Bilingual Robots to Explore Cognitive Effects of Bilingualism thru Human-Robot Interaction

Casey C. Bennett†
Department of Intelligence Computing
Hanyang University
Seoul Korea
cabennet@hanyang.ac.kr

Say Young Kim
Department of English Language & Literature
Hanyang University
Seoul Korea
sayyoungkim@hanyang.ac.kr

Benjamin Weiss
Quality and Usability Lab
Technische Universität Berlin
Berlin Germany
benjamin.weiss@tu-berlin.de

ABSTRACT

Research on bilingualism over the past 50 years has revealed that learning a second language causes significant cognitive shifts in human brains, which affect how individuals express themselves during social interaction. There is tremendous potential for the human-robot interaction (HRI) community to contribute to that understanding. By building robots and virtual avatars that can communicate in multiple languages using purposely designed parallel speech systems, we can create replicable yet modifiable experimental tools to explore questions around the influence of human language on social communication. Doing so will allow us to overcome the limitations of using human confederates or other experimental paradigms in linguistics where replicability can be challenging. Moreover, such research can contribute back to the field of HRI, by allowing us to design robots in the future that can better communicate with human users across cultures and language barriers, as well as perhaps even develop novel forms of communication between humans and robots.

CCS CONCEPTS

Human-centered computing → Human computer interaction (HCI), **Computing Methodologies** → ARTIFICIAL INTELLIGENCE → Natural Language Processing, **Computer systems Organization** → Embedded and cyber-physical systems → Robotics

KEYWORDS

Human-Robot Interaction, Social Cognition, Bilingualism, Virtual Avatar, Linguistic Relativity, Experimental Design

ACM Reference format:

Casey C. Bennett, Say Young Kim, and Benjamin Weiss. 2023. Bilingual robots to explore cognitive effects of bilingualism thru human-robot interaction**.** *CUI@HRI Workshop at the ACM/IEEE Conference on Human-Robot Interaction (HRI'23).* ACM, Stockholm, Sweden, 4 pages.

†Corresponding Author

HRI '23 CUI@HRI Workshop, March 13–16, 2023, Stockholm, Sweden

1 Introduction

What does it mean to be bilingual, and what can the ways in which bilingual speakers converse in different languages tell us about human social cognition [1,2]? Those are important questions, especially for those of us who study how to create better, more life-like interactive technology [3]. This position paper contends that there is a significant role for robots, and in particular human-robot interaction (HRI), to contribute to our understanding of this problem.

The creation of multilingual conversational agents – whether physically-embodied robots or digitally-embodied agents (e.g. virtual avatars, chatbots) – can serve as an important experimental tool, where the speech systems can be purposely manipulated across multiple languages in order to explore the cognitive effects of language structure. The goal is to help overcome the challenges of studying human speech in response to artificial lab stimuli (e.g. response times to recorded audio) by replacing that with actual conversation. However, human-human conversation characteristics can be difficult to replicate, especially if one hopes to introduce subtle changes using a human confederate. In contrast, an HRI approach can allow for consistently replicable manipulation, so that we can compare the effects across languages with many participants.

This paper summarizes two key lines of evidence to support this position. First, we describe existing research from psycho-linguistics about the surprising effects of bilingualism on human cognition. Second, we describe some preliminary research on bilingualism using bilingual robots and virtual avatars, which were designed to have an identical speech system across two languages that can be manipulated in various ways.

2 Existing Research

2.1 Bilingual Research on Cognition from Linguistics

There is ample research on the topic of bilingualism over the past half century from the field of linguistics, including into what is known as *linguistic relativity*. The concept of linguistic relativity has historically been a somewhat controversial topic, contending that the language one speaks influences how one thinks [4,5]. Nevertheless, many researchers have investigated the topic in recent years with renewed interest. For instance, Wang & Wei (2021) conducted experiments with bilingual speakers of Chinese, Japanese, and English and found that learning second languages (L2) causes "cognitive re-structuring" that surprisingly affects the speaker's native language (L1) speech patterns [6]. Pavlenko (2011) found that such cognitive re-structuring due to learning a second L2 language affects both verbal and non-verbal behaviors in the L1 language during speech interactions, possibly indicating that such effects go beyond simple word choice and rather impact deeper thought processes [7]. Similarly, Park (2020) showed that although Korean-English bilinguals (native L1 Korean speakers) fell in between Korean and English monolinguals in their speech patterns with influences from both, they were actually closer to the English monolinguals in both languages [8]. That is of particular relevance for our research here, which also examines how Korean-English bilinguals interact with virtual agents in their two languages . While many papers in that line of research focus on motion and spatio-temporal events, Athanasopoulos & Aveledo (2012) found evidence that bilinguals shift their color categorization more towards their L2 language, rather than falling in between L1 and L2 monolinguals. Moreover, they found that, incredibly, L2 proficiency appears to be more important than real-life immersion in an L2-speaking environment [9]. In other words, cognitive shifts occur with advanced proficiency, even if an individual never lived in an L2-speaking country. In short:

> *"Bilinguals appear to have a much more complex conceptual organization than previously thought, and may exhibit behavior that is unlike that of their monolingual peers." [9]*

A related topic to the above is the concept of code-switching between one language and another during an ongoing speech interaction, e.g. bilinguals switching between their L1 and L2 languages. Code-switching can occur multiple times during the same interaction, and it may involve either a wholesale switch from one language to the second or in some cases the insertion of individual words or phrases from the second language into a sentence spoken in the first language [10,11]. For our purposes, code-switching is also a useful experimental method to investigate the concept of linguistic relativity in bilingual speakers by purposely triggering code-switching events during an experiment. The idea is to more precisely identify/quantify crossover effects, which are the specific ways that one learned language influences another during speech. Such effects may be due to long-term cognitive re-structuring described earlier in this section, or perhaps due to more short-term residual influences during code-switching events.

2.2 Preliminary HRI Evidence

There is also some research within the field of human-robot interaction (HRI) into the effects of language differences on human interaction with autonomous agents. For instance, Skantze (2021) looked at how turn-taking cues can affect interactions between humans and conversational agents (both physical robots or virtual agents), with such cues varying across language and culture [12]. More broadly, there are a variety of HRI studies that have investigated the differences across languages during interactions with robots and virtual avatars [13,14], effects of linguistic differences on children interacting with robots [15], and the use of robots to assist with second-language learning [16,17]. Suffice it to say, language and culture are intimately intertwined, and the fields of HRI and human-computer interaction (HCI) are actively exploring what that means for how we design interactive technology.

	Control (std dev)	Bilingual (std dev)	*p*-val	*Sign.*
English				
Human	61.34 (27.4)	56.9 (37.8)	0.6735	
Avatar	49.72 (9.0)	45.55 (14.7)	0.0752	
Avatar-self	39.67 (5.5)	35.55 (9.7)	0.1076	
Avatar-ASR	10.06 (5.6)	10 (7.3)	0.9770	
Korean				
Human	17.55 (9.2)	53.95 (31.3)	0.0001	***
Avatar	16.3 (8.8)	39.25 (15.1)	0.0001	***
Avatar-self	15.4 (8.4)	36.25 (15.1)	0.0001	***
Avatar-ASR	0.9 (0.7)	3 (3.1)	0.0054	**

Table 1. Utterance Counts for each experimental session for bilingual versus monolingual (Control) speakers

In our own research, we have also been exploring how bilingual human speakers interact with bilingual robots and virtual agents in cooperative game environments, particularly in comparison to monolingual speakers [14, 18-20]. Typically, the game environment will utilize something like *Don't Starve Together*, a social survival game where players need to collect resources, make tools, fight monsters, and cooperate with each other to survive longer. We developed context-specific speech systems in multiple languages for such environments, allowing an autonomous virtual avatar to play the game along with the person while socially interacting with them. Such speech interaction includes both self-generated utterances in response to game events, as well as automatic speech recognition (ASR) based responses to human questions/comments.

For instance, in a recently completed study with 20 Korean-English bilinguals (L1 Korean, L2 English) interacting with a bilingual virtual avatar, we found that bilingual speakers speak both Korean *and* English more like monolingual English speakers than monolingual Korean speakers (total n=40). That was true both in terms of the amount of speech utterances, as well as the frequency of speech interruptions and the speech sentiment. For example, the results for amount of speech can be seen in Table 1. The experiments comprised 30-minute game sessions interacting with the virtual avatar described above, comparing two conditions. The Control condition was conducted with monolingual speakers in one single language. We then conducted a Bilingual condition with bilingual speakers that involved switching the language around the 15-minute mark. The conditions were identical other than the languages involved. All participants were required to be native or "advanced proficiency" speakers in either one or both of the languages (as measured by the TOEIC/TOPIK), depending on the condition. For fair comparison, the utterance counts in Table 1 were scaled for both conditions to equate to 15 minutes of gameplay. We also tested for "order effects", i.e. whether it mattered if the participants started in Korean then switched to English, or vice versa. However, there were no significant order effects in our dataset.

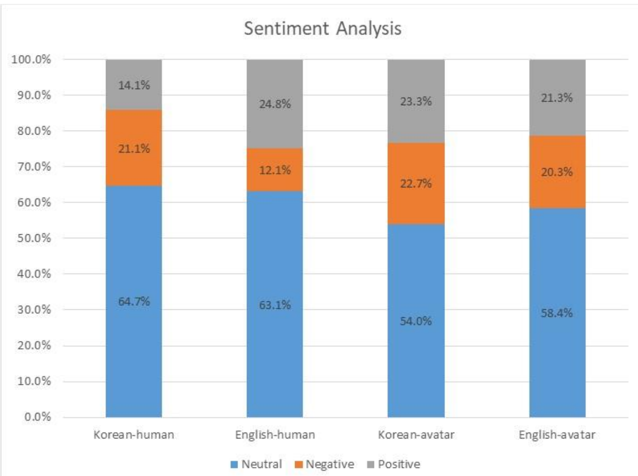

Figure 1. Speech Sentiment during Control condition

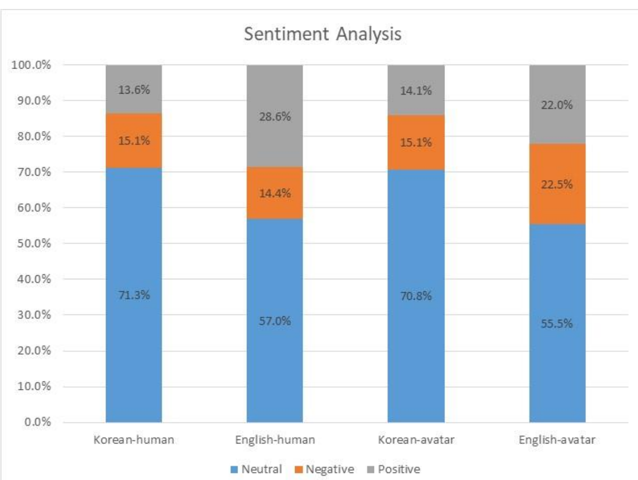

Figure 2. Speech Sentiment during Bilingual condition

Similarly, a comparison from the speech sentiment analysis can be seen in Figures 1 and 2. Sentiment was calculated using VADER in English [21] and Korean [22]. The main takeaway there was bilingual speakers exhibited less negative sentiment when speaking Korean than monolinguals, shifting into more neutral sentiment (compare "Korean-human" columns in both figures). In fact, their negative sentiment levels were closer to monolingual English speakers, though they were still not as positive. Interestingly, the bilingual speakers when speaking English were more positive and less neutral than English monolinguals (compare "English-human" columns in both figures), perhaps as some sort of subconscious "mental compensation" when speaking their non-native language. In short, the differences in the bilingual speech sentiment reflected a general shift towards their L2 second language, both when speaking the L2 language as well as their L1 native language.

Those results agree with the linguistic research findings described in Section 2.1, that learning a second language causes "cognitive shifts" in the speaker's brain that results in changes even when speaking in their native first language. What is interesting in the research here is that there appears to be evidence that the same phenomenon occurs during bilingual interactions with robots and virtual agents. Such results indicate that there may be enormous potential for the HRI research community to leverage interactive robots to better understand empirically the cognition behind how language and social communication in humans manifests. As suggested in the Introduction (see Section 1), that can be accomplished by designing HRI experiments to explore interactions across languages with the same robotic platform.

The key is in designing parallel context-specific speech systems, despite the syntactic or morphological differences in the languages. In other words, we must be aware of differences in how different languages handle "context", and differences in the use of implicit and explicit cues. Korean and English serve as good examples of this, with the former a high-context language where much is "left unsaid" to be inferred by the listener based on shared cultural understanding. That is averse to the latter (English) which is a low-context language requiring more explicit cues during speech interactions [23]. Suffice it to say, creating perfectly identical speech systems with direct translations would result in less-than-natural communication. **Rather, we should endeavor to create *parallel* speech systems**, where the "meaning" is the same in both languages even when how it is communicated is different. After doing so, we will have an experimental platform to further explore the cognitive effects of language on how humans think, which were described earlier in this section. Obviously, challenges exist for development of such bilingual robots and virtual agents, but such research holds tremendous potential for the HRI community to contribute to fundamental questions around the underlying mechanisms and evolution of human social communication.

3 Discussion

This position paper is at its essence a call for more research from the HRI community into bilingualism and/or multilingualism in humans, as an empirical method to better understand how language shapes the way humans think and behave. We lay out several lines of evidence from both linguistics as well as preliminary HRI evidence to support that call. We would be remiss though not to point out in conclusion that such research also holds huge potential for the field of HRI. If we can develop more sophisticated theories about social cognition in humans as it relates to how we communicate with others, then that may allow us to develop novel methods of communication between humans and robots [24].

To some degree, the methods currently used for communication between humans and robots are delimited by the ways humans communicate with each other even before the advent of modern technology (speech, gaze, gesture) or by traditions developed by the computing community for interacting with early computers (touchscreens, buttons, blinking lights). However, there is nothing that says that we must be limited by those methods. There could very well be better ways of communicating with autonomous robots and other interactive technology that we just haven't discovered yet. The cognitive effects of bilingualism (and more broadly *linguistic relativity*) may hold potential in that regard. In other words, making the effort to develop robots and virtual avatars to study bilingualism may in turn lead to better design of future HRI platforms.

3.1 Limitations & Future Challenges

There are some limitations and future challenges we should discuss, particularly in regards to the preliminary HRI evidence. First of all, there needs to be work comparing a broader array of languages. Many studies focus on comparisons of English versus some second language, but obviously the linguistic world is much broader than that. Moreover, given the global popularity of American pop culture, often even monolingual speakers of another language have been exposed to at least a small amount of English during their lifetime. Broadening the languages analyzed can mitigate those issues.

Beyond that, there is also a need for development of context-specific speech systems that can be used as replicable experimental platforms. Such context needs to be embedded within some goal-driven environment that *requires* the human and virtual agent/robot to communicate in order to complete some task. In other words, we are not just endeavoring to have some "small talk" casual conversation, but rather purposely engage the user's decision-making cognitive abilities in a socially-oriented manner. Doing so, though, requires a lot of work, especially if we hope to create replicable experiments [18]. This is an area where the HRI community can work together to address the challenge.

Finally, we would also note that there are limitations to only studying the cognitive effects of language in isolation, as language is also impacted by accompanying non-verbal communication in a myriad of ways [25,26]. As such, there is potential for multi-modal research that combines changes in the appearance/gestures of the robot or agent with speech system changes. How the HRI community may best go about tackling those challenges remains to be seen, as it is a multi-faceted problem. It is also very likely that non-verbal communication strategies may depend on the use setting, e.g. in user homes versus in a hospital, and that certain types of appearance/gesture may be more appropriate for one setting than another. Moreover, certain types of appearance/gesture may result in different cognitive effects depending on the setting. That is something that requires more research.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Research Foundation of Korea (NRF) (Grant number: 2021R1G1A1003801). We would also like to thank our various research collaborators over the years who have contributed to this work.

REFERENCES

[1] K. Gonzales, K. Byers-Heinlein, and A.J. Lotto, 2019. How bilinguals perceive speech depends on which language they think they're hearing. *Cognition*, 182, 318-330. DOI: https://doi.org/10.1016/j.cognition.2018.08.021

[2] P. Athanasopoulos, E. Bylund, G. Montero-Melis, L. Damjanovic, A. Schartner, A. Kibbe, et al., 2015. Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, 26(4), 518-526. DOI: https://doi.org/10.1177/0956797614567

[3] P.R. Doyle, L. Clark, and B.R. Cowan, 2021. What do we see in them? identifying dimensions of partner models for speech interfaces using a psycholexical approach. *CHI Conference on Human Factors in Computing Systems (CHI)*, 1-14. DOI: https://doi.org/10.1145/3411764.3445206

[4] P. Athanasopoulos, and A. Casaponsa, 2020. The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, 37(5-6), 393-412. DOI: https://doi.org/10.1080/02643294.2020.1769050

[5] L. Boroditsky, 2001. Does language shape thought?: Mandarin and English speakers' conception of time. *Cognitive Psychology*, 43(1), 1-22. DOI: https://doi.org/10.1006/cogp.2001.0748

[6] Y. Wang and L. Wei, 2021. Cognitive restructuring in the multilingual mind: language-specific effects on processing efficiency of caused motion events in Cantonese–English–Japanese speakers. *Bilingualism: Language and Cognition*, 24(4), 730-745. DOI: ttps://doi.org/10.1017/S1366728921000018

[7] A. Pavlenko, 2011. Thinking and Speaking in Two Languages: Overview of the Field. *Thinking and Speaking in Two Languages*, 77, 237. DOI: https://doi.org/10.21832/9781847693389

[8] H.I. Park, 2020. How do Korean–English bilinguals speak and think about motion events? Evidence from verbal and non-verbal tasks. *Bilingualism*, 23(3), 483-499. DOI: https://doi.org/10.1017/S1366728918001074

[9] P. Athanasopoulos, and F. Aveledo, 2012. Linguistic relativity and bilingualism. *Memory, language, and bilingualism: Theoretical and applied approaches*, 236-255.

[10] C.M. Scotton, and W. Ury, 1977. Bilingual Strategies: The Social Functions of Code-Switching. *Linguistics*, 193: 5-20. DOI: https://doi.org/10.1515/ijsl.1977.13.5

[11] P. Auer, 1998. *Code-Switching in Conversation: Language, Interaction and Identity (1st ed).* Routledge, London, United Kingdom. DOI: https://doi.org/10.4324/9780203017883

[12] G. Skantze, 2021. Turn-taking in conversational systems and human-robot interaction: a review. *Computer Speech & Language*, 67, 101178. DOI: https://doi.org/10.1016/j.csl.2020.101178

[13] S. Seok, E. Hwang, J. Choi, and Y. Lim, 2022. Cultural differences in indirect speech act use and politeness in human-robot interaction. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 470-477. DOI: http://dx.doi.org/10.5555/3523760.3523823

[14] C.C. Bennett and B. Weiss, 2022. Purposeful failures as a form of culturally-appropriate intelligent disobedience during human-robot social interaction. *International Conference on Autonomous Agents and Multiagent Systems (AAMAS): Best & Visionary Papers.* 84-90. Springer, Cham. DOI: https://doi.org/10.1007/978-3-031-20179-0_5

[15] Y. Kim, S. Marx, H.V. Pham, and T. Nguyen, 2021. Designing for robot-mediated interaction among culturally and linguistically diverse children. *Educational Technology Research and Development*, 69(6), 3233-3254. DOI: https://doi.org/10.1007/s11423-021-10051-2

[16] O. Engwall, J. Lopes, A. Åhlund, 2021. Robot interaction styles for conversation practice in second language learning. *International Journal of*

Social Robotics, 13(2), 251-276. DOI: https://doi.org/10.1007/s12369-020-00635-y

[17] V. Lin, H.C. Yeh, and N.S. Chen, 2022. A systematic review on oral interactions in robot-assisted language learning. *Electronics*, 11(2), 290. DOI: https://doi.org/10.3390/electronics11020290

[18] C.C. Bennett, B. Weiss, J. Suh, E. Yoon, J. Jeong, and Y. Chae, 2022. Exploring data-driven components of socially intelligent AI through cooperative game paradigms. *Multimodal Technologies and Interaction*, 6(2), 16. DOI: https://doi.org/10.3390/mti6020016

[19] J. Suh, C.C. Bennett, B. Weiss, E. Yoon, J. Jeong, and Y. Chae, 2021). Development of speech dialogue systems for social AI in cooperative game environments. *IEEE Region 10 Symposium (TENSYMP)*, 1-4. DOI: https://doi.org/10.1109/TENSYMP52854.2021.9550859

[20] C.C. Bennett, Y.H. Bae, J.H. Yoon, Y. Chae, E. Yoon, S. Lee, et al., 2022. Effects of cross-cultural language differences on social cognition during human-agent interaction in cooperative game environments. *Computer Speech & Language*. In Submission.

[21] C. Hutto and E. Gilbert, 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*. 8(1), 216-225. DOI: https://doi.org/10.1609/icwsm.v8i1.14550

[22] H.M. Park, C.H. Kim, and J.H. Kim, 2020. Generating a Korean sentiment lexicon through sentiment score propagation. *KIPS Transactions on Software and Data Engineering*, 9(2), 53-60. DOI: https://doi.org/10.3745/KTSDE.2020.9.2.53

[23] J.O. Yum, 1988. The impact of Confucianism on interpersonal relationships and communication patterns in East Asia. *Communications Monographs*, 55(4), 374-388. DOI: https://doi.org/10.1080/03637758809376178

[24] H.A Frijns, O. Schürer, and S.T. Koeszegi, 2021. Communication models in human–robot interaction: an asymmetric MODel of ALterity in human–robot interaction (AMODAL-HRI). *International Journal of Social Robotics,* 1-28. DOI: https://doi.org/10.1007/s12369-021-00785-7

[25] M. Knapp and J. Hall, 2010. *Nonverbal Communication in Human Interaction.* Thomas Learning, Wadsworth, Boston, MA USA.

[26] S.H. Tseng, Y.H. Hsu, Y.S. Chiang, T.Y. Wu, and L.C. Fu, 2014. Multi-human spatial social pattern understanding for a multi-modal robot through nonverbal social signals. *23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN),* 531-536. DOI: https://doi.org/10.1109/ROMAN.2014.6926307