Findings from Studies on English-Based Conversational AI Agents (including ChatGPT) Are Not Universal

Casey C. Bennett*
College of Computing & Digital Media
DePaul University
Chicago, IL, USA
cbenne33@DePaul.edu

ABSTRACT

A common, but largely untested, assumption in artificial intelligence (AI) and speech systems is that results from experiments using only the English language as the communication medium will hold true across any other language or cultural context. We argue here, based on emerging recent scientific evidence, that such an assumption appears to be invalid. In fact, there appear to be stark differences across languages and cultures when experiments are conducted using the same artificial speech system setup to be able to communicate in more than one language. Moreover, using those AI systems with bilingual human speakers shows that their behavior, social cues, and communication patterns change when language "code-switching" occurs within the same experiment session. To illustrate our point further, in the second half of the paper we give the specific example of ChatGPT (as the backbone speech content for artificial speech systems) being used for older adults with dementia and Alzheimer's, who often have altered speech patterns (e.g. slurred pronunciation). There are emerging reports from such research of severe limitations of ChatGPT in such contexts, which highlights the dangers of assuming findings from a narrow range of linguistic and/or cultural contexts can fully capture some universal truths about human communication with artificial agents. Finally, we point out that the reluctance of scientific journals and conferences to publish *negative results* means many of those emerging reports are only being reported anecdotally, which is problematic for the field of conversational user interfaces (CUI).

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; • **Computing methodologies** → **Natural language processing**; • **Computer systems organization** → *Robotics.*

KEYWORDS

Language, Artificial Agents, Speech Systems, Bilingualism, Dementia, Experimental Bias

*Corresponding Author

CUI '24, July 08–10, 2024, Luxembourg, Luxembourg

ACM ISBN 979-8-4007-0511-3/24/07
https://doi.org/10.1145/3640794.3665879

ACM Reference Format:
Casey C. Bennett. 2024. Findings from Studies on English-Based Conversational AI Agents (including ChatGPT) Are Not Universal. In *ACM Conversational User Interfaces 2024 (CUI '24), July 08–10, 2024, Luxembourg, Luxembourg*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3640794.3665879

1 INTRODUCTION

1.1 Overview

A common, but largely untested, assumption in artificial intelligence (AI) and artificial speech systems is that results from experiments using only the English language as the communication medium will hold true across any other language or cultural context. Such an assumption is rooted in the notion that, though differences across languages and cultures certainly exist, underneath those differences lies some *universal pattern* that serves as a substratum upon which linguistic/cultural variation manifests. As such, many studies across a range of fields from AI to linguistics to cognitive science are conducted in the English language, without comparing other languages. The results of such studies are then typically published without that single-language caveat, arguing that those findings represent some universal truth.

However, there is emerging scientific evidence that such an assumption of "English Language findings = universal truth" appears to be invalid (our primary provocation point). We provide evidence toward that point in Section 2 below. In short, there appear to be idiosyncrasies about the English language which are not representative of all languages [13]. Moreover, there appear to be stark differences across languages and cultures when experiments are conducted using the same artificial speech system setup to be able to communicate in more than one language. Those differences also manifest during artificial agent interactions with bilingual human speakers when code-switching occurs *within the same* experimental session, altering the communication styles and social cues significantly despite the language spoken being the only change. In other words, the way humans interact verbally with artificial agents appears to vary depending on the language being spoken at the time.

Interestingly, there are emerging anectodal reports of the use of large language models (LLMs) like ChatGPT in artificial speech systems that interact with older adults who have dementia and Alzheimer's, where similar problems are occurring (see Section 3). Such older adults often exhibit "changed speech" patterns, which alter their pronunciation, turn-taking, and other speech behaviors [11]. A fundamental problem is that many of those reports are only being reported anecdotally, due to reluctance of scientific journals and conferences to publish *negative results*, which is leading to a biased pool of research in the scientific literature on the use of such LLMs for artificial speech systems. Such bias is not simply a problem of bad knowledge, but can also serve to mislead the design of future research studies resulting in a "compounding effect".

That issue will ultimately lead to delays in the development of conversational user interfaces (CUIs) and speech systems for artificial agents/robots in the future. As such, we argue here that this is a fundamental problem for the CUI community to address at this point in time. The paper is structured as follows: first we lay out the evidence for our main provocation point in Section 2, then illustrate our point further in Section 3 via a specific example of using LLMs like ChatGPT for artificial speech systems that interact with older adults who have dementia and Alzheimer's.

2 PROVOCATION POINT EVIDENCE

As stated in Section 1, our primary provocation point is that emerging scientific evidence suggests that studies of speech systems for artificial agents using English-only as the language medium do not represent a "universal truth", and that in fact when other languages are used the results of research studies can be significantly different.

To that point, Blasi et al. (2022) has shown that there are number of notable idiosyncrasies in the English language that are not common in many other languages [13]. They have demonstrated that those idiosyncrasies impact not only speech patterns but also underlying cognitive processes in humans. In other words, the idiosyncrasies in English (or any language for that matter) affect not only how we speak but also how we think. Critically, the aforementioned work of Blasi et al. has shown that those idiosyncrasies have seriously affected the existing body of knowledge in cognitive science potentially leading to a number of misconceptions in that field, due to over-reliance on English-only studies. Indeed, some researchers have even dubbed English an "outlier" amongst world languages due to its unusual orthography and atypical grammatical structure, both of which have been linked to neural processing of language [14]. A recently released pre-print from Atari et al. (2023) used a battery of standardized psychological measure tests to show that LLMs respond to such tests in ways that most resemble people from Western countries where English is the dominant language, rather than as some generic human exemplar [5].

Elsewhere, Bennett et al. (2023a) conducted recent studies directly comparing English and Korean speakers interacting with a bilingual virtual avatar (Korean/English) in a cooperative video game environment. Their results showed significant differences in the speech patterns of the humans depending on the language spoken, including effects on the frequency of speech, turn-taking, and sentiment [9]. Interestingly, they found that those differences also affected the way the artificial agent spoke, even though its speech system *programming was exactly the same in both scenarios* other than the language spoken. In a follow-up study, Bennett et al. (2023b) ran the same experiment with bilingual participants (Korean/English) which involved the artificial agent "code-switching" (i.e. changing languages) during the experiment session with each participant, finding that bilingual speakers (while interacting with an artificial agent) were significantly different in their speech patterns in both languages in comparison to their mono-lingual peers, even those from their native language [10]. That is problematic for research on CUI and artificial speech systems using only English, given that although an estimated 1 out of 5 people in the world speak English with some level of proficiency, the vast majority of those (75%) are non-native bilingual speakers [16]. That raises the question of who exactly is being studied in such research.

Other studies have directly compared other languages during interactions between artificial agents and humans and found notable differences, including English vs. Arabic [4] and Japanese vs. Serbian [26]. A general review of cross-cultural differences during human-robot interaction (HRI) is given in [23]. Surprisingly, it is not only the production of verbal speech itself that is affected by linguistic and/or cultural differences, but also how the human brain processes things like music [24] and color perception [6].

More broadly, there is a wealth of research from the field of bilingualism on speech pattern differences *within* a single individual, depending on which language they are speaking at the time [7]. That research has shown how bilingual speakers shift cognitive processes during code-switching, as well as how their L2 language (i.e. non-native second language) affects speech patterns in their L1 language (i.e. native language) but not vice versa. In other words, the effects of bilingualism illustrate the complexities of how language influences both thought and human behavior in often surprising ways. For our purposes here though, they underscore the problems of primarily using one language for all research into artificial speech systems, CUIs, LLMs, etc. **To do so assumes that there is a simple one-to-one mapping of speech onto the underlying cognitive processes in humans that produce such speech, and that is clearly not the case.**

We would be remiss not to mention the non-verbal social cues that occur during speech interactions, which have also been shown to be language-specific. In fact, many current robots and other artificial agents incorporate language-specific social cues (typically English) that have been noted to significantly impact human interaction with the robot [29], while at the same time limiting their utility with second language learners (e.g. non-native English speakers) [15, 22]. Elsewhere, Skantze (2021) has looked specifically at how turn-taking cues can affect interactions between humans and conversational agents, with such cues differing significantly across language and culture [27]. Interestingly, other research has also shown that the *exact same* social cue in one culture may be interpreted differently in another culture, and that the expression of such cues is often involved in the covert signaling of one's social status [11]. In short, it appears that similar complexity exists across languages on the non-verbal side as well as verbal side.

A summary of the evidence presented in this section is provided in Table 1, along with relevant cited papers, though we note that this article was not intended as a full meta-review so additional examples may exist.

3 EXAMPLE - LLM'S FOR DEMENTIA SPEECH SYSTEMS

To illustrate the point of Section 2 further, it is useful to look at a specific example of how different styles of communication can impact speech interactions between humans and artificial agents. To do so, we can take the example of using LLMs like ChatGPT

Table 1: Evidence Summary

Category	Description	Example Papers
Cross-Language - Different Populations	Research using artificial agents speaking in multiple languages with populations of human native speakers of those languages	9,4,26
Cross-Language - Bilingual Speakers	Research using artificial agents speaking in multiple languages with the same human bilingual speaker	10,24,6,7
Cognitive Science Research	Studies from Cognitive Science on the idiosyncracies of the English language compared to other languages	13,14,5
Non-Verbal Social Cues	Research on how non-verbal social cues vary across languages, e.g. turn-taking signals	29,15,22,27,11

for artificial speech systems that interact with older adults who have dementia and Alzheimer's. Such older adults often exhibit "changed speech" patterns - such as slurred pronunciation, altered grammatical structures, prosody changes, pragmatic impairments (e.g. staying on topic), and disrupted turn-taking (e.g. elongated interpausal units [IPUs]), among others - that get progressively worse as their health condition progresses. Those changes are thought to be due to nerve cell failure, and result in significant communication difficulties that are almost as if speaking another language [8, 18]. Some researchers have even reported that changed speech characteristics can be used to clearly differentiate between dementia sub-types, such as Alzheimer's and Lewy-body dementia [31]. Critically to our point in Section 2, it has been reported that those **speech dysfunctions in dementia may vary depending on the patient's spoken language and that English speakers offer a sub-optimal benchmark for other language groups** [17].

Over the past decade, a number of researchers (including ourselves) have begun exploring development of speech systems for conversational artificial agents specifically to interact with individuals living with dementia and Alzheimer's, due to those speech changes. In the last few years, researchers have begun to use LLMs like ChatGPT in such efforts, attempting to utilize their transformer-based approach to create more flexible speech interactions. However, a slew of recent reports in the past 12 months have highlighted numerous problems with such LLM transformer-based dementia speech systems, such as keeping the artificial agent on-topic, appropriate turn-taking (not interrupting speech), integration of multiple modalities (e.g. non-verbal cues), etc. For instance, Irfan et al (2023) conducted a human user study with dementia participants in Sweden interacting with such a speech system deployed on a robot, finding that the robot frequently interrupted the users, responded slowly and repetitively, and engaged in meandering superficial conversations [21]. Sabanovic's research group at Indiana University found similar "quirkiness" with their I.R.I.S. system for the QT robot [19], which they have attempted to overcome with an intensive co-design process [20]. However, as of writing of this paper, those problems remain unresolved (S. Sabanovic, personal communication, April 5, 2024). Another recent preprint from Xygkou (2024) reported similar problems involving "superficial" conversations with an LLM-based chatbot that meandered from one topic to another without consistency, leading to shallow interactions that dementia patients found confusing [30]. Elsewhere, Addlesee et al. (2024) have reported multiple recent attempts to specially engineer solutions for conversational agents to address the challenges due to changed speech patterns (e.g. turn-taking), with partial success [1, 2]. However, the need to specially engineer solutions to design LLM-based conversational agents for dementia flies in the face of the general hype around LLM tools such as ChatGPT in healthcare and beyond [3]. Part of that hype was that manually creating things like dialogue state trackers and other components would no longer be necessary, that rather the LLM would just figure it all out itself using its "transformer magic", based on emergent statistical properties of the language. It could be argued that, if we still need all those other components, then LLMs essentially are just a new way to produce speech content, rather than an actual "speech system" in and of itself.

We stress here that many of those above reports are coming via pre-prints or as anecdotal comments at the end of larger papers, partially due to the reluctance of journals and high-impact conferences in regards to publishing *negative results*. That is extremely problematic, as it means those findings are not well-publicized and hard to get a full picture of. In fact, we discovered the same problems with LLMs as the above reports in our own research on conversational agents in 2023, and halted our study in order to revise our system, rather than attempt to publish the negative results.

Linking all of this back to Section 2, we note that the vast majority of the research being done in this domain (dementia speech systems) is being done in the English language, which is problematic for a number of reasons [17]. Indeed, it is well-established that one's native language directly affects how speech is encoded in the brain, e.g. English speakers are extremely sensitive to pitch changes but relatively insensitive to melodic/rhythm changes in speech whereas other language speakers are the exact opposite [13]. The idiosyncrasies of the English language thus may contribute to how speech dysfunction plays out in people living with dementia and Alzheimer's as their health condition progresses. For instance, a recent paper by Mazzeo et al. (2024) found that Italian speakers and English speakers with dementia manifest speech dysfunctions quite differently, positing those differences were directly tied to their native language [25]. A comprehensive review of further examples is given by Garcia et al. (2023), and we refer interested readers there [17]. We also note that similar reports of differences in semantic understanding loss are also being reported in other neurological disorders, such as Parkinson's Disease, leading to calls for studies of non-English populations in those disorders [28].

4 DISCUSSION

In summary, in this paper we address the common, but largely untested, assumption that experimental evidence based on English-only artificial speech systems represent some "universal truth" about how humans interact with such artificial agents. We present a slew of evidence from cross-language research on the topic suggesting that that assumption appears to be invalid. In short, there appear to be idiosyncrasies about the English language which are not representative of all languages [11]. That evidence includes experiments using artificial agents in multiple languages with the same human bilingual speaker, as well as studies across populations of human speakers of different languages. There is additional evidence coming from the field of cognitive science related to idiosyncrasies of the English language, along with research on differences in non-verbal social cues. Beyond that, we present evidence using the example of LLM-based artificial speech systems for people living with dementia, with various emerging reports of problems occurring from several different research groups world-wide causing those systems to fail in various ways. More concerning is the fact that many of those reports are only occurring anecdotally or buried at the end of larger papers, due to the reluctance journals and high-impact conferences to publish *negative results*. That means it is hard to get a good view of the full scope of the problem, and that many within the CUI field may not yet be aware of the problem.

The evidence presented throughout this paper challenges the notion that speech systems (including LLM-based ones) developed using English-only data and experiments represent a form of "intelligence" if, in fact, they only really represent the idiosyncrasies of one particular form of communication rather than some underlying universal principles. Indeed, it calls into question what we even mean by the word "intelligence", if such systems lack adaptability to other forms of communication. **Yet we would hope this paper serves not only as a *challenge* on that point to the research community, but also an opportunity.** Bilingual and/or multilingual research is certainly much harder to conduct than using a single language only, but there exists a rich area of research when doing so that may help us further illuminate the underpinnings of human intelligence (and subsequently AI) through the field of CUI.

Accomplishing the above may require some changes to how research is conducted in the field. For instance, it may be necessary to establish international research collaboration networks, as finding native speakers of different languages within a single country can often be challenging. Moreover, due to the influences of L2 language on L1 language in bilingual speakers (see Section 2), that would rule out bilingual speakers in many cases (e.g. fluent speakers of English and French in Canada) unless the study was specifically about the effects of bilingualism. For instance, currently we would need to, say, find a bunch of native speakers of French and Japanese in Canada, who did not speak the other language. That could be challenging, so a much easier approach may be to simply repeat the same experiment in different countries (e.g. having collaborators in Quebec and Japan), using the exact same platform and study design [12]. Doing so is, of course, difficult in its own way (e.g. logistics). The other hurdle is financial, as research funding agencies often limit themselves to a specific geographic region or single nation, and sharing research funding across international borders is commonly restricted in various ways by governments. As such, we contend that it is incumbent that such international research collaboration networks are established, and that the CUI field lobbies their national/regional research funding agencies to provide more funding for such research. **Doing so will allow us to develop a deeper understanding of human language and communication particularly as it pertains to CUI and artificial speech systems, which is ultimately the goal of the CUI field, is it not?**

As a final aside, we posit an interesting, though speculative, theory here related to the dementia speech systems discussed in Section 3. Namely, there are some striking parallels between the "changed speech" patterns in human individuals living with dementia and the problems of LLM-based conversational agents designed to interact with those individuals (difficulty staying on topic, turn-taking disruptions, etc.). Indeed, an intriguing hypothesis for future research is that the current generation of LLMs like ChatGPT essentially speak like dementia patients. If so, that may allow us to use LLM-based conversational agents as models of speech dysfunction in dementia and Alzheimer's, and the solutions developed to enhance the conversational fluency of such artificial agents (across multiple languages) may even help shed light on the underlying mechanics of how that speech dysfunction manifests in humans. That remains another exciting area for potential future research in the CUI field.

ACKNOWLEDGMENTS

This work draws from research across many different projects, but no grant specifically. We would like to thank S. Sabanovic and others for personal communication included in this paper. We would also like to thank our various research collaborators from over the years, who contributed to these ideas.

REFERENCES

[1] Angus Addlesee, Neeraj Cherakara, Nivan Nelson, Daniel Hernández Garcia, Nancie Gunson, Weronika Sieińska, Christian Dondrup, and Oliver Lemon. 2024. Multi-party multimodal conversations between patients, their companions, and a social robot in a hospital memory clinic. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. 62–70.

[2] Angus Addlesee and Arash Eshghi. 2024. You have interrupted me again!: making voice assistants more dementia-friendly with incremental clarification. *Frontiers in Dementia* 3 (2024), 1343052.

[3] Amir Masoud Afsahi, Seyed Ahmad Seyed Alinaghi, Ayoob Molla, Pegah Mirzapour, Shima Jahani, Armin Razi, Paniz Mojdeganlou, Elaheh Karimi, Mohammad Mehrtak, Omid Dadras, et al. 2024. Chatbots Utility in Healthcare Industry: An Umbrella Review. *Frontiers in Health Informatics* 13 (2024), 200.

[4] Sean Andrist, Micheline Ziadee, Halim Boukaram, Bilge Mutlu, and Majd Sakr. 2015. Effects of culture on the credibility of robot speech: A comparison between english and arabic. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*. 157–164.

[5] Mohammad Atari, Mona J Xue, Peter S Park, Damián Blasi, and Joseph Henrich. 2023. Which humans? (2023).

[6] Panos Athanasopoulos and Fraibet Aveledo. 2012. Linguistic relativity and bilingualism. *Memory, language, and bilingualism: Theoretical and applied approaches* (2012), 236–255.

[7] Panos Athanasopoulos and Aina Casaponsa. 2020. The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology* 37, 5-6 (2020), 393–412.

[8] Silva Banovic, Lejla Junuzovic Zunic, and Osman Sinanovic. 2018. Communication difficulties as a result of dementia. *Materia socio-medica* 30, 3 (2018), 221.

[9] Casey C. Bennett, Young-Ho Bae, Jun Hyung Yoon, Yejin Chae, Eunseo Yoon, Seeun Lee, Uijae Ryu, Say Young Kim, and Benjamin Weiss. 2023. Effects of Cross-Cultural Language Differences on Social Cognition During Human-Agent Interaction in Cooperative Game Environments. *Computer Speech & Language* 81 (2023), 101521.

[10] Casey C Bennett, Say Young Kim, Benjamin Weiss, Young-Ho Bae, Jun Hyung Yoon, Yejin Chae, Eunseo Yoon, Uijae Ryu, Hansae Cho, and Yesung Shin. 2023. Cognitive Shifts in Bilingual Speakers Affect Speech Interactions with Artificial Agents. *International Journal of Human–Computer Interaction* (2023), 1–12.

[11] Casey C Bennett and Minha Lee. 2023. Would People Mumble Rap to Alexa?. In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI)*. 1–5.

[12] Casey C Bennett, Selma Sabanovic, Cedomir Stanojevic, Zachary Henkel, Seongcheol Kim, Jinjae Lee, Kenna Baugus, Jennifer A Piatt, Janghoon Yu, Jiyeong Oh, et al. 2023. Enabling robotic pets to autonomously adapt their own behaviors to enhance therapeutic effects: a data-driven approach. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1625–1632.

[13] Damián E Blasi, Joseph Henrich, Evangelia Adamou, David Kemmerer, and Asifa Majid. 2022. Over-reliance on English hinders cognitive science. *Trends in cognitive sciences* 26, 12 (2022), 1153–1170.

[14] Peter T Daniels and David L Share. 2018. Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading* 22, 1 (2018), 101–116.

[15] Ö Ece Demir-Lira, Junko Kanero, Cansu Oranç, Sümeyye Koşkulu, Idil Franko, Tilbe Göksun, and Aylin C Küntay. 2020. L2 vocabulary teaching by social robots: The role of gestures and on-screen cues as scaffolds. In *Frontiers in Education*, Vol. 5. Frontiers Media SA, 599636.

[16] David M Eberhard, Gary Francis Simons, and Charles D Fenning. 2020. Ethnologue: Languages of the world. Online version: http://www. ethnologue. com.

[17] Adolfo M García, Jessica de Leon, Boon Lead Tee, Damián E Blasi, and Maria Luisa Gorno-Tempini. 2023. Speech and language markers of neurodegeneration: a call for global equity. *Brain* 146, 12 (2023), 4870–4879.

[18] Amandine Geraudie, Petronilla Battista, Adolfo M García, Isabel E Allen, Zachary A Miller, Maria Luisa Gorno-Tempini, and Maxime Montembeault. 2021. Speech and language impairments in behavioral variant frontotemporal dementia: a systematic review. *Neuroscience & Biobehavioral Reviews* 131 (2021), 1076–1095.

[19] Long-Jing Hsu, Waki Kamino, Weslie Khoo, Katherine Tsui, David Crandall, and Selma Šabanović. 2023. Working Together Toward ikigai: Co-Designing Robots That Can Help Us Achieve Meaning and Purpose in Life. *XRDS: Crossroads, The ACM Magazine for Students* 30, 1 (2023), 38–45.

[20] Long-Jing Hsu, Weslie Khoo, Peter Lenon Goshomi, Philip B Stafford, Manasi Swaminathan, Katherine M Tsui, David J Crandall, and Selma Sabanović. 2024. Is Now a Good Time? Opportune Moments for Interacting with an Ikigai Support Robot. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 549–553.

[21] Bahar Irfan, Sanna-Mari Kuoppamäki, and Gabriel Skantze. 2023. Between reality and delusion: challenges of applying large language models to companion robots for open-domain dialogues with older adults. *ResearchSquare PrePrint* (2023). https://doi.org/10.21203/rs.3.rs-2884789/v1

[22] Yanghee Kim, Sherry Marx, Hung Viet Pham, and Tung Nguyen. 2021. Designing for robot-mediated interaction among culturally and linguistically diverse children. *Educational Technology Research and Development* 69 (2021), 3233–3254.

[23] Velvetina Lim, Maki Rooksby, and Emily S Cross. 2021. Social robots on a global stage: establishing a role for culture during human–robot interaction. *International Journal of Social Robotics* 13, 6 (2021), 1307–1333.

[24] Jingxuan Liu, Courtney B Hilton, Elika Bergelson, and Samuel A Mehr. 2023. Language experience predicts music processing in a half-million speakers of fifty-four languages. *Current Biology* 33, 10 (2023), 1916–1925.

[25] Salvatore Mazzeo, Chris Hardy, Jessica Jiang, Carmen Morinelli, Valentina Moschini, Ella Brooks, Jeremy CS Johnson, Anthipa Chokesuwattanaskul, Anna Volkmer, Jonathan D Rohrer, et al. 2024. Primary progressive aphasia in Italian and English: a cross-linguistic cohort study. *medRxiv* (2024), 2024–03.

[26] Ognjen Rudovic, Jaeryoung Lee, Lea Mascarell-Maricic, Björn W Schuller, and Rosalind W Picard. 2017. Measuring engagement in robot-assisted autism therapy: A cross-cultural study. *Frontiers in Robotics and AI* 4 (2017), 36.

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

[28] Felipe Diego Toro-Hernández, Joaquín Migeot, Nicolás Marchant, Daniela Olivares, Franco Ferrante, Raúl González-Gómez, Cecilia González Campo, Sol Fittipaldi, Gonzalo M Rojas-Costa, Sebastian Moguilner, et al. 2024. Neurocognitive correlates of semantic memory navigation in Parkinson's disease. *npj Parkinson's Disease* 10, 1 (2024), 15.

[29] Sebastian Wallkötter, Silvia Tulli, Ginevra Castellano, Ana Paiva, and Mohamed Chetouani. 2021. Explainable embodied agents through social cues: a review. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 3 (2021), 1–24.

[30] Anna Xygkou, Ang Chee Siang, Panote Siriaraya, Jonasz Kopecki, Alexandra Covaci, Eiman Kanjo, and She Wan-Jou. 2024. MindTalker: Navigating the Complexities of AI-Enhanced Social Engagement for People with Early-Stage Dementia. *ACM conference on Human Factors in Computing Systems (CHI)* (2024), In Press. https://kar.kent.ac.uk/105061/.

[31] Yasunori Yamada, Kaoru Shinkawa, Miyuki Nemoto, Miho Ota, Kiyotaka Nemoto, and Tetsuaki Arai. 2022. Speech and language characteristics differentiate Alzheimer's disease and dementia with Lewy bodies. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 14, 1 (2022), e12364.