–881 (2018)
[33] Si, M.: A virtual space for children to meet and practice chinese. International Journal of Artificial Intelligence in Education **25**(2), 271–290 (2015)
[34] Wik, P., Hjalmarsson, A.: Embodied conversational agents in computer assisted language learning. Speech Communication **51**(10), 1024–1037 (2009), ISSN 0167-6393, spoken Language Technology for Education
[35] Xu, Y., Wang, D., Collins, P., Lee, H., Warschauer, M.: Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. Computers & Education **161**, 104059 (2021)