A special interest group on developing theories of language use in interaction with conversational user interfaces

A Special Interest Group on Developing Theories of Language Use in Interaction with Conversational User Interfaces

Paola Raquel Peña Huerta
paola.pena@ucdcoonect.ie
University College Dublin
Dublin 4, Ireland

Philip R. Doyle
philip.doyle1@ucdconnect.ie
University College Dublin
Dublin 4, Ireland

Emily Y.J. Ip
ipem@tcd.ie
Trinity College Dublin
Dublin, Ireland

Giovanni M. Di Liberto
gdiliber@tcd.ie
Trinity College Dublin
Dublin, Ireland

Darragh Higgins
higgind3@tcd.ie
Trinity College Dublin
Dublin, Ireland

Rachel McDonnell
RAMCDONN@tcd.ie
Trinity College Dublin
Dublin, Ireland

Holly Branigan
University of Edinburgh
Edinburgh, United Kingdom
Holly.Branigan@ed.ac.uk

Joakim Gustafson
KTH Royal Institute of Technology in Stockholm
Stockholm, Sweden
jkgu@kth.se

Donald McMillan
Stockholm University
Stockholm, Sweden
donald.mcmillan@dsv.su.se

Robert J. Moore
IBM Research-Almaden Lab
San Jose, United States of America
rjmoore@us.ibm.com

Benjamin R. Cowan
bejamin.cowan@ucd.ie
University College Dublin
Dublin 4, Ireland

ABSTRACT

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models**; **Interaction design theory, concepts and paradigms**; • **Applied computing** → **Psychology**.

KEYWORDS

conversational user interfaces, human-machine dialogue, speech agents, psycholinguistic models

ACM Reference Format:
Paola Raquel Peña Huerta, Philip R. Doyle, Emily Y.J. Ip, Giovanni M. Di Liberto, Darragh Higgins, Rachel McDonnell, Holly Branigan, Joakim Gustafson, Donald McMillan, Robert J. Moore, and Benjamin R. Cowan. 2023. A Special Interest Group on Developing Theories of Language Use in Interaction with Conversational User Interfaces. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '23 Extended Abstracts).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3544549.3583179

1 INTRODUCTION

Through the increased growth of speech agents, text based chatbots and social robots, language interactions with machine dialogue partners are now commonplace. Discovering what drives the way we converse with machines is fundamental to understanding our interaction with such automated dialogue partners. However, understanding of what governs user language choices in such human-machine dialogues (HMD) is sparse [14]. This Special Interest Group aims to be a catalyst for discussing and building fundamental theories of how people produce language when engaged in conversation with conversational user interfaces (CUIs). The main objective is to bring together researchers across CHI and related communities (e.g. HRI, CUI, cognitive science, linguistics and speech technology) to map the grand challenges required to be addressed to generate evidence-based theories to explain what impacts our linguistic interactions with CUIs.

2 NEED FOR A SIG

Theory work on language production in CUI interaction is scarce [15]. Current approaches rely on applying existing theoretical accounts of human-human dialogue (HHD) and language use from disciplines such as psycholinguistics and sociolinguistics to help understand mechanisms that lead users to adopt specific language choices and behaviours. Using these accounts as a foundation is useful as it gives us initial methods and concepts that can be applied to CUI interaction. Recent efforts to understand user language production in CUI have taken this approach, directly applying methods and concepts from such disciplines [7, 18, 24]. Yet such methods and concepts also need to be tailored so as to be sensitive to the fundamental differences between human and machine dialogue interaction in terms of partner capabilities, the nature and aims of the dialogue [15, 45].

This SIG will outline 1) the existing approaches and theoretical accounts currently adopted from social science disciplines as well as devise 2) a roadmap of work that needs to be undertaken to build CUI centered theories, sensitive to the nuances of HMD and CUI

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3583179

interaction. In the following sections, we outline some example theoretical and methodological frameworks employed in HHD that have been influential in recent CUI interaction research. These are by no means exhaustive. We propose to focus on these frameworks as a basis for our discussion, mapping out how these approaches (and other identified during the SIG) can be applied to generate more cohesive theoretical frameworks focused on linguistic interactions with CUIs more specifically.

3 EXAMPLES OF RELEVANT THEORETICAL AND METHODOLOGICAL FRAMEWORKS

3.1 Audience Design, Common Ground and Partner modelling

Research in HHD proposes that people engage in audience design [4] when speakers adapt their utterances based on perceptions of their addressees' knowledge, dialogue capabilities and beliefs [11, 12]. The concept of common ground is critical to facilitating audience design, in that interlocutors use contextual information, information or descriptions co-constructed during dialogue along with presumed knowledge of their partner to plan utterances so as to optimally design them for their audience [10]. Other HHD accounts such as the monitor and adjust model [3] state that, rather than audience design being the sole driver of language choices, speakers also rely on their own knowledge for utterance construction, only using their partner's perspective when absolutely necessary [26, 27, 47]. Both of these have been applied in CUI research. A significant amount of work has demonstrated that audience design (also termed recipient design [2]) occurs in HMD, whereby language choices are adaptive and considered based on preconceptions of a machine partner's abilities and knowledge [5, 8, 29, 32, 34]. Machines tend to be seen as "at risk" listeners [42] in conversation, which leads people to adapt language choices to ensure that communication is successful [7] echoing HHD audience design accounts [4, 10, 13]. Indeed CUI design is also likely important in this regard, with aspects such as agent accent or perceived nationality [18, 28, 46] may act as a cue for informing people's perceptions of partner capabilities (recently termed partner models [24] in HMD research) that guide language use. Yet, recent studies have questioned the ubiquity of such an account, either finding no audience design effects [16–18] or observing more egocentric production [23].

3.2 Computers are Social Actors

The Computers Are Social Actors (CASA) paradigm asserts that people *mindlessly* apply social heuristics from human-human interaction in human-computer interaction, leading them to respond to computer systems as they respond to other people. This assertion is founded on studies that showed users applying gender stereotypes to male- and female-voiced computers [39], using social categorisation to inform behaviours and judgements of computers in collaborative tasks [36], being polite to a computer when asked to evaluate it [38], being more helpful to computers that were perceived as more helpful [25], and applying expertise judgements to televisions [37]. The earliest experiments were on computer systems with voice output [40], but this was later generalised to other types of interfaces and devices. This account has been highly influential in the field of HCI, and gives an important potential mechanism which may govern the construction of initial partner models of CUIs as dialogue partners. That is, users may use cues from the interaction to inform the use of social heuristics when in dialogue with CUIs.

3.3 Neurophysiological approaches of CUI Interaction

The study of human communication in neurophysiology entails probing underlying brain mechanisms for metrics that can validate existing psycholinguistic models. Although initially focusing on simple speech listening tasks far from dialogue scenarios (e.g., listening to isolated syllables), the field has undergone a rapid development of naturalistic methodologies in research [20, 22, 31], propelled by the discovery that neural signals track acoustic and linguistic features of a speech input [19]. This robust relationship between specific input features and corresponding neural signal [41] is thought to open up new potential to study the processing of various speech constituents simultaneously in ecologically-valid settings. Such naturalistic paradigms support predictive processing, a long-standing theoretical framework [9, 33] whereby sensory information coupled with the brain's predictions of upcoming sensory events highlights the "active" nature of perception. These neural architectures can elucidate how humans contextualise speech sounds in dialogue scenarios, informing both HMD and HHD. Indeed new hyperscanning technology has brought us closer to researching more realistic dialogue scenarios enabling the simultaneous recording of EEG signals from multiple participants. This line of work has already demonstrated the possibility of measuring the impact of social cues (e.g., reflecting attention) on neural synchrony across participants [6, 21]. Although dialogue presents challenges for neurophysiological study, extracting objective indices of brain activity during dialogue in interactive scenarios [44] may constitute promising foundations for new insights into CUI based dialogue interaction.

3.4 Naturalistic observation of CUI Language Use

Along with controlled experiment approaches to understanding language production in CUI interaction (e.g. [7, 18] work on understanding language production also uses more naturalistic approaches through the statistical and conversation analysis of real world human-machine dialogues [30, 45]. Conversation analysis in particular has gained significant traction as a practice in CUI research [1, 35, 43, 45], being used to identify and interpret people's language choices in CUI interaction. Rather than focusing on the frequency of language-based phenomena, as is more common in statistical and controlled experiment approaches, CA gives a rich, in-depth exploration of linguistic effects and behaviours in dialogue. Such approaches and findings, in tandem with controlled and statistical approaches are likely to be integral to the development of fundamental theories of language production in CUI contexts.

4 OPEN QUESTIONS

The SIG will aim to use the outlined theories and methodological approaches as a starting point to discuss how as a community we can develop models and accounts that interpret and explain the mechanisms within language production in CUI interaction. The questions below have been identified as key ones to address as part of the SIG

- What theoretical lenses are the most appropriate to anchor attempts to devise theoretical insight to language use in CUI interaction?
- What are the key methodological approaches and challenges in ensuring the viability and robustness of theoretical frameworks for language production in CUI interaction and how can these work together?
- What concepts are critical to devising theories in this area?
- What aspects about CUI interaction need to be considered when adapting theories and concepts from other disciplines?
- What are some of the key unknowns around these concepts that need significant work?

REFERENCES

[1] Saul Albert and Magnus Hamann. 2021. Putting wake words to bed: We speak wake words with systematically varied prosody, but CUIs don't listen. In *CUI 2021-3rd Conference on Conversational User Interfaces*. 1–5.

[2] Sungeun An, Robert Moore, Eric Young Liu, and Guang-Jie Ren. 2021. Recipient Design for Conversational Agents: Tailoring Agent's Utterance to User's Knowledge. In *CUI 2021-3rd Conference on Conversational User Interfaces*. 1–5.

[3] Dale J Barr and Boaz Keysar. 2002. Anchoring comprehension in linguistic precedents. *Journal of Memory and Language* 46, 2 (2002), 391–418.

[4] Allan Bell. 1984. Language style as audience design. *Language in society* 13, 2 (1984), 145–204.

[5] Linda Bell and Joakim Gustafson. 2000. Positive and negative user feedback in a spoken dialogue corpus. In *nternational Conference on Spoken Language Processing*.

[6] Dana Bevilacqua, Ido Davidesco, Lu Wan, Kim Chaloner, Jess Rowland, Mingzhou Ding, David Poeppel, and Suzanne Dikker. 2019. Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: Evidence from a real-world classroom electroencephalography study. *Journal of cognitive neuroscience* 31, 3 (2019), 401–411.

[7] Holly P Branigan, Martin J Pickering, Jamie Pearson, Janet F McLean, and Ash Brown. 2011. The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition* 121, 1 (2011), 41–57.

[8] Susan E Brennan. 1998. The grounding problem in conversations with and through computers. *Social and cognitive approaches to interpersonal communication* (1998), 201–225.

[9] Andy Clark. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences* 36, 3 (2013), 181–204.

[10] Herbert H Clark. 1996. *Using language*. Cambridge university press.

[11] Herbert H Clark. 2020. Common ground. *The International Encyclopedia of Linguistic Anthropology* (2020), 1–5.

[12] Herbert H Clark. 2021. Anchoring utterances. *Topics in Cognitive Science* 13, 2 (2021), 329–350.

[13] Herbert H Clark and Susan E Brennan. 1991. Grounding in communication. (1991).

[14] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. *Interacting with Computers* 31, 4 (2019), 349–371.

[15] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[16] Benjamin R Cowan and Holly P Branigan. 2015. Does voice anthropomorphism affect lexical alignment in speech-based human-computer dialogue?. In *Sixteenth Annual Conference of the International Speech Communication Association*.

[17] Benjamin R Cowan, Holly P Branigan, Mateo Obregón, Enas Bugis, and Russell Beale. 2015. Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human- computer dialogue. *International Journal of Human-Computer Studies* 83 (2015), 27–42.

[18] Benjamin R Cowan, Philip Doyle, Justin Edwards, Diego Garaialde, Ali Hayes-Brady, Holly P Branigan, João Cabral, and Leigh Clark. 2019. What's in an accent? The impact of accented synthetic speech on lexical choice in human-machine dialogue. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–8.

[19] Giovanni M Di Liberto, Jens Hjortkjaer, and Nima Mesgarani. 2022. Neural Tracking. Closing the Gap Between Neurophysiology and Translational Medicine. *Frontiers in Neuroscience* (2022), 307.

[20] Giovanni M Di Liberto, Varghese Peter, Marina Kalashnikova, Usha Goswami, Denis Burnham, and Edmund C Lalor. 2018. Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *Neuroimage* 175 (2018), 70–79.

[21] Suzanne Dikker, Lu Wan, Ido Davidesco, Lisa Kaggen, Matthias Oostrik, James McClintock, Jess Rowland, Georgios Michalareas, Jay J Van Bavel, Mingzhou Ding, et al. 2017. Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current biology* 27, 9 (2017), 1375–1380.

[22] Nai Ding, Monita Chatterjee, and Jonathan Z Simon. 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88 (2014), 41–46.

[23] Judit Dombi, Tetyana Sydorenko, and Veronika Timpe-Laughlin. 2022. Common ground, cooperation, and recipient design in human-computer interactions. *Journal of Pragmatics* 193 (2022), 4–20.

[24] Philip R Doyle, Leigh Clark, and Benjamin R Cowan. 2021. What do we see in them? identifying dimensions of partner models for speech interfaces using a psycholexical approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[25] BJ Fogg and Clifford Nass. 1997. How users reciprocate to computers: an experiment that demonstrates behavior change. In *CHI'97 extended abstracts on Human factors in computing systems*. 331–332.

[26] Boaz Keysar. 2007. Communication and miscommunication: The role of egocentric processes. (2007).

[27] Boaz Keysar, Shuhong Lin, and Dale J Barr. 2003. Limits on theory of mind use in adults. *Cognition* 89, 1 (2003), 25–41.

[28] Sara Kiesler. 2005. Fostering common ground in human-robot interaction. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* IEEE, 729–734.

[29] Dimosthenis Kontogiorgos, Andre Pereira, and Joakim Gustafson. 2021. Grounding behaviours with conversational interfaces: effects of embodiment and failures. *Journal on Multimodal User Interfaces* 15, 2 (2021), 239–254.

[30] Dimosthenis Kontogiorgos, Sanne van Waveren, Olle Wallberg, Andre Pereira, Iolanda Leite, and Joakim Gustafson. 2020. Embodiment effects in interactions with failing robots. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.

[31] Edmund C Lalor and John J Foxe. 2010. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European journal of neuroscience* 31, 1 (2010), 189–193.

[32] Ludovic Le Bigot, Jean-François Rouet, and Eric Jamet. 2007. Effects of speech- and text-based interaction modes in natural language human-computer dialogue. *Human Factors* 49, 6 (2007), 1045–1053.

[33] Gary Lupyan and Andy Clark. 2015. Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science* 24, 4 (2015), 279–284.

[34] Elizabeth J Meddeb and Patricia Frenz-Belkin. 2010. What? I Didn't Say THAT!: Linguistic strategies when speaking to write. *Journal of Pragmatics* 42, 9 (2010), 2415–2429.

[35] Robert J Moore, Margaret H Szymanski, Raphael Arar, and Guang-Jie Ren. 2018. Studies in Conversational UX Design. (2018).

[36] Clifford Nass, BJ Fogg, and Youngme Moon. 1996. Can computers be teammates? *International Journal of Human-Computer Studies* 45, 6 (1996), 669–678.

[37] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56, 1 (2000), 81–103.

[38] Clifford Nass, Youngme Moon, and Paul Carney. 1999. Are people polite to computers? Responses to computer-based interviewing systems 1. *Journal of applied social psychology* 29, 5 (1999), 1093–1109.

[39] Clifford Nass, Youngme Moon, and Nancy Green. 1997. Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of applied social psychology* 27, 10 (1997), 864–876.

[40] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 72–78.

[41] Jonas Obleser and Christoph Kayser. 2019. Neural entrainment and attentional selection in the listening brain. *Trends in cognitive sciences* 23, 11 (2019), 913–926.

[42] Sharon Oviatt. 2022. Multimodal Interaction, Interfaces, and Analytics. In *Handbook of Human Computer Interaction*. Springer, 1–29.

[43] Hannah RM Pelikan and Mathias Broth. 2016. Why that nao? how humans adapt to a conventional humanoid robot in taking turns-at-talk. In *Proceedings of the*

2016 CHI conference on human factors in computing systems. 4921–4932.

[44] Alejandro Pérez, Manuel Carreiras, and Jon Andoni Duñabeitia. 2017. Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Scientific reports* 7, 1 (2017), 1–12.

[45] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[46] Aaron Powers, Adam DI Kramer, Shirlene Lim, Jean Kuo, Sau-lai Lee, and Sara Kiesler. 2005. Eliciting information from people with a gendered humanoid robot. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* IEEE, 158–163.

[47] Hadas Shintel and Boaz Keysar. 2009. Less is more: A minimalist account of joint action in communication. *Topics in Cognitive Science* 1, 2 (2009), 260–273.

Received 15 December 2022