Bennett CC, et al. (2023) Cognitive shifts in bilingual speakers affect speech interactions with artificial agents. *International Journal of Human–Computer Interaction*. In Press. https://doi.org/10.1080/10447318.2023.2261728

Cognitive Shifts in Bilingual Speakers Affect Speech Interactions with Artificial Agents

Casey C. Bennett[1,2,*], Say Young Kim[3], Benjamin Weiss[4], Young-Ho Bae[1], Jun Hyung Yoon[1], Yejin Chae[1], Eunseo Yoon[1], Uijae Ryu[1], Hansae Cho[1], Yesung Shin[1]

[1]Department of Data Science, Hanyang University, Seoul, Korea

[2]Department of Computing & Digital Media, DePaul University, Chicago, IL, USA

[3]Department of English Language & Literature, Hanyang University, Seoul, Korea

[4]Quality and Usability Lab, Technische Universität Berlin, Berlin, Germany

*** Corresponding Author:** Casey C. Bennett, cabennet@hanyang.ac.kr, ORC ID: 0000-0003-2012-9250

Abstract

A major research question in psycholinguistics relates to the phenomenon of *linguistic relativity*, which contends that the language one speaks influences how one thinks. Of particular interest is whether bilingual speakers shift cognitive paradigms when speaking different languages. Here, we conducted a human-agent interaction (HAI) study using a bilingual virtual avatar capable of autonomous speech during cooperative gameplay in two languages (Korean and English). We ran 40 participants, including 20 monolingual speakers (10 Korean, 10 English) and 20 Korean/English bilingual speakers, engaging the avatar during 30-minute game sessions. Comparison of speech patterns showed that bilingual speakers exhibited notable "cognitive shifts" in *both* languages while interacting with the avatar, which were markedly different from their monolingual counterparts. Interestingly, the virtual avatar's own speech behavior also significantly changed during interaction with bilingual speakers, despite identical programming. As evidenced here, such cognitive shifts appear to impact the way humans interact with artificial agents.

Keywords: human-robot interaction, social cognition, bilingualism, virtual avatar, linguistic relativity, cross-cultural robotics

1. Introduction

1.1 Background

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? Those are important, yet challenging, questions to answer.

Prior research has shown that bilingual speakers often feel as if "schizophrenic" with two completely different selves manifesting, depending on the language they were speaking at the time (Pavlenko, 2006, 2014). However, an open question is whether such differences would substantially impact the way people interact with virtual avatars, social robots, and other interactive technology capable of spoken or written communication. That has numerous implications for not only the way we design technology, but also for envisioning innovative future modes of communication between humans and technology that go beyond the limits of current language-based modes (e.g. speaking or typing in an existing human language such as English) (Cowan, 2014).

Likewise, other research has shown that bilinguals subconsciously adjust their listening strategies depending on which language they are cued to (Gonzales et al., 2019), and that such differences lead to context-specific differences in behavior (Athanasopoulos et al., 2015). Such findings suggest there is extreme flexibility in the communication strategies humans employ, but that flexibility is often confined within the linguistic boundaries we have learned in the past (Doyle et al., 2021). Getting humans outside those confines consciously is often difficult at best, but exploring those boundaries is critical for answering the questions outlined at the start of this section.

One possible strategy for such exploration is experimenting with bilingual speakers during interaction with bilingual artificial agents (e.g. virtual avatars or social robots) in cooperative environments, so that we can test multiple languages during actual conversation in a *replicable* experimental environment. The idea is to deliberately trigger the hypothesized phenomenon known as *linguistic relativity* in a controlled fashion (see next Section), within context-specific scenarios where cooperation between the human and agent is required to complete some task. Such an approach can overcome the replicability challenges of studying human speech in response to artificial lab stimuli (e.g. recorded audio) or trained human confederates. The bilingual speakers could then be compared back to monolingual speakers of both languages in the same experimental setup, allowing us to elucidate the linguistic boundaries and, more importantly, their effects on human-computer interaction (HCI). Such an approach could illustrate further how bilingual experiences affect social interaction via cognitive flexibility.

1.2 Linguistic Relativity Effects

The concept of *linguistic relativity* has historically been a somewhat controversial topic (Athanasopoulos & Casaponsa, 2020; Boroditsky, 2001), contending that the language one speaks influences how one thinks (as described in the previous Section 1.1). Nevertheless, many researchers

have investigated the topic in recent years with renewed interest. For instance, Wang & Wei (2021) found that learning second languages (L2) causes "cognitive re-structuring" that surprisingly affects the speaker's native language (L1) speech patterns with bilingual speakers of Chinese, Japanese, and English. 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. 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. That is of particular relevance for our research here, which also examines how Korean-English bilinguals interact with artificial 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. 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" (*Athanasopoulos & Aveledo, 2012, p.251)

Typically, such studies as those above have examined cognitive shifts of bilingualism by looking at either changes in mental categorization (e.g. color, motion), changes in speech pattern behavior (e.g. verbal interference), or neuro-imaging data. Prior studies also primarily have used artificial lab stimuli, either recorded audio or a human confederate, though for the latter human-human conversation can be difficult to replicate across participants. Excellent overviews of such experimental approaches to study bilingualism can be found in Pavlenko (2011) and Athanasopoulos & Casaponsa (2020).

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 another or in some cases the insertion of individual words or phrases from the second language into a sentence spoken in the first language, or vice versa (Scotton & Ury, 1977; Auer, 1998). 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 that linger for a few seconds/minutes.

1.3 Language Research in HRI & HAI

We would be remiss not to mention briefly the language-related research within the field of human-robot interaction (HRI), and more broadly HCI, into the effects of cross-cultural differences on human interaction with technology. HRI is intricately linked to the human-agent interaction (HAI), which we explore in this paper, with significant overlap between the two fields in terms of research questions and methodology. In particular, some of that prior research has sought to understand the influence of variation in verbal communication on both HAI and HRI. For instance, Skantze (2021) looked at how turn-taking cues can affect interactions between humans and conversational agents (whether physical robots or virtual agents), with such cues differing significantly across language and culture. More broadly, there are a variety of HRI studies that have investigated the differences across languages during interactions with robots and virtual avatars (Seok et al., 2022; Bennett & Weiss, 2022), effects of linguistic differences on children interacting with robots (Kim et al., 2021), and the use of robots to assist with second-language learning (Engwall et al., 2021; Lin et al., 2022). Suffice it to say, language and culture are intimately intertwined, with significant implications for how we design interactive technology.

1.4 Research Aims

The focus of this study is to adopt the strategy described in Section 1.1 above, where we investigate linguistic relativity in bilingual speakers while interacting with an artificial agent in multiple languages. To do so, we created a virtual avatar capable of autonomous speech during cooperative gameplay, with an identical context-specific speech system in two languages (Korean and English). We then conducted experiments with bilingual human speakers to investigate the effects on speech interactions when the languages were switched during the same experiment session (i.e. code-switching). The experiments took place in a customized cooperative game paradigm that we constructed, so that the virtual avatar and human had to meaningfully interact in both languages in order to accomplish some task. A separate control condition was conducted with monolingual speakers in only one language, for comparison. Our focus here was on changes to speech behavior patterns as a measure of cognitive shifts in bilingual speakers (see Section 1.2).

Our central hypothesis was that if bilingual speakers do indeed *think* differently when speaking different languages, then that should manifest in changes to the speech interactions that occur. To measure that objectively, we analyzed the speech behavior in multiple ways (amount of speech, frequency of interruptions, speech sentiment patterns), as well human cognitive perceptions of the virtual avatar. Our aims here are two-fold: 1) developing better experimental methods via robots and artificial agents in order to better test linguistic relativity in humans, and 2) examining linguistic relativity during HRI and HAI *specifically* in order to better design interactive robots and artificial agents in the future.

2. Methods

2.1 Virtual Avatar & Cooperative Game Environment

To investigate the hypotheses described in our Research Aims (Section 1.4), we developed a virtual avatar capable of autonomous speech during a cooperative survival game. The virtual avatar and speech system (henceforth the "Social AI") was the subject of extensive development and testing over several years, which has been described in detail previously elsewhere (Bennett, Bae, et al., 2023, Bennett, Weiss, et al., 2022; Suh et al.; 2021; Bennett & Weiss, 2022). The Social AI was capable of hundreds of different speech utterances covering 46 different utterance categories, each related to a particular game situation (e.g. collecting resources, fighting monsters, deciding where to go next) organized as a hierarchy with several levels. The utterances were derived from a series of human-human experiments in the same game environment, based on a conversation analysis to identify frequent conversational topics and verbal responses that humans made. The resulting speech utterances were both self-generated based on internal logic of the Social AI, as well as responses to human player speech via automatic speech recognition (ASR). The speech responses were purposely similar in both English and Korean to simulate a single individual who can speak both languages (i.e. the AI was bilingual, in essence). A newer "speech system 2.0" using GPT-3 for utterance generation is under development, but was not available yet at the time of these experiments.

As mentioned above, the system has been extensively described elsewhere, but in short it was implemented as custom code written in Python, using locally-installed (Windows or Mac) voice packages as part of the Text-to-Speech (TTS) module, with the audio output redirected to an internal "virtual" microphone jack. The ASR component used the Microsoft Azure speech-to-text API for human speech recognition in both English and Korean, using the Universal Language Model trained on language-specific data embedded within Azure for real-time detection. The speech output was then sent to the Loomie application (https://www.loomielive.com/), where we created a visual avatar capable of moving its lips synchronously with the speech running through the virtual microphone. The Loomie avatar was also capable of some basic built-in gestures, though we did not attempt to modify or enhance those for these experiments.

In the current study, we utilized a video game called *Don't Starve Together* for our cooperative game environment (https://www.klei.com/games/dont-starve-together), which can be downloaded from online sources such as Steam. The *Don't Starve Together* game is a social survival game where players need to collect resources, make tools, fight monsters, and cooperate with each other to survive longer. Similar to other social survival games (e.g. Minecraft), *Don't Starve Together* requires players to collect specific combinations of resources in order to build things, without which they will be vulnerable to various dangers and likely lose the game via player death, though there are multiple strategies that can be pursued (i.e. free-form). Moreover, it has cooperative multi-player gameplay modes (used here), which allow the players to cooperate on such tasks to survive. The tasks are under time constraints, however, as the level of danger gradually increases over time. As such, the game represents a free-form yet goal-oriented cooperative gameplay environment.

For the study here, the *Don't Starve Together* game was customized through the creation of a "Game Mod" for the experiment using LUA programming (https://www.lua.org/). That allowed us to modify the game conditions and create specific repeatable scenarios for each game session. Moreover, the customization allowed us to record game data from the backend game engine during the experiment, for later analysis. More details can be found elsewhere (Bennett, Weiss, et al., 2022; Suh et al., 2021).

2.2 Experimental Design

2.2.1 Participants

For the experiments here, we recruited 40 participants, with 20 in a Bilingual condition and 20 in a monolingual Control condition. For the Bilingual condition, all participants were required to be either native speakers or have advanced proficiency in ***BOTH*** English (TOEIC Level B2) and Korean (TOPIK Level C1, aka "level 5"). Given that the study was conducted in South Korea, that resulted in all bilingual speakers being L1 native Korean speakers who spoke English as a second language (L2). Additionally, we ran a monolingual Control condition with 20 participants (10 English, 10 Korean). The Korean monolinguals were local Koreans, while the English monolinguals consisted primarily of university exchange students attending school that semester in Korea. The Control participants had to meet the same proficiency requirements as the bilinguals listed above, but only for the single language in which they participated in the study. Across both conditions, the genders were balanced comprising 17 males and 23 females, with an average age of roughly 23.2 years. All participants were provided a brief 5-minute tutorial for how to play the game prior to the start of the experiment, in either Korean or English. The experiments and sample size calculations were approved by the Hanyang University IRB (#HYU-2021-138).

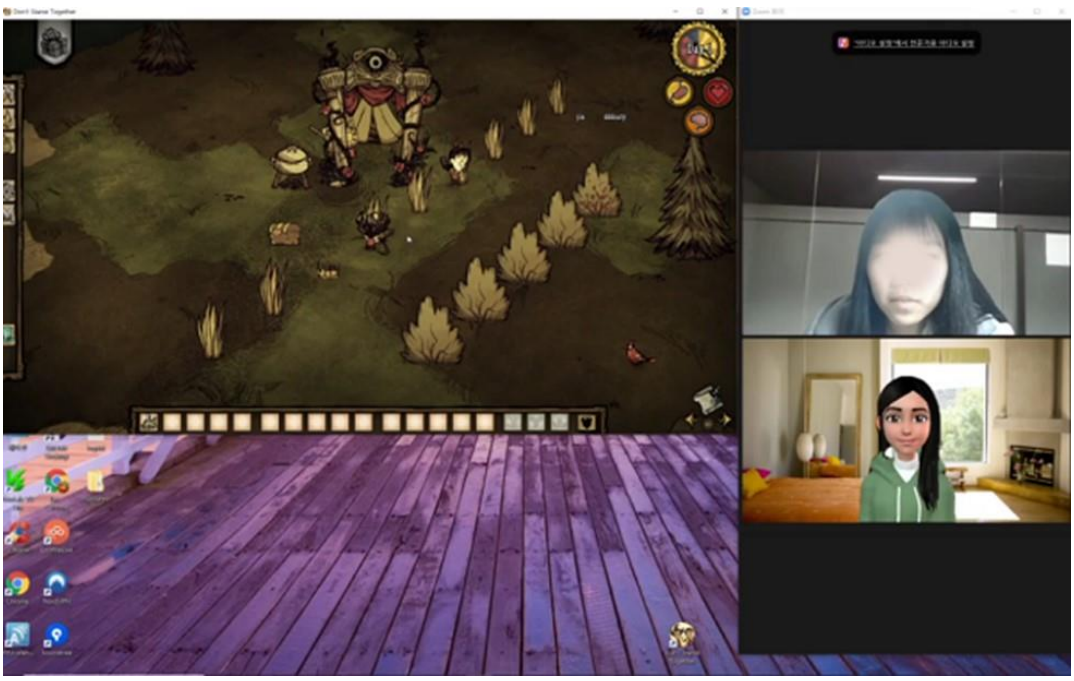
2.2.2 Experiment Setup & Procedure

The experimental setup involved two computers in two separate rooms, one for the human participant ("player computer") and one for the virtual avatar where its code was run ("confederate computer"), both linked to the same online game server. The player computer was further equipped with an HD camera, headphones, and Blue Snowball microphone for high-quality audio-visual input/output. Each game session involved one human participant and the virtual avatar, engaging in a 30-minute game session on a private server in 2-player cooperative gameplay mode. Zoom was used to allow direct audio-visual communication between the human and avatar while playing the game, in a side-by-side configuration. An example of this can be seen in Figure 1.

For the Control condition, the entire experiment session occurred in one language (either Korean or English, depending on the participant). For the Bilingual condition, the language was switched *exactly once* around the 15-minute mark when there was a pause in speaking. Bilingual participants were informed that this switch would occur prior to the experiment beginning and were informed to always speak to the avatar in whichever language it was currently speaking. However,

once the experiment had begun, participants were given no warning prior to language-switching. To check for "order effects", half the participants started the experiment in English then switched to Korean, while the other half started in Korean then switched to English.

Figure 1: Gameplay example during experiment (human vs avatar)



2.2.3 Description of Collected Data

During each experiment, we collected three main types of data: 1) audio-visual recordings of the gameplay, 2) written game data, and 3) instrument data of human perceptions. We used OBS Studio (https://obsproject.com/) to record the entire computer screen during gameplay, including the game window and the Zoom window of simultaneous social interactions. We used those to later extract the speech from the recordings for both the avatar and human player synced with in-game gameplay events. Written game data was also collected to later analyze how different gameplay events influenced the interactions. Finally, we collected several common HRI instruments at the end of each 30-minute game session, including the Godspeed scale (Bartneck et al., 2009) for measuring general perceptions of a robot/agent and the Networked Minds instrument (Biocca et al., 2001) for measuring social presence (Oh et al., 2018). The Godspeed is useful for evaluating various components that contributed to the human's overall perception of the robot/agent, such as lifelikeness and perceived intelligence. Meanwhile, Networked minds is useful for understanding how "immersive" an interactive experience was, which is referred to in the literature as "social presence".

2.2.4 Data Analysis Approach

To analyze differences between conditions in this paper, we first extracted the speech data from the OBS recordings of the experiments in order to create data for NLP analysis. This entailed using speaker diarization via Google Cloud services to automatically identify avatar and human participant speech in the recorded video of each game session, resulting in output transcripts with timestamps (so they

could be synced with in-game gameplay events). A few short snippets of some example conversations between the human participants and avatar from those transcripts (in both English and Korean) can be found in Tables S1-S4 in the online Supplementary Material. It was necessary to perform some post-diarization manual cleanup of those transcripts to ensure accuracy. For all analyses, the avatar speech and human speech were analyzed *separately*, though for the interruption analysis that did involve looking at the transcripts to identify speech overlaps between the avatar and human (see below). However, even then, we calculated the interruption frequency of the avatar interrupting the human and the human interrupting the avatar separately based on who was speaking first.

The speech data was then analyzed in multiple ways listed below, by condition. That entailed various statistical methods (two-tailed independent-samples *t*-tests) and data visualizations performed in either Python or R, which are described in the relevant sections in the results below (see Section 3). We also checked for "order effects" during the Bilingual condition, to see whether the language the experiment started in influenced the second language spoken after the switch (i.e. Korean to English versus English to Korean).

First, to understand the basic frequency of speech, utterance counts were calculated for both the human and avatar. Utterance counts for the avatar were further separated into two categories: self-generated speech and ASR responses to human speech. After that, we conducted an interruption analysis, looking for places where either the avatar interrupted the human participant by speaking while the human was still speaking (i.e. inter-pausal unit, or IPU), or vice versa the human interrupted the avatar (Skantze, 2021). This was done through manual annotation of the speech transcript data, by identifying places where timestamps of utterances overlapped without any pause between. Third, we also conducted sentiment analysis using lexical parsing via VADER (Hutto & Gilbert, 2014). For English, VADER was used directly, while a scientifically-validated Vader-like dictionary was used in Korean (Park et al., 2020). Finally, we analyzed the instrument data to evaluate whether there were differences in the human perceptions of the avatar social interaction during gameplay.

3. Results

3.1 Order Effects

We first tested for "order effects", i.e. whether it mattered if the participants started speaking in Korean during the game session then switched to English, or vice versa. The question was if the order of the languages would have any impact on the various results observed below (e.g. utterance count, interruption frequency, etc.). In order test this, our experimental design was purposely setup so that during the Bilingual condition half of the participants started in Korean then switched to English halfway through the game session, while the other half of the participants started in English then switched to Korean halfway. We then compared those two groups for all the same analyses described in Section 2.2.4. However, we **found no significant order effects in our dataset**, which suggests that any the effects of bilingualism seen in the results of this paper would be due to more long-term

cognitive re-structuring rather than short-term carryover at the time of code-switching (see Section 1.2). Thus, for all the results below, order effects are not considered.

3.2 Utterance Counts

After checking for any order effects, we subsequently compared the total amount of speech (i.e. utterance counts) in the different conditions during each game session, averaged across participants. Results can be seen in Table 1, with graphic visualizations of the same data in Figures 2 and 3 (for English and Korean, respectively). For fair comparison, the utterance counts were scaled for both conditions to equate to 15 minutes of gameplay, since the language was shifted halfway through the game session in the Bilingual condition.

Table 1: Overall Utterance Counts by Condition, with means (standard deviation)

	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 (14.8)	0.0752	
Avatar-self generated	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 generated	15.4 (8.4)	36.25 (15.1)	0.0001	***
Avatar-ASR	0.9 (0.7)	3 (3.1)	0.0054	**

Figure 2: Utterance Counts Visualized (English)

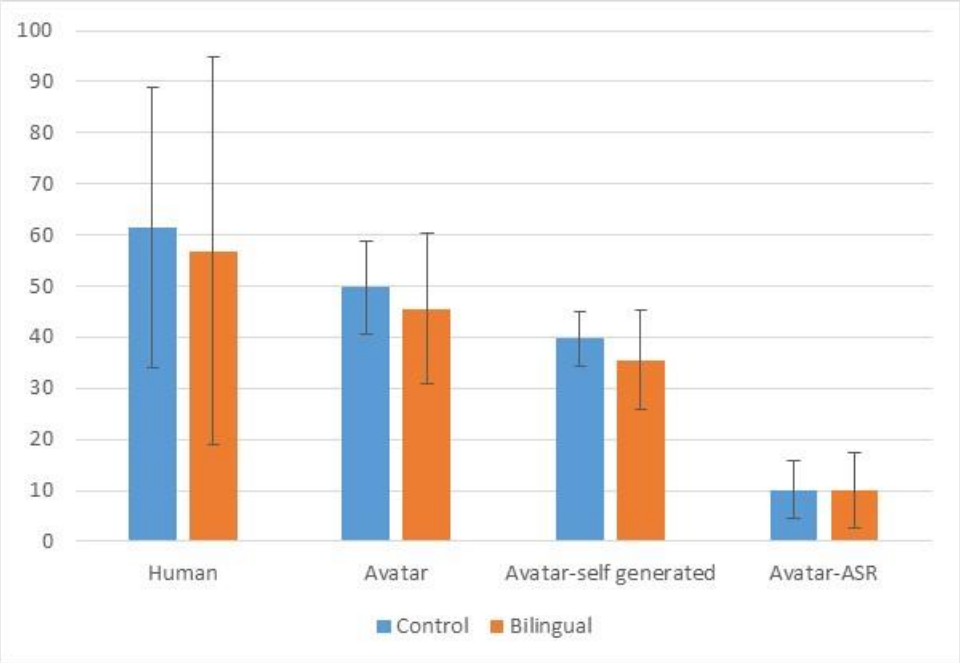
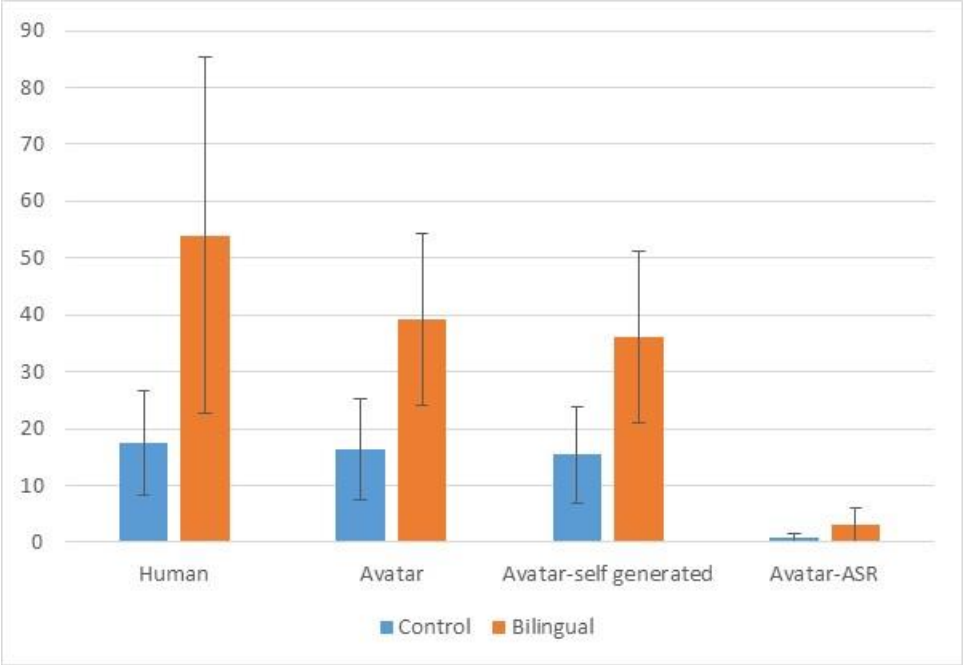


Figure 3: Utterance Counts Visualized (Korean)



As can be seen in the table, the significant differences were all on the Korean side. In short, the **bilingual speakers spoke *both* Korean and English more like monolingual English speakers,** while being quite distinct from monolingual Korean speakers. Those results agree with the linguistic research findings described in Section 1.2, 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. Interestingly, the Korean speakers increased speech amounts also affected the verbal interaction with the virtual avatar, causing the avatar to also speak more frequently in Korean as well, in similar amounts as when speaking in English. In other words, the cognitive effects of bilingualism appear to have an effect on speech interactions between humans and artificial agents, affecting the behavior of both parties.

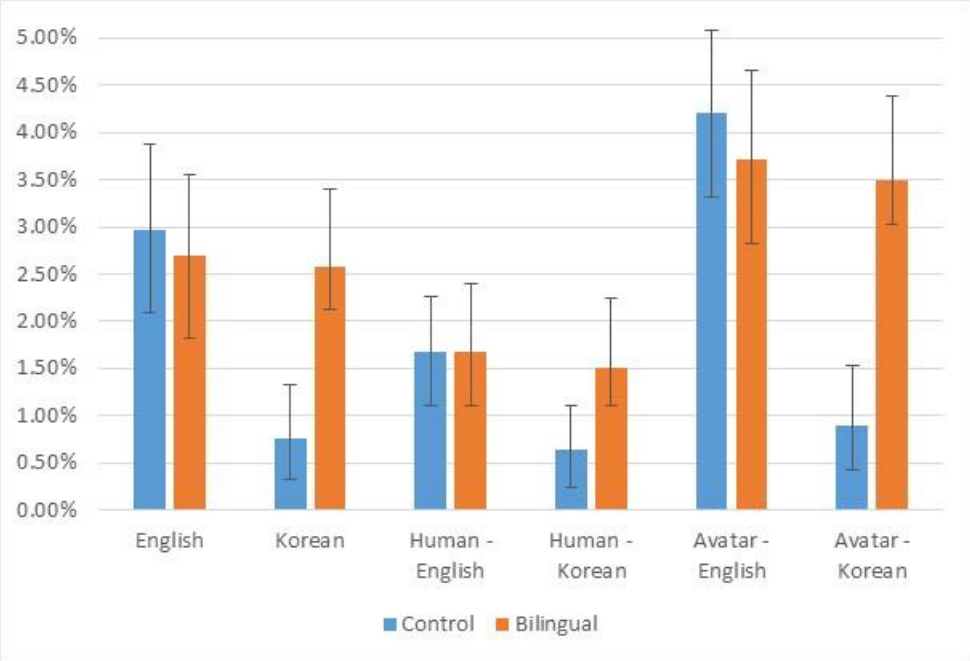
3.3 Interruption Frequency

We also evaluated the frequency of interruptions when one speaker interrupted the other, i.e. "turn-taking failures" as defined by Skantze (2021). Those were compared by speaker (avatar, human) across conditions and languages. Given the different utterance counts for different conditions and languages though (see Section 3.2), the interruption counts were calculated as a percentage of the total utterance count within each category, for fair comparison. Results can be seen in Table 2, as well as visualized in Figure 4.

Table 2: Interruption Frequency by Condition, with means (standard deviation)

	Control (std dev)	Bilingual (std dev)	*p*-val	*Sign.*
Language				
English	2.96% (3.3)	2.69% (3.0)	0.7877	
Korean	0.76% (1.3)	2.57% (2.8)	0.0122	*
Overall	1.80% (2.8)	2.59% (2.8)	0.3796	
Speaker				
Human - English	1.68% (1.3)	1.68% (2.1)	1.00000	
Human - Korean	0.64% (0.9)	1.51% (2.2)	0.11130	
Avatar - English	4.20% (3.1)	3.71% (3.6)	0.68040	
Avatar - Korean	0.90% (1.6)	3.50% (3.1)	0.00180	**

Figure 4: Interruption Frequency Visualized



As can be seen in the table and figure, interruption frequencies were much higher for the bilinguals when speaking Korean averse to Korean monolingual speakers. This is quite noticeable in Figure 4, where the bilinguals (orange bars) match for Korean and English for both the human and avatar. That is not the case for the monolingual speakers in the Control condition (blue bars). This is similar to what was observed for utterance counts in Section 3.2, where the bilingual speakers shifted their speech patterns to be more like English monolingual speakers, regardless whether they were speaking in their L1 language (Korean) or L2 language (English). Note that the standard deviations across participants were relatively high relative to the average values though, which resulted in less statistical significance in the differences compared to the utterance counts.

Overall, the avatar was more likely to interrupt the human than vice versa, which was likely due to the avatar agent's limited turn-taking prediction capabilities at the present time (Bennett, Bae, et al., 2023). Enhancing those abilities is something being explored in ongoing follow-up studies, but

regardless the bilingual effects on HAI were apparent on both the human and avatar side. What is quite interesting is that the avatar interrupted the bilingual participants more in Korean compared to the Control condition, even though no change was made to the avatar's speech system from the Control condition. That suggests that the **cognitive shifts in bilingual speakers toward their L2 language (English) resulted in speech interactions with the artificial agent that were fundamentally different than their native language monolingual peers**, even when speaking in their native language. That has potentially deep implications for the design of robots and other interactive speech-based devices in the future ... a bilingual speaker and a monolingual speaker may not interact the same way with such technology in their native language.

3.4 Sentiment Analysis

We further investigated whether there would be any effect due to cognitive shifts in bilingual speakers on the sentiment of spoken content (e.g. the frequency of positive versus negative utterances). A comparison from the speech sentiment analysis based on VADER can be seen in Figures 5 and 6, for the Control and Bilingual conditions respectively.

The main takeaway here was that bilingual speakers exhibited less negative sentiment when speaking Korean than monolinguals, shifting into more neutral sentiment (compare "Korean-human" columns in both figures, 71.1% vs. 64.7%). In fact, bilingual negative sentiment levels were closer to monolingual English speakers, though they were still not as positive. One may also note that there were notable shifts in the speech sentiment of the avatar when speaking Korean (more neutral sentiment, less positive/negative), but not on the English side. Similar to what was seen with the interruption frequency results in Section 3.3, the cognitive shifts in bilingual speakers appear to affect verbal interactions with artificial agents in multiple ways, even in their native language.

Figure 5: Speech Sentiment during Control Condition

Figure 6: Speech Sentiment during Bilingual Condition

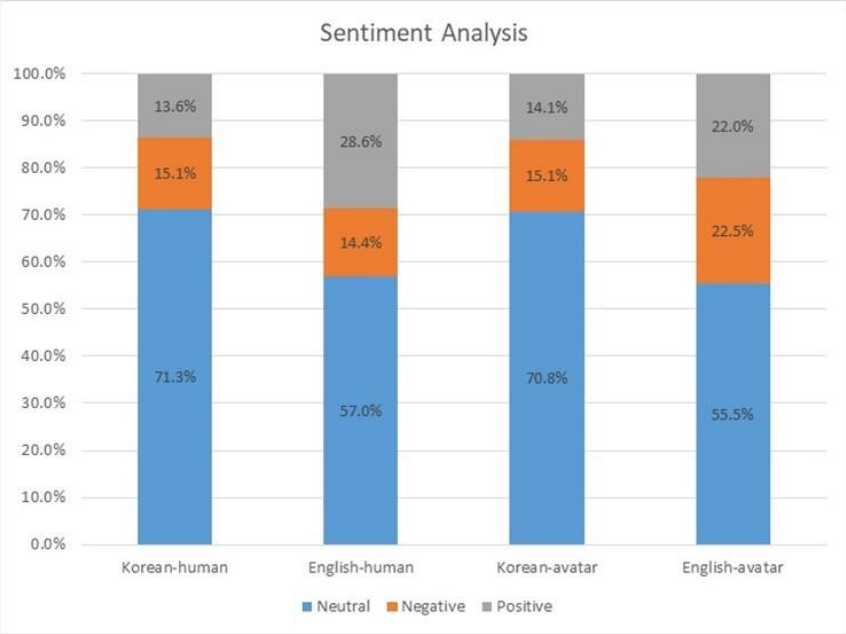


Table 3: Sentiment Analysis – Biliinguals vs Monolinguals (t-tests)

	English				Korean			
Sentiment	*Control*	*Bilingual*	***p*-val**	***Sign.***	*Control*	*Bilingual*	***p*-val**	***Sign.***
Positive	24.8%	28.6%	0.71020		14.1%	13.6%	1.00000	
Neutral	63.1%	57.0%	0.58830		64.7%	71.3%	0.63410	
Negative	12.1%	14.4%	0.55390		21.1%	15.1%	0.23700	

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. English monolinguals expressed significantly more positive utterances and less negative utterances than Korean monolingual speakers in the Control condition (24.8% vs. 14.1%, two-tailed independent t-test *p*-value=0.0012), as reported previously (Bennett, Bae, et al., 2023). Since the bilingual speakers here largely fell in between the monolingual groups, the bilingual differences to either group did not obtain statistical significance (see Table 3). Regardless, the differences in the bilingual speech sentiment appear to reflect a general shift towards their L2 second language, both when speaking the L2 language as well as their native language, similar to what was seen in previous sections (but not as strongly).

3.5 User Perceptions Based on Instrument Data

Additionally, we were curious if bilingual speakers would have any differences with monolingual speakers in their perceptions of the artificial agent, as measured using standardized HRI instruments. That focused on general perceptions of the virtual avatar (using the Godspeed instrument) as well as perceptions of its social presence (using the Networked Minds instrument), which are described in Section 2. The results of that analysis can be seen in Table 4.

Table 4: Instrument Analysis by Condition

	Control			Bilingual	*p*-val	*Sign.*
	Total	*ENG only*	*KOR only*	**Total**		
Godspeed (total)	3.44	*3.57*	*3.30*	2.93	0.00020	***
Subscales						
Anthropomorphism	3.00	*3.10*	*2.90*	2.45	0.00377	**
Animacy	3.37	*3.45*	*3.28*	2.83	0.00738	**
Likeability	3.81	*4.02*	*3.59*	3.28	0.02532	*
Perc. Intelligence	3.58	*3.70*	*3.46*	3.03	0.00517	**
Perc. Safety	3.45	*3.63*	*3.27*	3.13	0.03824	*
NM Self	3.20	*3.22*	*3.18*	3.05	0.20873	
NM Other	3.31	*3.35*	*3.28*	3.16	0.14098	

The main takeaway here was that bilingual speakers rated the interaction with the virtual avatar consistently lower on all instruments than either the monolingual Korean or English speakers during the Control condition. However, the only statistically significant difference was in the Godspeed scores related to lower cognitive perceptions of robot during Bilingual condition, specifically for ratings of likeability, intelligence, and anthropomorphism. One could hypothesize that perhaps seeing the virtual avatar speak in both languages (Korean and English) made its flaws more apparent, but it is not exactly clear at this point why the ratings were lower. Regardless, interacting with an artificial agent speaking in multiple languages appears to have an effect on human user perceptions of the agent, which merits further research.

4. Discussion

4.1 Summary of Results

This study explored the concept of *linguistic relativity* by having humans interact with a bilingual virtual avatar in a cooperative game environment. The virtual avatar was equipped with an identical context-specific speech system in two languages (Korean and English). Participants included both bilingual human speakers (L1 Korean, L2 English) with whom the avatar's spoken language was switched mid-experiment, as well as monolingual speakers in both languages who participated in the experiment in only one language for comparison. Results showed that there the bilingual speakers' interactions with the virtual avatar were significantly different from the monolingual speakers.

More specifically, we found that bilingual speakers spoke *both* Korean and English more like monolingual English speakers. In other words, learning a second language appears to cause "cognitive shifts" in humans that have an effect even when speaking their native first language. That was true for both their overall amount of speech (i.e. utterance count), as well in their turn-taking behavior (i.e. interruption frequency) and speech sentiment. Interestingly, the virtual avatar's own speech behavior also significantly changed during interaction with bilingual speakers, even though no change was made to the avatar's speech system from the Control condition with monolingual speakers. Thus the

cognitive changes related to language seem to spill over into interactions between humans and technology in a number of unexpected ways.

To summarize, cognitive shifts in bilingual speakers toward their L2 language resulted in speech interactions with the artificial agent that were fundamentally different than their native language monolingual peers. This aligns with previous research from psycholinguistics of the cognitive effects of bilingualism on humans (Wang & Wei, 2021; Pavlenko, 2011; Park, 2020; Athanasopoulos & Aveledo, 2012), extending those findings into the realm of HAI and HRI.

4.2 Implications for Bilingual Avatars & Robots

This research has a number of potential long-term implications. There is a distinct possibility that we may be able to develop a novel communication "language" for human interaction with technology that is independent of natural human languages or other modes of communication (Frijns et al., 2021). We can see from the results here that language itself shapes how we think, so it may be a limiting factor in taking full advantage of technology. Such novel communication modes may still be symbolic like natural languages, but stripped of any cultural baggage that might interfere with effective communication with machines. Likewise, that kind of approach could also be used to reduce the coding length of communicated messages (in terms of Shannon's "information theory"), thereby increasing the information density and making HRI/HAI more efficient (Shannon, 1948). Bilingual speakers are thought to have better attentional control especially in noisy environments, which further supports the idea that irrelevant parts of the message could be trimmed out if a technology-specific language was developed (Zhou & Krott, 2016; Hilchey & Klein, 2011).

Along the same lines, we would be remiss not to point out that this area of research also holds huge potential for the field of HRI from a "robot design" standpoint. 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 how we physically design social robots 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 (Honig & Oron-Gilad, 2018; Hellström & Bensch, 2018). The cognitive effects of bilingualism (and more broadly linguistic relativity) may hold potential in that regard, by elucidating the behavioral changes that occur when the same individual speaks in a different language. It may very well be that those changes extrapolate to other forms of communication, such as across different dialects of the same language and non-verbal cues. If so, it may be possible to design robot communication systems in a way in order to trigger certain human interaction styles, e.g. via code-switching by the robot in certain scenarios, such as changing from standard language into dialect in order to trigger more in-group behavior towards the robot (Bennett & Lee, 2023). Moreover, many robots now include "digital interfaces" (e.g. screens) that incorporate both embodied interaction as well as virtual interaction on the same platform (i.e. "mixed reality" or AR), blurring the lines between

physical and digital communication. In other words, making the effort to develop robots and virtual avatars to study bilingualism may in turn lead to better physical design of future social robots and HRI platforms.

We also note there is some argument in the psycho-linguistics field about the purported "cognitive advantage" of bilingualism in a general sense beyond specific improved abilities (e.g. attentional control), with evidence for (Marian & Shook, 2012) and against that (Von Bastian et al., 2016). We should point out that the findings from our research here are independent of that debate. Regardless of whether there are advantages of being bilingual or not in general, there are certainly cognitive changes that occur in human brains in response to second language learning, which appear to extend into interactions with artificial agents based on the results here. One question that naturally might arise from that is whether we need to design robots and artificial agents differently for bilingual vs monolingual speakers? That is indeed a very good question. Our take is that the appropriate design may depend on whether the agent/robot is intended to be deployed in multilingual and/or multicultural environments or not, i.e. the situated context of use (Lee & Sabanovic, 2014). Most robots and artificial agents nowadays are either built for a particular cultural/linguistic setting, or treat various diverse cultural/linguistic settings as the same, or assume scenarios where a robot/agent is used in multilingual/multicultural environments are equivalent to a monolingual environment. However, none of those may be valid assumptions, as our bilingual research here suggests. In our increasingly interconnected global world, such multilingual and/or multicultural environments are becoming more common, so this is a question we should consider carefully.

4.3 Limitations & Future Challenges

There are a number of limitations to this study, many of which also represent future research challenges. First of all, there should be more exploration of potential factors that might mitigate the effects of bilingual cognitive shifts during speech interactions between humans and artificial agents. For instance, in the current study we only switched the languages one time, but repeatedly switching the languages back and forth during the same experiment might produce other code-switching "crossover effects". We are currently working on an experiment where the languages will be switched multiple times per game session without warning. Beyond 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 (Tseng et al., 2014; Knapp & Hall, 2010). As such, there is potential for multi-modal research that combines changes in the appearance/gestures of the robot or agent with changes the spoken language. Furthermore, certain types of appearance/gesture may result in different cognitive effects depending on the setting. That is something that requires more research.

Second, there needs to be work comparing a broader array of languages. Many studies in psycho-linguistics and cognitive science focus on comparisons of English versus some second language, but obviously the linguistic world is much broader than that (Blasi et al., 2022). 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, as well perhaps expose other types of cognitive shifts that don't occur with English.

Finally, beyond the above, there is also a need for development of more context-specific speech systems that can be used as replicable experimental platforms in HRI/HAI. 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 to 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 that can be manipulated to explore different hypotheses (Bennett, Weiss, et al., 2022). This is an area where the HRI and HAI communities can work together to address the challenge (Baxter et al., 2016).

Acknowledgements

We would also like to thank our various collaborators, including Jaeyoung Suh, Jihong Jeong, and Seeun Lee (Hanyang University), for their assistance in this work.

Funding

This work was supported by a grant from the National Research Foundation of Korea (NRF) under grant number: 2021R1G1A1003801.

Disclosure Statement

The authors report there are no competing interests to declare.

Data Availability Statement

The data used in this study included audiovisual recordings of participants during the experiment. However, it may be made available in de-identifiable form upon reasonable request.

References

Athanasopoulos, P., & Aveledo, F. (2012). Linguistic relativity and bilingualism. *Memory, Language, and Bilingualism: Theoretical and Applied Approaches*, 236-255.

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., ... & Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, 26(4), 518-526.

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, 37(5-6), 393-412.

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

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Tobotics*, 1(1), 71-81. https://doi.org/10.1007/s12369-008-0001-3

Baxter, P., Kennedy, J., Senft, E., Lemaignan, S., & Belpaeme, T. (2016). From characterising three years of HRI to methodology and reporting recommendations. *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (pp. 391-398).

Bennett, C.C., Weiss, B., Suh, J., Yoon, E., Jeong, J., & Chae, Y. (2022). Exploring data-driven components of socially intelligent AI through cooperative game paradigms. *Multimodal Technologies and Interaction*, 6(2), 16.

Bennett, C. C., & Weiss, B. (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*, (pp. 84-90). Springer, Cham.

Bennett, C.C., Bae, Y.H., Yoon, J.H., Chae, Y., Yoon, E., Lee, S., Ryu, U., Kim, S.Y, & Weiss, B. (2023) Effects of cross-cultural language differences on social cognition during human-agent interaction in cooperative game environments. *Computer Speech & Language*, 81, 101521. https://doi.org/10.1016/j.csl.2023.101521

Bennett, C. C., & Lee, M. (2023). Would People Mumble Rap to Alexa? *Proceedings of the 5th ACM International Conference on Conversational User Interfaces (CUI)*, (pp. 1-5). https://doi.org/10.1145/3571884.3603757

Biocca, F., Harms, C., & Gregg, J., (2001). The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. *4th Annual International Workshop on Presence*, (pp. 1-9).

Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, 26(12): 1153-1170.

Boroditsky L. (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

Cowan, B. R. (2014). Understanding speech and language interactions in HCI: The importance of theory-based human-human dialogue research. *Designing Speech and Language Interactions Workshop, ACM Conference on Human Factors in Computing Systems (CHI)*, (Vol. 10, No. 2559206.2559228).

Doyle, P.R., Clark, L., & Cowan, B. R. (2021). What do we see in them? identifying dimensions of partner models for speech interfaces using a psycholexical approach. *ACM Conference on*

Human Factors in Computing Systems (CHI), (pp. 1-14). https://doi.org/10.1145/3411764.3445206

Engwall, O., Lopes, J., & Åhlund, A. (2021). Robot interaction styles for conversation practice in second language learning. *International Journal of Social Robotics*, 13(2), 251-276. https://doi.org/10.1007/s12369-020-00635-y

Frijns, H. A., Schürer, O., & Koeszegi, S. T. (2021). Communication models in human–robot interaction: an asymmetric MODel of ALterity in human–robot interaction (AMODAL-HRI). *International Journal of Social Robotics*, 15(3), 473-500.

Gonzales, K., Byers-Heinlein, K., & Lotto, A. J. (2019). How bilinguals perceive speech depends on which language they think they're hearing. *Cognition*, 182, 318-330.

Hellström, T., & Bensch, S. (2018). Understandable robots-what, why, and how. *Paladyn, Journal of Behavioral Robotics*, 9(1), 110-123.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, 18, 625-658.

Honig, S., Oron-Gilad, T. (2018). Understanding and resolving failures in human-robot interaction: Literature review and model development. *Frontiers in Psychology*, 9, 861. https://doi.org/10.3389/fpsyg.2018.00861

Hutto, C., Gilbert, E. (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.

Kim, Y., Marx, S., Pham, H. V., & Nguyen, T. (2021). Designing for robot-mediated interaction among culturally and linguistically diverse children. *Educational Technology Research and Development*, 69(6), 3233-3254.

Knapp, M. & Hall J. (2010). *Nonverbal Communication in Human Interaction*. Thomas Learning, Wadsworth.

Lee, H.R. & Šabanović, S. (2014). Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. *9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, (pp. 17-24). https://doi.org/10.1145/2559636.2559676

Lin, V., Yeh, H. C., & Chen, N. S. (2022). A systematic review on oral interactions in robot-assisted language learning. *Electronics*, 11(2), 290.

Marian, V., & Shook, A. (2012). The cognitive benefits of being bilingual. In *Cerebrum: The Dana Forum on Brain Science* (Vol. 2012, pp. 13). Dana Foundation.

Oh, C.S., Bailenson, J.N., Welch, G.F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 114. https://doi.org/10.3389/frobt.2018.00114

Park, H. I. (2020). How do Korean–English bilinguals speak and think about motion events? Evidence from verbal and non-verbal tasks. *Bilingualism*, 23(3), 483-499.

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

Pavlenko, A. (2006). Bilingual selves. *Bilingual education and bilingualism* (pp.1-33). Multilingual Matters LTD.

Pavlenko, A. (2011). Thinking and Speaking in Two Languages: Overview of the Field. *Thinking and Speaking in Two Languages* (pp.237-257). Multilingual Matters LTD.

Pavlenko, A. (2014). *The Bilingual Mind: And What It Tells Us About Language and Thought*. Cambridge University Press.

Scotton, C. M., & Ury, W. (1977). Bilingual Strategies: The Social Functions of Code-Switching. *Linguistics*, 193: 5-20.

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

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system Technical Journal*, 27(3), 379-423.

Skantze, G. (2021). Turn-taking in conversational systems and human-robot interaction: a review. *Computer Speech & Language*, 67, 101178.

Suh, J., Bennett, C. C., Weiss, B., Yoon, E., Jeong, J., & Chae, Y. (2021). Development of speech dialogue systems for social AI in cooperative game environments. *IEEE Region 10 Symposium (TENSYMP)*, (pp. 1-4).

Tseng, S.H., Hsu, Y.H., Chiang, Y.S., Wu, T.Y., & Fu, L.C. (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)*, (pp. 531-536).

Von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, 145(2), 246.

Wang, Y., & Wei, L. (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.

Zhou, B., & Krott, A. (2016). Data trimming procedure can eliminate bilingual cognitive advantage. *Psychonomic Bulletin & Review*, 23, 1221-1230.