Dialogue & Discourse 16(3) (2025) 1–7 doi:10.5210/dad.2025.301

Embodied Conversational Systems in Human–Robot Interaction: Introduction to the Special Issue

Dimitra Gkatzia* D.GKATZIA@NAPIER.AC.UK
Edinburgh Napier University, Edinburgh, UK

Hendrik Buschmeier* HBUSCHME@UNI-BIELEFELD.DE
Bielefeld University, Bielefeld, Germany

Mary Ellen Foster MARYELLEN.FOSTER@GLASGOW.AC.UK
University of Glasgow, Glasgow, UK

Carl Strathearn C.STRATHEARN@NAPIER.AC.UK
Edinburgh Napier University, Edinburgh, UK

Editor: Barbara Di Eugenio, David Traum

1. Introduction

In recent years, conversational systems such as chatbots and virtual assistants have become increasingly popular. The underlying technology has the potential to enhance human–robot interaction (HRI; Bartneck et al., 2020) and improve its user experience. However, designing and implementing effective conversational systems for HRI presents significant challenges that need to be addressed (cf. Devillers et al., 2020; Marge et al., 2022). This special issue of Dialogue & Discourse on "Embodied Conversational Systems in Human–Robot Interaction", brings together researchers and practitioners to explore the opportunities and challenges of developing conversational systems for HRI.

Conversational systems and natural language generation (NLG; Reiter and Dale, 2000; Reiter, 2025) are central to human–robot interaction, enabling natural and intuitive speech-based communication. Advances in these areas, as well as in related fields such as multimodal interaction, can make robots more accessible, usable, and engaging in domains such as healthcare, education, services, assistive living, and entertainment. By integrating speech, facial expressions, and other non-verbal cues, conversational systems allow robots to better infer users' emotions and social signals and to tailor their responses accordingly. This adaptability enables personalized interactions that reflect individuals' needs, preferences, and characteristics, resulting in more meaningful and natural exchanges. This capability is especially valuable in contexts such as personalized tutoring or explanation-giving (Stange et al., 2022), where effective communication depends on sensitivity to the user. Conversational systems provide a foundation for these capabilities by enabling natural language interaction, an intuitive and familiar means of communication for humans.

Human–robot interaction is a complex, interdisciplinary field that requires expertise in robotics, artificial intelligence, (computational) linguistics, psychology and human factors, among others. Conversational systems integrate many of these areas as well, representing a challenging and ever evolving area of research that has the potential to advance HRI technology. This special issue brings

*. Equal contribution.

together novel research in dialogue systems designed to enhance or support the interaction with robots. A key objective within the active research domain of HRI is to develop robotic agents capable of emulating socially intelligent behavior when interacting with humans. Despite the clear relationship between social intelligence and fluent, flexible linguistic interaction, interactive robots have only recently begun to utilize anything beyond a basic dialogue manager and template-based response generation process in practice (van Deemter et al., 2005). This means that, thus far, social robot systems have been unable to exploit the flexibility offered by dialogue systems and natural language generation when managing conversations between humans and robots in dynamic environments, or when the conversation needs to adapt to different contexts or multiple target languages. Conversely, end-to-end systems based on large language models (LLMs) enable robots to communicate; however, they must be integrated into the overall robotics architecture in order to take into account the situated, embodied nature of robots (Lison and Kennington, 2023).

2. Building bridges between NLG, HRI and dialogue

These issues have been explored in a series of workshops organized by the editors of this special issue: the first Workshop on Natural Language Generation for Human Robot Interaction ("NLG for HRI"; Foster et al., 2018) took place at the 2018 International Natural Language Generation Conference (INLG 2018); conversely, the second workshop in the series (Buschmeier et al., 2020) was meant to take place in March 2020 at the ACM/IEEE International Conference on Human–Robot Interaction (HRI 2020), but had to be postponed to INLG 2020 due to the emergence of the COVID-19 pandemic; finally, a special session "Natural Language in Human Robot Interaction" (NLiHRI) took place at the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial 2022). A wide range of topics were discussed at the workshops, but most of the papers presented dealt with one or more of four central themes: multimodal generation, (visual) grounding and contextual knowledge, robust interaction, and adaptation.

Robots are embodied agents with the ability to move. Depending on their physical configuration, they may be able to move their actuators, sensors or even their entire body. These movements must be planned for action and perception; however, each movement also has the potential to be interpreted as a non-verbal communication behaviour. These non-verbal behaviors need to be planned alongside (or even integrated with) the robot's non-communicative actions, as well as its verbal acts. This makes questions of **multimodal generation** central to the intersection of NLG, dialogue research, and human–robot interaction and thus a topic discussed across all three workshops (e.g., Cass et al., 2018; Bailly and Elisei, 2020; Gella et al., 2022).

Robots are situated in real-world environments and may share space with people. Therefore, a robot needs to be able to perceive its environment, including objects and people, and ground its language in the environment. It also needs to be able to reason about and refer to the environment using verbal and non-verbal means. **(Visual) grounding and contextual knowledge** is central for language generation in robots and thus discussed across all workshops (e.g., Pomarlan et al., 2018; Wallbridge et al., 2020; Torres-Fonseca et al., 2022).

Although robot actions and human–robot interaction are often much slower than human action and interaction, they share many of the pressures present in human face-to-face communication. These include the timing of turn-taking, miscommunication, conversational failure and repair. This requires a certain level of robustness in the robot's interaction capabilities, making **robust interaction** a topic at each workshop (e.g., Zarrieß and Schlangen, 2018; Doğan and Leite, 2020; Li et al., 2022).

Finally, several contributions across the workshops argued that robots interacting with human conversation partners should **adapt** to their characteristics (e.g., personality), emotional states (as expressed, e.g., through social signals), and needs (e.g., Ritschel and André, 2018; Shenoy and Dugan, 2020; Fernau et al., 2022).

3. Overview of the papers in the special issue

Continuing the discussion of these central themes of the workshops, this special issue comprises four papers that cover empirical, computational modeling, and engineering perspectives on embodied conversational systems in human–robot interaction. Topics that recur throughout the papers include multimodality and conversational behavior, the representation of dialogue state using knowledge graphs, and the integration of language models into conversationally capable robot architectures.

The first paper **"Laughter use by virtual agents increases task success"** (Ludusan and Wagner, 2025) studies the influence of synthetic sound-based nonverbal expressions in human–agent interaction – specifically the use of laughter in dialogue – on participants' agent social perception and, importantly, task success. It shows that an agent that laughs is rated higher in its social perception and has a higher task success. This is novel evidence for the discussions earlier in the workshops which dealt with the importance of modeling humor and laughter capabilities for interactive robots (Ritschel and André, 2018; Ritschel et al., 2020). The paper also shows that being able to technically integrate para-verbal behavior with the verbal behavior of these systems (both on the level of planning as well as on the level of synthesis) is useful and effective (Aylett, 2020).

Two contributions in this issue make use of knowledge graphs to improve the interaction capabilities in interactive robots (a line of research that was already visible in workshop contributions that made use of ontologies to link and ground a robot's physical and conversational actions, e.g., Pomarlan et al., 2018, 2020).

The contribution **"A modular architecture for creating multimodal embodied agents with an episodic knowledge graph as an explainable and controllable long-term memory"** (Baier et al., 2025) uses an 'episodic' knowledge graph in order to create coherence and continuity across interactions. The article presents a broader architectural framework for interactive agents consisting of two core components: the aforementioned episodic knowledge graph, and a component for the time-aligned management of multimodal signals encountered (and produced) by an agent during an interaction. These core components must be supplemented by elements that interpret and annotate signals, make decisions and generate behavior. The overarching goal of the architectural framework is to combine flexibility with control exerted through modeling higher-level intentions representing an agent's goals.

In contrast to this, the contribution **"A graph-to-text approach to knowledge-grounded response generation in human–robot interaction"** (Walker et al., 2025) proposes and evaluates a conversation model for a robot which represents the dialogue state in a graph-based representation. This knowledge graph combines linguistic with situated and multimodal information and is continuously updated from sensors of the robot as well as other system information, preserving temporal as well as probabilistic aspects. This representation is then used for generating an intermediate textual representation, which forms the basis for the generation of the robot's conversational actions using large language models.

Finally, the contribution **Prior lessons of incremental dialogue and robot action management for the age of language models** (Kennington et al., 2025), addresses the important topic of incremen-

tal processing in dialogue (see also Dialogue & Discourse, Vol. 2 No. 1; Rieser and Schlangen, 2011) and analyses its implications for the "age of LMs". The authors argue that incremental dialogue processing, particularly dialogue management, is essential for human–robot interaction (a point also been made in several of the workshop contributions: Zarrieß and Schlangen, 2018; Bailly and Elisei, 2020; Li et al., 2022). They argue that this presents challenges for systems in which dialogue capabilities are primarily driven by LLMs. The contribution introduces incremental dialogue processing, reviews the state of the art in incremental dialogue management, and discusses challenges and requirements.

This special issue presents contributions written when LLMs first emerged as a novel technological development and were swiftly incorporated into robotics and human–robot interaction. The contributions reflect this technological shift by utilizing LLMs and/or discussing how they can be integrated in light of the well-understood theoretical and engineering challenges at the intersection of conversational systems and human–robot interaction.

References

Matthew P. Aylett. Mixing speech and semantic free utterances: A challenge for natural language generation. In Hendrik Buschmeier, Mary Ellen Foster, and Dimitra Gkatzia, editors, *2nd Workshop on Natural Language Generation for Human–Robot Interaction*, Online, 2020. URL https://purl.org/nlghri2020/aylett.

Thomas Baier, Selene Báez Santamaría, and Piek Vossen. A modular architecture for creating multimodal embodied agents with an episodic knowledge graph as an explainable and controllable long-term memory. *Dialogue & Discourse*, 16(3):25–59, 2025. doi:10.5210/dad.2025.303.

Gérard Bailly and Frédéric Elisei. Speech in action: Designing challenges that require incremental processing of self and others' speech and performative gestures. In Hendrik Buschmeier, Mary Ellen Foster, and Dimitra Gkatzia, editors, *2nd Workshop on Natural Language Generation for Human–Robot Interaction*, Online, 2020. URL https://purl.org/nlghri2020/bailly.

Christoph Bartneck, Tony Belpaeme, Friederike Eyssel, Takayuki Kanda, Merel Keijsers, and Selma Šabanović. *Human-Robot Interaction: An Introduction*. Cambridge University Press, Cambridge, UK, 2020. doi:10.1017/9781108676649.

Hendrik Buschmeier, Mary Ellen Foster, and Dimitra Gkatzia. Second workshop on natural language generation for human–robot interaction. In *Companion of the 2020 ACM/IEEE International Conference on Human–Robot Interaction*, pages 646–647, Cambride, UK, 2020. Association for Computing Machinery. doi:10.1145/3371382.3374853.

Aaron G. Cass, Kristina Striegnitz, and Nick Webb. A farewell to arms: Non-verbal communication for non-humanoid robots. In Mary Ellen Foster, Hendrik Buschmeier, and Dimitra Gkatzia, editors, *Proceedings of the Workshop on NLG for Human–Robot Interaction*, pages 22–26, Tilburg, The Netherlands, 2018. Association for Computational Linguistics. doi:10.18653/v1/W18-6905.

Kees van Deemter, Emiel Krahmer, and Mariët Theune. Real versus template-based natural language generation: A false opposition? *Computational Linguistics*, 31:15–23, 2005. doi:10.1162/0891201053630291.

Laurence Devillers, Tatsuya Kawahara, Roger K. Moore, and Matthias Scheutz. Spoken language interaction with virtual agents and robots (SLIVAR): Towards effective and ethical interaction (Dagstuhl Seminar 20021). *Dagstuhl Reports*, 10(1):1–51, 2020. doi:10.4230/DagRep.10.1.1.

Fethiye Irmak Doğan and Iolanda Leite. Open challenges on generating referring expressions for human–robot interaction. In Hendrik Buschmeier, Mary Ellen Foster, and Dimitra Gkatzia, editors, *2nd Workshop on Natural Language Generation for Human–Robot Interaction*, Online, 2020. doi:10.48550/arXiv.2104.09193.

Daniel Fernau, Stefan Hillmann, Nils Feldhus, Tim Polzehl, and Sebastian Möller. Towards personality-aware chatbots. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 135–145, Edinburgh, UK, 2022. Association for Computational Linguistics. doi:10.18653/v1/2022.sigdial-1.15.

Mary Ellen Foster, Hendrik Buschmeier, and Dimitra Gkatzia, editors. *Proceedings of the Workshop on NLG for Human–Robot Interaction*, Tilburg, The Netherlands, 2018. Association for Computational Linguistics. doi:10.18653/v1/W18-69.

Spandana Gella, Aishwarya Padmakumar, Patrick Lange, and Dilek Hakkani-Tur. Dialog acts for task driven embodied agents. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 111–123, Edinburgh, UK, 2022. Association for Computational Linguistics. doi:10.18653/v1/2022.sigdial-1.13.

Casey Kennington, Pierre Lison, and David Schlangen. Prior lessons of incremental dialogue and robot action management for the age of language models. *Dialogue & Discourse*, 16(3):96–130, 2025. doi:10.5210/dad.2025.305.

Siyan Li, Ashwin Paranjape, and Christopher Manning. When can I speak? Predicting initiation points for spoken dialogue agents. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 217–224, Edinburgh, UK, 2022. Association for Computational Linguistics. doi:10.18653/v1/2022.sigdial-1.22.

Pierre Lison and Casey Kennington. Who's in charge? Roles and responsibilities of decision-making components in conversational robots. In *Workshop on Human–Robot Conversational Interaction*, Stockholm, Sweden, 2023. doi:10.48550/arXiv.2303.08470.

Bogdan Ludusan and Petra Wagner. Laughter use by virtual agents increases task success. *Dialogue & Discourse*, 16(3):8–24, 2025. doi:10.5210/dad.2025.302.

Matthew Marge, Carol Espy-Wilson, Nigel G. Ward, Abeer Alwan, Yoav Artzi, Mohit Bansal, Gil Blankenship, Joyce Chai, Hal Daumé, Debadeepta Dey, Mary Harper, Thomas Howard, Casey Kennington, Ivana Kruijff-Korbayová, Dinesh Manocha, Cynthia Matuszek, Ross Mead, Raymond Mooney, Roger K. Moore, Mari Ostendorf, Heather Pon-Barry, Alexander I. Rudnicky, Matthias Scheutz, Robert St. Amant, Tong Sun, Stefanie Tellex, David Traum, and Zhou Yu. Spoken language interaction with robots: Recommendations for future research. *Computer Speech & Language*, 71:101255, 2022. doi:10.1016/j.csl.2021.101255.

Mihai Pomarlan, Robert Porzel, John Bateman, and Rainer Malaka. From sensors to sense: Integrated heterogeneous ontologies for natural language generation. In Mary Ellen Foster, Hendrik

Buschmeier, and Dimitra Gkatzia, editors, *Proceedings of the Workshop on NLG for Human–Robot Interaction*, pages 17–21, Tilburg, The Netherlands, 2018. Association for Computational Linguistics. doi:10.18653/v1/W18-6904.

Mihai Pomarlan, Vanja Sophie Cangalovic, Robert Porzel, and John Bateman. Human, we have a problem: What to say when things go wrong. In Hendrik Buschmeier, Mary Ellen Foster, and Dimitra Gkatzia, editors, *2nd Workshop on Natural Language Generation for Human–Robot Interaction*, Online, 2020. URL https://purl.org/nlghri2020/pomarlan.

Ehud Reiter. *Natural Language Generation*. Springer, Cham, Switzerland, 2025. doi:10.1007/978-3-031-68582-8.

Ehud Reiter and Robert Dale. *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge, UK, 2000. doi:10.1017/CBO9780511519857.

Hannes Rieser and David Schlangen. Introduction to the special issue on incremental processing in dialogue. *Dialogue & Discourse*, 2(1):1–10, 2011. doi:10.5087/dad.2011.001.

Hannes Ritschel and Elisabeth André. Shaping a social robot's humor with natural language generation and socially-aware reinforcement learning. In Mary Ellen Foster, Hendrik Buschmeier, and Dimitra Gkatzia, editors, *Proceedings of the Workshop on NLG for Human–Robot Interaction*, pages 12–16, Tilburg, The Netherlands, 2018. Association for Computational Linguistics. doi:10.18653/v1/W18-6903.

Hannes Ritschel, Thomas Kiderle, Klaus Weber, and Elisabeth André. Multimodal joke presentation for social robots based on natural-language generation and nonverbal behaviors. In Hendrik Buschmeier, Mary Ellen Foster, and Dimitra Gkatzia, editors, *2nd Workshop on Natural Language Generation for Human–Robot Interaction*, Online, 2020. URL https://purl.org/nlghri2020/ritschel.

Sudhir Shenoy and Joanne Dugan. Challenges and opportunities for NLG in persuasive robotics. In Hendrik Buschmeier, Mary Ellen Foster, and Dimitra Gkatzia, editors, *2nd Workshop on Natural Language Generation for Human–Robot Interaction*, Online, 2020. URL https://purl.org/nlghri2020/shenoy.

Sonja Stange, Teena Hassan, Florian Schröder, Jacqueline Konkol, and Stefan Kopp. Self-explaining social robots: An explainable behavior generation architecture for human-robot interaction. *Frontiers in Artificial Intelligence*, 5(866920):1–19, 2022. doi:10.3389/frai.2022.866920.

Josue Torres-Fonseca, Catherine Henry, and Casey Kennington. Symbol and communicative grounding through object permanence with a mobile robot. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 124–134, Edinburgh, UK, 2022. Association for Computational Linguistics. doi:10.18653/v1/2022.sigdial-1.14.

Nicholas Thomas Walker, Stefan Ultes, and Pierre Lison. A graph-to-text approach to knowledge-grounded response generation in human–robot interaction. *Dialogue & Discourse*, 16(3):60–95, 2025. doi:10.5210/dad.2025.304.

Christopher D. Wallbridge, Alex Smith, Manuel Giuliani, Chris Melhuish, Tony Belpaeme, and Séverin Lemaignan. Towards the effectiveness of ambiguous spatial descriptions in human

robot interaction. In Hendrik Buschmeier, Mary Ellen Foster, and Dimitra Gkatzia, editors, *2nd Workshop on Natural Language Generation for Human–Robot Interaction*, Online, 2020. URL https://purl.org/nlghri2020/wallbridge.

Sina Zarrieß and David Schlangen. Being data-driven is not enough: Revisiting interactive instruction giving as a challenge for NLG. In Mary Ellen Foster, Hendrik Buschmeier, and Dimitra Gkatzia, editors, *Proceedings of the Workshop on NLG for Human–Robot Interaction*, pages 27–31, Tilburg, The Netherlands, 2018. Association for Computational Linguistics. doi:10.18653/v1/W18-6906.