A Survey on Dialogue Management in Human-Robot Interaction

Reimann, Merle M.; Kunneman, Florian A.; Oertel, Catharine; Hindriks, Koen V.

DOI
10.1145/3648605
Publication date
2024
Document Version
Final published version
Published in
ACM Transactions on Human-Robot Interaction

A Survey on Dialogue Management in Human–robot Interaction

MERLE M. REIMANN, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
FLORIAN A. KUNNEMAN, Utrecht University, Amsterdam, Netherlands
CATHARINE OERTEL, Delft University of Technology, Delft, Netherlands
KOEN V. HINDRIKS, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

As social robots see increasing deployment within the general public, improving the interaction with those robots is essential. Spoken language offers an intuitive interface for the human–robot interaction (HRI), with dialogue management (DM) being a key component in those interactive systems. Yet, to overcome current challenges and manage smooth, informative, and engaging interaction, a more structural approach to combining HRI and DM is needed. In this systematic review, we analyze the current use of DM in HRI and focus on the type of dialogue manager used, its capabilities, evaluation methods, and the challenges specific to DM in HRI. We identify the challenges and current scientific frontier related to the DM approach, interaction domain, robot appearance, physical situatedness, and multimodality.

CCS Concepts: • **Computing methodologies** → *Discourse, dialogue and pragmatics*; • **Computer systems organization** → *Robotics;*

Additional Key Words and Phrases: Spoken interaction, dialogue management, social robots

ACM Reference Format:
Merle M. Reimann, Florian A. Kunneman, Catharine Oertel, and Koen V. Hindriks. 2024. A Survey on Dialogue Management in Human–robot Interaction. *ACM Trans. Hum.-Robot Interact.* 13, 2, Article 22 (June 2024), 22 pages. https://doi.org/10.1145/3648605

1 INTRODUCTION

For humans, spoken communication is a natural way of interacting with each other, their smart speakers, and even their pets. Social robots are robots that are designed specifically to interact with their human users [16], for example, by using spoken dialogue. For social robots, the interaction with humans plays a crucial role [8, 32], for example, in the context of elderly care [18] or education [10]. Robots that use speech as a main mode of interaction do not only need to understand the user's utterances but also need to select appropriate responses given the context. **Dialogue management (DM)**, according to Traum and Larsson [104], is the part of a dialogue system that

Authors' addresses: M. M. Reimann and K. V. Hindriks, Vrije Universiteit Amsterdam, De Boelelaan 1111, Amsterdam, North Holland, 1081 HV, Netherlands; e-mails: m.m.reimann@vu.nl, k.v.hindriks@vu.nl; F. A. Kunneman, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands; e-mail: f.a.kunneman@uu.nl; C. Oertel, Delft University of Technology, P.O. Box 5031, Delft, South Holland, 2600 GA, Netherlands; e-mail: c.r.m.m.oertel@tudelft.nl.

ACM 2573-9522/2024/06-ART22
https://doi.org/10.1145/3648605

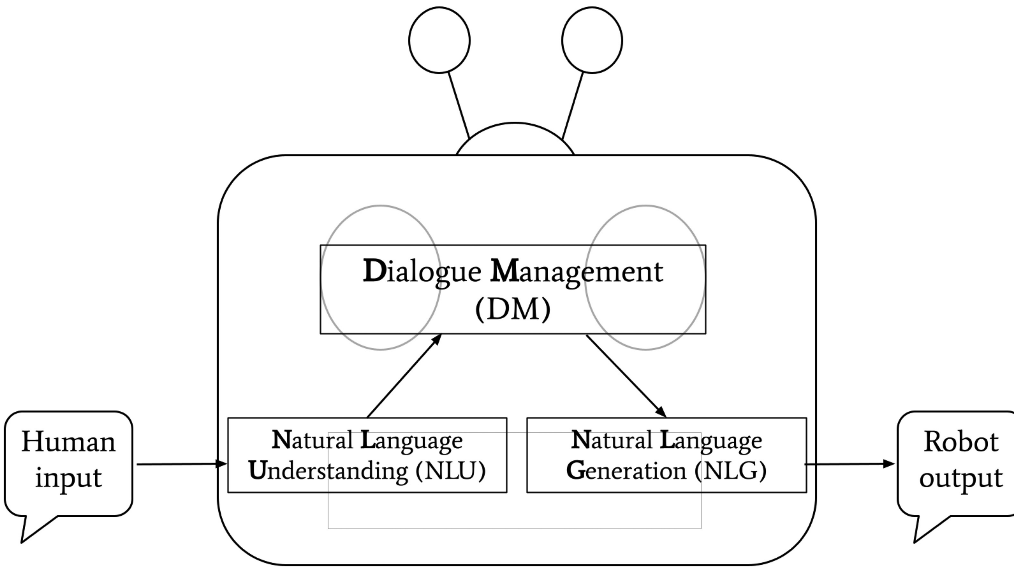

Fig. 1. The integration of the dialogue manager into a spoken dialogue system.

performs four key functions: (1) it maintains and updates the context of the dialogue, (2) it includes the context of the utterance for interpretation of input, (3) it selects the timing and content of the next utterance, and (4) it coordinates with (non-)dialogue modules. In spoken dialogue systems, the dialogue manager receives its input from a **natural language understanding (NLU)** module and forwards its results to a **natural language generation (NLG)** module, which then generates the output (see Figure 1).

In contrast to general DM, DM in **human–robot interaction (HRI)** has to also consider and manage the complexity added by social robots (see Figure 2). The concentric circles of the figure describe decisions that have to be made when designing a dialogue manager for HRI. From each circle, one or more options can be chosen and combined with each other. The robot appearance, modalities of the interaction, interaction scenarios, and physical environment influence the DM. Their combination leads to high variability, but also great complexity. While pure neural networks are used for DM in non-HRI contexts [15, 112], this approach is not adopted generally for HRI, where sparse data, need for robustness, and control of high-stakes interactions pose additional constraints.

While DM and HRI have not been studied extensively in combination, this is not the case for the fields individually. Harms et al. [43], Brabra et al. [15], Deriu et al. [25], Zhao et al. [112], and Trung [105] investigate DM, but not from a robotics perspective, and Skantze [98] looks at turn-taking, a part of DM. While there are reviews on DM in HRI, these have a specific user-group focus, such as patients suffering from dementia [92]. An overview of the present and future of natural language in HRI, without a specific focus on DM, is provided by [69], [74], and [72]. In contrast to those reviews, we focus on dialogue managers that are used in physical robots and provide a general overview of DM in HRI. With this review, we aim at giving HRI researchers an overview of the currently used DM systems and their capabilities, to help them make a more informed decision when choosing a dialogue manager for their system.

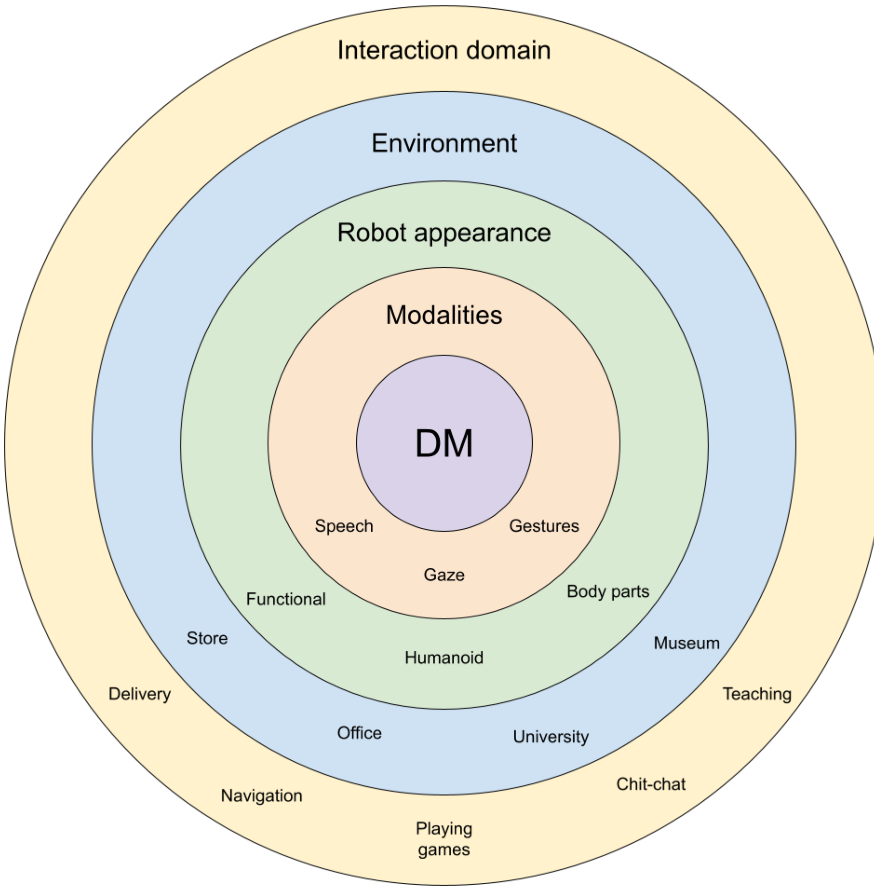

Fig. 2. Dialogue management in human–robot interaction is influenced by multiple factors that show high variability. Different robot appearances can be combined with varying modalities, interaction domains, and environments. Some examples are given for each factor to illustrate it.

2 REVIEW

We conducted a systematic review using the PRISMA protocol [76]. We determined inclusion and exclusion criteria, mentioned below, which guided the selection steps, based on our research questions:

RQ1: Which robots are commonly used for DM? (Section 2.1).
RQ2: Which types of dialogue managers are commonly used in HRI, and what are the reasons? (Section 2.2).
RQ3: Which capabilities do current dialogue managers used in HRI have? (Section 2.3).
RQ4: How are dialogue managers in HRI evaluated? (Section 2.4).
RQ5: What are the main challenges for DM in HRI? (Section 2.6).

We used Scopus, IEEE, and ACM for the literature search using the search terms "Robot AND 'mixed initiative"' and "Robot AND ('dialog manager' OR 'dialogue manager')," resulting in 949 papers (ACM: 556, Scopus: 278, IEEE: 115). Performing a second search with the more general search terms "Robot AND (dialog OR dialogue)" but a restriction to relevant conferences (HRI, IUI, ICMI, AAMAS, SIGGRAPH, HAI, SIGDIAL, RO-MAN, IROS, INTERSPEECH, DIS, MHAI, AAAI) led to another 504 papers (ACM: 62, Scopus: 350, IEEE: 92). To make sure also that papers that use the term "interaction" instead of "dialogue" in the abstract and title are included, we performed a third search with the search term "'social interaction' AND robot AND speech." This led to 209 (ACM: 20, Scopus: 134, IEEE: 55) papers. All papers were imported into Rayyan [81] for further processing.

The formal exclusion criteria were chosen to filter out papers shorter than four pages, reviews, demonstrator papers, and results that are not scientific papers or did not include a physical robot. Furthermore, we excluded papers written before 2005 since the NLU modules and DM capabilities have improved a lot since then. Additionally, the papers needed a focus on spoken DM or DM capabilities of the dialogue manager. DM capabilities are the conversational abilities the dialogue manager possesses. We decided to scope the review to only include papers that describe a whole dialogue manager. While other components like dialogue state trackers do help the dialogue manager, by providing information that the dialogue manager can then use to decide what its next action should be, they are out of scope for this survey. After duplicate deletion and keyword-based filtering ("dialog," "dialogue," "dialogs," "dialogues," "conversation," "conversational," "conversations," and "discourse"), all authors independently performed a manual abstract screening of the remaining 753 papers, excluding papers that do not sufficiently focus on spoken HRI. Conflicting decisions were discussed until an agreement was reached. A subsequent paper screening led to a total of 68 papers analyzed for this review.

2.1 Robot Appearance

We classify the 69 robots found in the 68 papers by their appearance using the taxonomy in [8]. Figure 3 shows the resulting distribution of the various types of robots found. *Humanoid robots* (41%) are one of the most dominant categories in spoken HRI. Within this category, the NAO robot, used in 10 papers, is the most commonly used humanoid robot [4, 13, 21, 24, 26, 36, 64–66, 79]. The Pepper robot [50, 78, 109] and Maggie [2, 3, 41] are used in three papers each. A robot bartender [33, 84], Armar 3 [47, 87], and the PR2 robot [37, 108] are mentioned twice, while all other humanoid robot appearances occur just once [22, 27, 52, 59, 62, 73, 103]. The second biggest category consists of *functional robot* appearances (30%); the design of those robots is influenced by the task they are used for. The three Segway-based robots in this category are used for delivery and navigation [5, 102, 111]. Mobility plays a role for all robots in this category as they are used to help with information and action requests based on instructions [1, 7, 23, 30, 58, 71, 88–90, 93], act as a guide [35, 62, 63, 70, 91, 107], or learn location names [34, 77].

For the robots based on body parts, two robotic heads were used: Flobi [20, 54] and Furhat [19, 53, 99]. In contrast to robotic heads, which do not have manipulators, robotic arms are used for object manipulation and grasping. We found robotic arms in five of the papers [82, 83, 95, 96, 101]. Robots based on other body parts were not found in any of the papers. All of the artifact-shaped robots found in the papers are robotic wheelchairs [29, 44, 45, 106], which are used for navigation. The papers using android robots all use ERICA as the robot [49, 56, 57, 75]. One of the robots was not described sufficiently to classify it [38].

Based on the diversity of robot types used, we can conclude that the use of spoken dialogue systems is not restricted to specific robots with certain shapes or features, which adds to the high variability of factors influencing the DM (see Figure 3).

2.2 Types of DM in HRI

While the first DM approaches starting in the 1960s were purely handcrafted, there has been a development into the direction of probabilistic and hybrid approaches. We use the framework provided by Harms et al. [43] as a basis for the classification of dialogue managers into handcrafted,

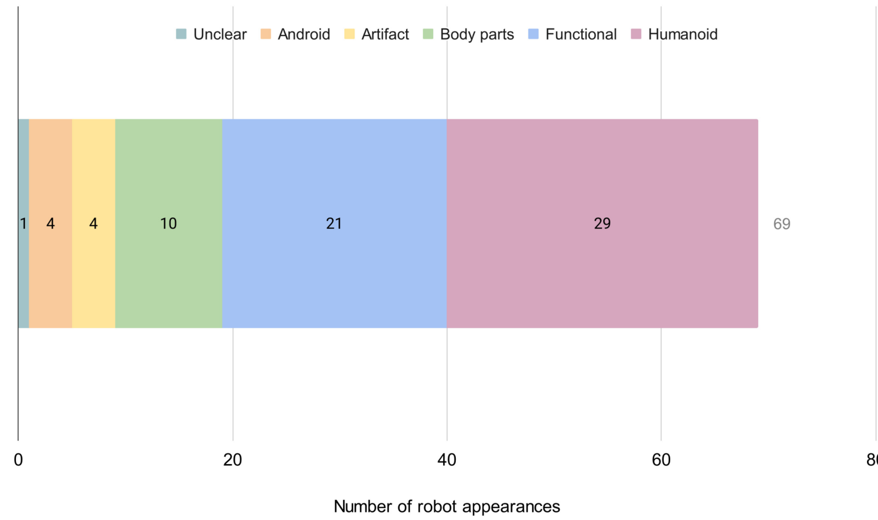

Fig. 3. Appearance of the robots used in the surveyed papers.

probabilistic, and hybrid approaches. For handcrafted approaches, the rules and states of the dialogue manager are defined by experts and developers, whereas the rules and states are learned from data by probabilistic approaches. Hybrid approaches are a combination of handcrafted and probabilistic ones. In total, 45 of the 68 papers use handcrafted approaches, 12 use probabilistic ones, and 11 make use of a hybrid approach (see Figure 4). From 2006 to 2010, 17 of the 23 included dialogue managers were handcrafted, with model-based approaches making up over 50% in total. Model-based approaches stayed popular from 2011 to 2015, but **(partially observable) Markov decision processes ((PO)MDPs)** saw an increased use as well. From 2016 onward, those two approaches were surpassed by hybrid ones that can combine the advantages of both.

2.2.1 Handcrafted Approaches. The main difference between the different handcrafted approaches lies in the information encoded in the states of the dialogue manager. **Finite-state machines (FSMs)** are a simple approach to handcrafted DM, where the states of the FSM directly correspond to the dialogue states. A simple FSM approach is used in [93], where the focus of the interaction lies in the learning of tasks through dialogue. [58] uses a composition of multiple FSMs to manage the different interactions their robot is offering. However, nowadays they are mostly used as building blocks in combination with other additional functionalities in the surveyed papers. For example, [30] uses an FSM with predefined dialogues for knowledge acquisition but aims at adding an autobiographical memory to the robot. Dialogue managers based on state charts, which are used to implement interaction patterns, are used in [20, 54, 82, 83]. Interaction patterns are based on general sequences that can be found during dialogues and are defined before the interaction. Another dialogue manager based on state charts is IrisTK, which adds recursion [53, 99]. Building upon IrisTK, [53] adds a data-driven turn-taking approach to the system.

To allow the human user to take more initiative and provide information more flexibly in dialogues, *frame-based systems* have been used [12]. Frame-based approaches employ empty slots, typically information units required to fulfill the user's request, that can be filled at any point of the conversation. [41] presents such a system, where multiple slots can be filled at a time, for action requests for the robot Maggie. A frame-based approach is also adopted in [109], where the user's food preferences are determined in an interview-like conversation by using slot-filling.

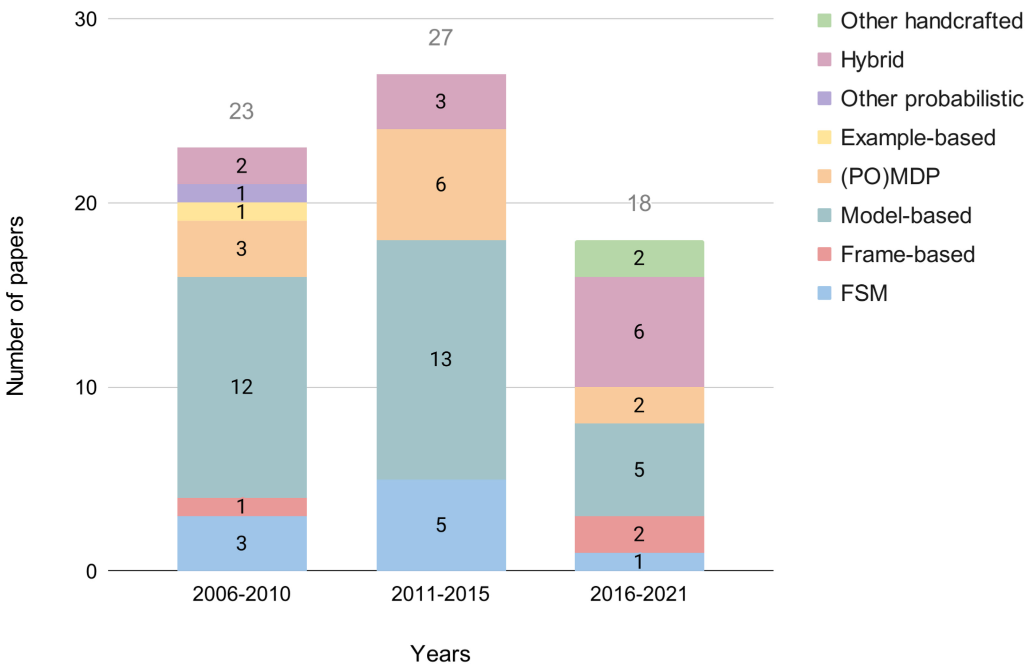

Fig. 4. The dialogue management approaches over the years used in the surveyed papers.

The third handcrafted approach is the *model-based approach*. In addition to the frames with slots, a model-based dialogue manager has more complex states. Those states include some form of model, for example, about the user, the environment, the situation, or the context. Model-based dialogue managers often consider the user model, including the user's goals and intention [21, 88, 90, 91, 102, 107, 108]. User models that influence the DM can also be based on the displayed emotion [3, 26] or personality [4] of the user. For long-term interaction, including a model of the conversation history is especially relevant and enables the dialogue manager to learn from past interactions. Cobot's dialogue manager described in [91] is designed for a long-term interaction and therefore uses the conversation history to avoid repetitive dialogue or unnecessary offers. In [27], the robot collects and stores user data that influences later conversations with the user. [62, 63] use a stack to keep track of unfinished conversation sequences for tracking the context during the interaction. The dialogue manager used in [78] groups utterances into contexts, which then can be focused during different points of the conversation.

The *Information State Update* approach [104] used by [88–90] stores and uses among other things the dialogue context, information about the possible referents of pronouns, and information about the current turn and initiative. [1, 7, 95] also use the dialogue context and user's intention, in addition to objects referenced in the environment, to determine the next dialogue move. Multiple expert modules, activated based on domain scores, are used by [77] and [34].

Plan-based systems, which belong to the model-based approaches, treat the conversation as a planning problem and aim at reaching a goal state by selecting from predefined actions. Plan-based DM is used by [84], [2], and [33], who use the planner for the dialogue and action generation. Other plan-based systems are used by [106] for navigation of a robotic wheelchair and by [36], with additional rule-based speech overlap management, which is integrated in the cognitive DIARC architecture. [23] uses answer set programming for their planning module, which can generate a high-level plan based on the user's request and a low-level plan for the motion planner. In [19], a goal, based on novelty, dialogue history, relevance, and diversity, is used for the utterance selection, while [35] performs logical inference over the goals and beliefs to this end.

The dialogue manager used by [52] can be used as a model-based one, but a simplified version can also be used as a frame-based or FSM dialogue manager. Two dialogue managers do not fit the classification and were therefore classified as *other handcrafted*. [22] used timed Petri nets to improve multimodal interaction, especially in respect to turn-taking. Petri nets are state-transition networks that have tokens, which determine the current configuration state and whether a transition is enabled. In addition to utilizing this approach, timed Petri nets use real-time transitions of the tokens to be able to model the configuration according to the input. [79] uses a dialogue manager that uses XML files for generating the dialogue in Slovac languages.

2.2.2 Probabilistic Approaches. In contrast to handcrafted approaches, probabilistic approaches learn the dialogue policy directly from data. A probabilistic approach used in only one surveyed paper is the *example-based approach* [103], which looks up the most similar utterance in its database, consisting of numerous conversations, and uses the answer given there as its own.

A popular probabilistic approach used in DM is POMDPs. MDPs model the dialogue as a Markov process, moving from one state to another based on learned transition probabilities. For POMDPs, the current state is not known and is estimated based on observations (e.g., perceived actions and utterances) and their probabilities. The robotic wheelchairs used in [45] and [44] use dialogue to check their spatial semantic representation and make sure that it corresponds to the real environment by asking questions about the environment. The POMDP dialogue manager in [5] rewards successful conversations, leading to task completion, while additional questions, confirmations, and unsuccessful conversations are penalized. Two POMDPs, one for the DM and another one for deciding when to update the knowledge base, are used for improving the augmentation of the knowledge on an as-needed basis. [24] allows flexible navigation between the different subdialogues to enable the user to change their mind about the order in which they want to do certain tasks. A Bayesian approach to learning the best dialogue strategy is used by [29] for a robotic wheelchair. Integrating reasoning with probabilistic common-sense knowledge enables the robot in [111] and [70] to ignore irrelevant variables while planning. Another approach for limiting the state space is used in [64], where probabilistic rules are used for Bayesian networks. (PO)MDPs are also used for multimodal dialogue, where they learn the strategies from multimodal instead of pure dialogue data [47, 71, 87].

We did not find a system purely based on an *end-to-end learning approach*, but end-to-end approaches are in fact used as parts of hybrid systems (see Section 2.2, Hybrid Approaches). A probabilistic approach that does not fit into either category described earlier is used by [101]. They generate all possible linguistic expressions for an action, which is possible since they use a simplified sentence structure. Then they use expected log loss reduction and Bayesian logistic regression, predicting which utterance should be selected. Using the best score, an utterance is selected and produced. After the production, the appropriateness of the utterance is rated based on the user's reaction and the example is added to the training set. This leads to active learning during the conversation.

2.2.3 Hybrid Approaches. Instead of focusing on either a handcrafted or data-driven approach, in recent years it has become more popular to use a combination of the two. These so-called hybrid approaches can be fitted according to the needs of the target application and the availability of data. The option to combine approaches can compensate for their weaknesses and utilize their strengths. Hybrid approaches are often used when the dialogue manager is distributed in a variety of submodules that focus on different aspects of the DM, like turn-taking, engagement management, or

Table 1. Common Conversational Capabilities Observed in the 68 Surveyed Papers

Conversational Capability	Amount
Confirm	13
Ask for confirmation	24
Ask for repetition/rephrasing	10
Ask clarification questions	17
Reference resolution/spatial grounding	24
Acquire knowledge	21

problem handling. [56] combines a finite-state turn-taking machine with a deep neural network in order to improve the dialogue manager's turn-taking capabilities. Turn-taking is also integrated into the otherwise model-based dialogue manager in [37] by adding a combination of incremental processing and lexical prediction. In [13], an FSM approach is combined with self-supervised learning to handle disengagement. To handle problems occurring during the dialogue more efficiently, [38] adds an FSM with four states (Start, Normal, Help, Error) to the dialogue manager ARIADNE [46]. [96] also uses an FSM for tracking the general state of the dialogue system but makes use of external services for questions and declarative utterances.

[57] uses logistic regression to select so-called experts, which can have different strategies, while [49] relies on a priority system in combination with additional triggers for a backchannel and backup question module. Backchannels are short responses that are integrated into the conversation while listening, to show engagement (e.g., "hmm," "yeah"). In [75], a dialogue act tagger based on supervised machine learning (Support Vector Machines) is used to decide whether a statement response should be generated, based on either a decision tree or a response using an example-based approach. If neither have a sufficiently high confidence score, a backchannel is produced instead.

It is also possible to make use of probabilistic rules for DM to reduce the parameter space by including expert knowledge [66]. A combination of a POMDP and a model-based approach is used in [65], where the user's goal and actions are modeled using Bayesian inference. A combination of an example-based and agenda-based approach is used in [59]. While the agenda graph is designed manually, the examples that are mapped to the nodes are learned from human–human dialogues.

2.3 Communication Capabilities in HRI

We list common capabilities that were explicitly mentioned in the reviewed papers in Table 1.

Confirm. Explicit confirmations by the robot for establishing common ground are important, especially for task-based interaction, to increase task success [102]. Common ground describes shared beliefs, for example, about the environment or the state of the conversation [100], which is relevant if the human and robot have to achieve a task together. Example use-case scenarios include movement instructions [35, 41], receptionist tasks [47, 52, 90], grasping [83], delivering items [102], or the establishment of common ground in situated interactions [21].

Ask for confirmation. A clear trend across more than one-third of all papers is the robot's ability to ask for confirmation. This can happen either explicitly (e.g., "Do you want...?", with an option for the user to (dis-)confirm) or implicitly (e.g., "I will now...", with an option to cancel the action if it is not the desired one). [82, 83] use a correctable information request pattern for implicit confirmations and an explicit confirmation after an Information Request pattern for integrating confirmation questions into their task-based interaction. Confirmations by the human can be used to ensure common ground about the environment [21, 47, 99], to elicit information about user preferences

[70, 109], to confirm the correctness of past actions [64], or to make sure that the conversation is not terminated too early [93]. At the same time, unnecessary confirmation or clarification questions can make the conversation inefficient and increase user frustration and are therefore often associated with a cost in reinforcement learning settings [5, 29, 44, 64, 111]. For example, robots that navigate or deliver items [5, 29, 66, 111] or make appointments and book something [52, 90] often ask for confirmation, since a later correction after performing the action is more costly. To avoid unnecessary confirmation questions, [65] assigns a negative value to the reward function of their reinforcement learning dialogue strategy for additional confirmations, while the robot in [101] learns when and how to confirm through active learning. Since robots have to deal with challenging environments with noise, confirmation questions are used when the confidence that the utterance was understood correctly is low [41, 77, 89, 90, 108].

Ask for repetition/rephrasing. In case of speech recognition problems, a repetition of the utterance can already help resolve them, while rephrasing gives the human the option to produce a different utterance that is easier to understand for the robot than the first one. To determine whether repetition is needed, different strategies are available. Robots ask users to repeat an utterance when the recognition is below a threshold [41], when no match for the input was found [20, 35], or in case of underspecification [65, 66]. If no grounding is achieved for object names that are referred to by a gesture accompanying a pronoun, the robot in [63] asks for repetition. Like repetition, rephrasing can be used when the dialogue manager finds no match for the input [35, 50, 59] or if the dialogue manager is uncertain either about the dialogue state [5] or about the correct recognition of the utterance [41].

Ask clarification questions. Clarification questions are asked in those cases where the robot is unsure if it has understood the utterance correctly [5, 22, 88, 89, 102] or when it is missing information [35, 38, 41, 47, 109]. Additional information to make complete plans [23] or determine the current context [108] can be acquired through clarification questions. If most parts of a user request are already specified and only some need additional clarification, [20] suggests to let the robot already start the action while clarifying the missing parts.

Reference resolution/spatial grounding. Something that distinguishes robots from conversational agents is that they are physically embodied and situated in an environment. That is why situational grounding is important for robots that, for example, engage in navigation [44, 45, 66, 91] or deliveries [102] or give a tour in the environment [63, 82, 93]. Situational grounding refers to the robot's ability to ground concepts and referenced objects in the real world ("Give me the red cup" would refer to a specific red cup that is present in the environment, not to the concept of a red cup in general). The robot in [21] engages in a naming game of objects in the environment with the user and focuses on establishing common ground. To do so, it makes its knowledge explicit, such that the human can correct it if necessary. Robotic arms that grasp and move objects need to identify the object the user is talking about correctly [83, 95, 101]. For multimodal dialogue where the human refers to objects in the environment that are related to the robot's task, the robot also needs to be aware of which object the human is referencing using the available modalities (e.g., gestures, speech, gaze) [1, 7, 47, 87].

Acquire knowledge. If the human is talking to the robot about previously unseen or unheard objects or people, the robot should be able to make sense of them. Knowledge augmentation allows for new knowledge to be added to the knowledge base of the robot. For navigating robots, questions about their environment improve their knowledge about the spatial semantic representation [44, 45]. In the delivery domain, knowledge augmentation is used to learn names of objects or people that have to be delivered/should receive the delivery but are not yet in the knowledge base [5, 102]. Another context in which knowledge from previous interactions and users is used is for recommendations based on (assumed) user preferences [70, 107]. The robot in [90] is able to acquire and store knowledge, which is explicitly declared. In some cases the interaction with the robot is used to teach it new actions [95], for grasping methods in combination with object names [83], or for following instructive tasks [93]. Other forms of knowledge acquisition include active learning from past interactions [101], the internet [78], or updates of the world model [23] in order to avoid failures of the interaction. Another case of knowledge acquisition takes place when the robot learns personal information about the human and stores it in a knowledge base [3, 30].

2.4 Evaluation Methods of DM in HRI

Evaluation metrics. Different techniques have been used for evaluating the performance of dialogue managers in general. A detailed survey of them can be found in [25]. For the evaluation of dialogue managers in HRI, subjective and objective measures can be used (see Table 2).

Dialogue and tasks are often tightly coupled in task-based systems, and the dialogue is needed to achieve the goal. For task-based studies we therefore observed a high number of papers reporting the task success rate, a metric also observed as a common evaluation method by [25]. Since 58 of the 68 papers report on task-based interactions, task success offers an easily accessible option to evaluate the interaction. Number of turns is a metric that can be used independent of the system and would aid in the comparability of different systems and tasks if reported for all studies. However, the evaluation of the system still depends on the type of system used [25]. Even though this can partly account for the high number of different evaluation metrics, it would be advantageous if common metrics, like the number of turns, would be integrated into the evaluation of all studies to make them more comparable.

Method of evaluation. Performing user studies for the evaluation has the advantage that they allow to evaluate the performance of the dialogue manager and robot in a real interaction, while allowing the users to give their opinions in questionnaires, like the Godspeed questionnaire [9]. That way, user frustration and other subjective measurements can be included in the evaluation. However, user studies are resource and time intensive. In the reviewed papers, 35 user studies were reported, with a few (3) of those being in-the-wild studies, which take place outside of the lab [13, 19, 53]. Other evaluation methods include simulations (7) and evaluations based on a corpus of previously gathered data (10), while some papers just describe possible interaction scenarios (17). Crowdsourcing can be used to evaluate the dialogue manager independently from the robot [5, 102] or the interaction from a third person point of view.

The participant numbers reported for the user studies differ greatly, ranging from 2 to 97 participants, with an average of 21 participants (SD 18.4). Three studies [41, 70, 101] mention tests with users but do not report any specifics about the number of participants. For in-the-wild studies it is more difficult to report exact participant numbers, since it is not always clear if every user interacted only once, which is why [13, 19, 53] report the total number of recorded interactions instead.

Evaluation of parts vs. the whole system. Due to the integration of the dialogue manager into a bigger system, the dialogue manager is often evaluated in combination with the other modules of the robot. For example, the task success can include the dialogue but also the motor functions to achieve the task. While the integration of dialogue managers into a robot opens up the possibility of using evaluation methods that assess the whole system, it complicates the evaluation. An example is that the dialogue manager gets the information from the speech recognition module, which can introduce misunderstandings that the dialogue manager then has to deal with. To avoid those problems, simulations or corpus-based evaluations can be used. An evaluation of the dialogue

Table 2. Subjective and Objective Measures Used for Evaluating DM in HRI

Subjective measures
User satisfaction [19, 26, 102]
User frustration [5, 102]
Perceived interaction pace [24]
Perceived understanding [66, 102]
Scales from social psychology/communication [103]
Naturalness and pleasantness [66, 71]
Ease of use [24, 102]
Likability of the robot [27, 38, 57, 63, 103]
(Modified version of) the Godspeed questionnaire [33, 75]
Items from the Interpersonal Communication Satisfactory Inventory [19]
Perceived common ground [21]
Likelihood of usage in the future [5, 24, 102]
Appropriateness of the answer [35, 66]

Objective measures
Number of turns [21, 33, 38, 47, 59, 65, 66, 87, 102, 111]
Number of repair turns [62, 66]
Number of fallback utterances [75]
Number of rejected utterances [90]
Precision and recall [56, 75]
Accuracy [5, 44, 56, 64, 96, 111]
F1 scores [5, 56]
Task success rate [5, 21, 22, 33, 38, 47, 59, 71, 87, 90, 99, 102]
Entropy reduction [44, 45]
Latency and false cut-in rate [56]
Reward/cost functions [5, 29, 70, 87, 111]
Dialogue duration [66]

manager alone does not necessarily reflect how it would work in combination with the robot's other parts, especially if the robot makes use of multimodal data.

2.5 Action and Dialogue Coordination

As most robots used for spoken interaction not only engage in dialogue but also perform actions, the dialogue and the actions have to be coordinated to ensure a coherent experience. [67] presents an overview of the methods for end-to-end and modular action and DM coordination to help system designers to make an informed decision about which coordination strategy to choose. A collection of the approaches for this coordination that we observed in the surveyed papers can be found in Table 3. Even though they have the physically situated embodiment, not all robots perform non-dialogue actions. For those that perform actions and engage in dialogue, we observed four different ways in which the action and dialogue managers were coordinated in the system.

As the most common coordination method (43%), we observed dialogue managers also being responsible for the action selection. Actions can be, for example, selecting and navigating to locations [44, 45] or gestures of the robot [3]. The dialogue manager thus not only is able to perform dialogue moves but also can select actions as the next interaction move. In contrast to this, only three papers have an action manager that controls the dialogue moves as well.

Table 3. The Relation of the Dialogue and Action Manager to Each Other, as Observed in the Surveyed Papers

Action and dialogue coordination	
Only dialogue	[52, 54, 56, 57, 75, 87, 109]
Dialogue manager and action manager interact on one level/have a central interaction manager controlling both	[22, 24, 53, 62, 99, 101]
Dialogue manager is also in control of actions	[1, 3, 7, 20, 29, 30, 35, 41, 44, 45, 47, 60, 62–66, 70, 71, 77, 78, 82, 83, 88–90, 96, 107, 111]
Action manager is also in control of dialogue	[26, 58, 93]
Action manager makes decision based on input received from dialogue manager	[2, 4, 13, 21, 23, 27, 33, 34, 36, 37, 49, 50, 73, 79, 84, 95, 102, 106, 108]

Instead of having one module control the output for both action and dialogue, it is possible to have separate modules for both that interact on one level or are both controlled by a middleware acting as a central component.

The second most common coordination strategy observed is to feed the output of the dialogue manager into an action manager, which then, based on the received dialogue input, decides what action to do next. The dialogue manager can, for example, create goals that are then passed to a module that can choose an appropriate action for that goal [36, 37, 102, 108]. In cases where the dialogue manager and action manager are chained, the dialogue manager does not directly have information on which action will be produced based on its output. This approach allows to modify the possible output actions and modalities without needing to change the dialogue manager itself.

2.6 Challenges

Dialogue managers that are integrated into a robot do not act independently but are part of a bigger system architecture. While this is also true for dialogue managers used without robots, robots often make use of additional sensors and actuators to integrate multimodal input and output into the system. Currently, there is no commonly used off-the-shelf solution for spoken HRI, but instead there is a variety of different frameworks, not only for DM, but also for the related modules. This variety makes it challenging to select the most appropriate solution. Due to the tight coupling of modules in a robot, the dialogue managers are also influenced by the problems of other modules, such as speech recognition errors [34, 49, 57, 90]. Apart from improving the speech recognition module directly, it is possible to include mechanisms into the dialogue manager that are responsible for dealing with speech recognition errors [34].

Expectation management. Talking to a robot raises expectations about their conversational capabilities. When those expectations are not met, the user can become frustrated [5]. Expectations are shaped in part by the morphology of the robot [55], making the consideration of the morphology on the expected conversational capabilities necessary. For example, [75] uses the Android robot ERICA and states that "Erica's realistic physical appearance implies that her spoken dialogue system must have the ability to hold a conversation in a similarly human-like manner by displaying conversational aspects such as backchannels, turn-taking and fillers." The robot morphology does not only affect the expectations, for example, regarding the human likeness of its conversational capabilities, but also limits the tasks the robot can do, impacting the domains the robot has to be able to converse about.

In human–human conversations, common ground plays an important role and helps to reduce miscommunication [100]. In spoken HRI this is even more challenging since the perception of humans and robots is not the same and the human does not necessarily know what the robot is able to perceive and understand. Expectations based on human–human conversations do not necessarily hold for human–robot conversations. This can lead to a gap between the perceived versus the real common ground [21].

Multimodality and situational grounding. Interactions with robots are not purely speech based, but they can also make use of multimodal cues like gestures, gaze, or facial expressions. When the robot is using multimodal input for deciding the next dialogue move, it is not enough to simply detect the multimodal cues; they also need to be integrated for further processing. A separate model for multimodal fusion can be used that then forwards the fused information to the dialogue manager, as is done, for example, in [3, 47, 62, 71]. Before integrating multimodal cues, a decision has to be made regarding the required modalities and which ones are necessary or expected in the specific type of interaction and should be used by the robot. The robot morphology is impacting the multimodal cue generation as well, as robotic heads, for example, cannot generate pointing gestures. Therefore, the robot's appearance is adding both restrictions and expectations and should be chosen carefully.

HRI offers the additional challenge of grounding the interaction in the environment the robot is physically located in. This is especially the case for multimodal interaction where the referral of objects can happen, for example, through gestures [1, 7] or gaze [99]. A robot that is referring to locations around itself needs to know its own position as well as the names and places of the locations around it [34, 42, 45, 77]. Grounding interactions in the environment is especially challenging since the environment can change over time, with objects being placed and removed, people appearing and disappearing, and the robot itself moving in the environment. The physical environment of the robot also contains objects, which the robot or human can refer to. Especially for robots that can perform object manipulation, it is important to be able to correctly understand and produce references to those specific objects [62, 65, 83, 95, 101].

Multi-party and long-term interactions. When a robot is placed in an environment, it can encounter the same person multiple times or have longer interactions with the people, for example, when acting as a guide [70, 107]. In longer conversations or long-term interactions with multiple conversations, repetitions over time can annoy the user [91], and a lack of conversational memory will prevent the conversation from moving away from superficial chit-chat or strictly task-based conversations. To solve the problem of repetitions in long-term interactions, [91] suggests different options, for example, producing a number of different utterances for the same context that could reoccur and remembering previous utterances or even whole conversations. To understand the robot's behavior and its decisions, additional explanations might be required [27, 95]. Especially if the robot fails without an explanation, the user might misinterpret the situation or cause the failure again due to the lack of explanation.

Resources needed. However, this leads to the next problem, which involves the resources needed for designing complex dialogue managers. For example, for handcrafted approaches the structure has to be defined a priori [41, 52], knowledge has to be integrated [50], and even then it will not be able to cover unknown situations [45, 90]. In contrast to handcrafted approaches, data-driven ones need less human effort but rely on data for learning conversational behavior. However, data collection in HRI is expensive [29], and for supervised learning expert feedback is needed [64]. It is possible that the human and the robot follow different conversational strategies; for example, one tries to exploit previous conversation topics, while the other one tries to explore new topics. If the robot does not notice those differences in strategies and rigidly follows its own strategy, it can lead to high tension in the dialogue [19]. Collecting data in an HRI setting is a challenging task, since the data are often multimodal and the interaction has to take place in a situated setting. This means that the data collection requires time, human participants, and a system that can record all the required signals. However, if a specific robot is used for recording data, the generalizability to other robots is still unclear.

Variety of metrics and approaches. The reporting of user studies and their results is still lacking uniformity. Exemplary of this are the multiple ways in which length can be assessed in the context of dialogues. Some papers report the number of turns [19, 21, 47], while others focus on the length of turns [27, 59] or the total amount of time [53]. An additional difficulty with using length as a metric is that it is highly context dependent on whether a high or low number of turns is desirable in the conversation. While the heterogeneous reporting on the length can make comparisons challenging, there are also still papers that do not report on the interaction length at all.

To date, handcrafted DM approaches are still very common in HRI, due to lack of data and the aforementioned challenges due to the physical embodiment, but also because of the advantages it offers. If the robot's conversation domain is small and transparency is required, a handcrafted approach might be a good option, whereas a probabilistic approach has the advantage that the actions can be learned from real interaction data, leading to less expert knowledge being required.

2.7 Scoping

While the survey aims at giving an overview of the past and current state of DM in HRI, there is still some related work that is not captured for various reasons. The related work we look at in this section can be grouped into three categories: (1) end-to-end interaction management, (2) cognitive architectures, and (3) end-user programming. One of the main reasons is the variety in terminology used for DM in HRI, while another reason is that we only used papers with physically situated robots where an interaction of a human and robot is part of the paper. The variety in terminology is increased by HRI papers that include DM but do not specifically name the DM part. Instead, they just describe how the dialogue is modeled in the overall dialogue system [14, 48].

For robots where end-to-end models are used, those models are not just used for the dialogue manager but often to control the whole interaction [17, 28, 68]. In those cases the DM on its own is the main focus of the research but is just a (desired) byproduct of the interaction output. Because of this, we did not focus on such studies in this survey on DM in HRI. Papers that focus on non-verbal modalities and evaluate the addition of those to already existing DM systems [31] were also out of scope for this survey. This multimodal focus on the behavior generation that is accompanying speech is something that can be found in end-to-end or data-driven systems, as discussed in a survey paper by [80].

Besides the already mentioned DIARC architecture [36, 51, 107, 108], other cognitive architectures are deployed in HRI that often feature plan-based DM. One example is ACT-R [6, 14], where a procedural system is used to create output based on memories, goals, and other modules.

Not only the developer of the system but also the users can be involved in the development of the dialogue design [39, 40, 85, 86]. In those cases the focus of the interaction and the evaluation is not so much on the conversational abilities, but more on the interface and the user's interaction with it.

3 DISCUSSION

Our survey of literature that details DM in the context of HRI research revealed a variety of used DM approaches, influenced by the task the robot should perform and the needed reliability and capabilities. In this section, we will discuss the key challenges that have to be faced to make progress in DM for social robots.

Types of dialogue managers. For task-based human–robot dialogues, the majority of systems are using handcrafted approaches (see Section 2.3), among which model-based approaches are the most popular (see Figure 4). End-to-end approaches, like they are used for non-HRI task-based DM [112], are mainly used in hybrid approaches in task-based HRI. The resources needed for designing complicated dialogue managers constitute a challenge. For handcrafted approaches where the structure has to be defined *a priori*, human effort is needed for the creation of the dialogue. In contrast to handcrafted approaches, data-driven dialogue managers need less human effort but rely on data for learning conversational behavior, especially if it is combined with multimodal output. However, data collection in HRI is expensive [29], and expert feedback is needed for supervised learning [64].

Transferability from human–human conversation. In human–human conversations, common ground plays an important role and helps to reduce miscommunication [100]. In spoken HRI this is even more challenging since the perception of humans and robots is not the same, since robots might not be able to perceive the same modalities as a human. Expectations learned from human–human conversations do not necessarily hold for human–robot conversations. This can lead to a gap between the perceived versus the real common ground [21]. The physical presence of robots has an effect on the interaction [61]; however, it is not clear what exactly those influences look like for the different combinations as illustrated in Figure 2. This makes it challenging to transfer knowledge gained from human–human or even human–agent interaction directly to HRI.

Interaction domains. While the possible interaction domains for DM in HRI span a wide area and can be combined with varying robots and environments (see Figure 2), most of them are task based, for example, for deliveries or navigation. Avoiding unpredictable robot behavior seems to be an important factor for DM in HRI, which makes handcrafted approaches a compelling option, especially in task-based interaction domains, even though it does not provide the same flexibility as data-driven approaches. One trend we observed is that hybrid approaches divide the dialogue manager into smaller expert modules that are responsible for specific tasks within the dialogue manager [57, 75]. Different experts use different DM strategies and approaches themselves, based on the task that is assigned to them. This overall strategy of distributing the task to experts leads to more flexibility in the construction of the interaction, since the parts that are responsible for different tasks can use the approach that is best for that specific task.

Physical situatedness and embodiment. A capability that is specific to DM in robots is the grounding of the dialogue in the physical environment. For DM in HRI, the user and the robot are both physically situated in the environment, which can take different forms (see Figure 2). Because of this, they are exposed to changes, for example, due to people moving or objects being moved. A robot can encounter new or already known people [30] or unknown objects [83] during the conversation. This also relates to the question about endowing robots with a memory. More specifically, the physical situatedness of robots raises the question of how to manage such environmental dynamics in a robot's memory. Due to their situatedness in the real world, interactions can happen more than once or for a longer time interval with the same robot. In those interactions repetitions should be avoided [91]. The question then becomes how to endow a robot with an effective memory to sustain such long-term interactions.

Due to its physical presence in the environment, the robot's appearance has to be taken into account (see Figure 2). Due to its embodiment, the robot can manipulate its environment. Those robot actions and the changes in the environment that arise because of them need to be reflected in the dialogue.

When talking to a robot, the robot's physical appearance raises expectations about conversational capabilities, like in the case of Android robots that look similar to humans, leading to high expectations of their conversational capabilities [57, 75]. While the appearance influences expectations about capabilities, it does not make explicit which ones are actually present, and users can easily become frustrated when those expectations are not met. Therefore, to understand the robot's behavior and decisions, additional explanations will be required [27, 95]. To avoid future failures, the robot should be transparent about the reason for a failure when it happens, so that the user can try to avoid it in the future. The implications the chosen robot morphology has on the dialogue are rarely discussed in the surveyed papers. Without this information, it is difficult to judge if and how the obtained results are transferable to a different robot. If the choice of robot and the resulting effects on the interaction were more commonly discussed in research papers, this would help other researchers in choosing an appropriate robot appearance for their interactions.

DM as a part of a bigger (multimodal) system. Dialogue managers that are integrated into a robot are part of a bigger system. Social robots are typically multimodal and therefore depend on modules, such as sensors, while influencing other modules. Which modalities are included varies from case to case, as indicated in Figure 2. By influencing the actuators of the robot, the dialogue can have an immediate impact on the environment. The dialogue manager not only interacts with the NLU and NLG modules but also is often linked to other modalities that, for example, manage the perception and production of multi-modal cues, like gestures or gaze. The dialogue manager is in addition influenced by the problems of certain modules, such as speech recognition errors [34, 49, 57, 90]. Apart from improving the speech recognition module directly, it is possible to include mechanisms in the dialogue manager that are responsible for dealing with speech recognition errors [34]. Due to speech recognition difficulties that are especially common in in-the-wild studies, multiple common capabilities focus on types of repair (see Section 2.3). However, while repairing problems in speech recognition with DM techniques is an option, making the reasons for the failures explicit can help to improve the speech recognition in an HRI context. Speech recognition in HRI is impacted by noise from the robot's motor and the environment the robot is placed in, especially during experiments outside of the lab [94, 97]. Robots can encounter multiple people, both at the same time and at different times, whose age, accent, and way of speaking can vary. All of those factors influence the performance of the speech recognition module.

During conversations, it is problematic if the robot loses track of the user or is ignoring multimodal cues [63] while the human expects the robot to be able to process them. Since the dialogue manager is integrated into the robot's architecture, the whole system should be taken into account for the evaluation. This means that it is not enough to evaluate the dialogue manager in a decoupled manner. Rather, it is necessary to consider the effects of the integration into the robot. Using real users for evaluations comes with the advantage that it provides a clearer picture of how the system performs in actual interactions.

Recommendations and conclusion. Several of the challenges we outlined relate to the variance of interaction domains, appearances, modalities, and environments in HRI that need to be incorporated into the DM approach. One of these challenges is the integration into a system with multiple components, since problems in the other modules that relate to the multi-modal features of several robot types can impact the DM and need to be accounted for. Another challenge is the lack of datasets for HRI to further evolve probabilistic and hybrid approaches to DM in HRI. The data requirements are related not only to the domain of the interaction but also to the type of robot and the location of the interaction, since those can impact the conversation. The problem of lack of data could be addressed by using large language models that only have to be fine-tuned or can be directly prompted including recent turns in the conversation. However, they still have difficulties with situational awareness and the integration of the robot's sensors and actuators [11, 110]. Human–robot datasets cannot easily be substituted by human–human datasets, since robots have different abilities than humans and lack most of the common-sense abilities humans have. Datasets, moreover, even need to take specific robot shapes and forms into account, as the capabilities present in different robot platforms differ from each other. It is therefore important to maximize transparency about the robot's conversational capabilities.

HRI researchers new to DM can make an informed decision regarding the selection of a DM approach by assessing which of the presented approaches fits their requirements and limitations best. Based on our observations, hybrid approaches are a good option for more complex dialogue managers where the dialogue is an essential part of the interaction. If the interaction is simpler from a dialogue perspective or not enough data are available, handcrafted approaches, especially model-based ones, are a viable option. While a variety of robots with a range of different shapes and appearances has been used for spoken HRI so far, it would be helpful if the implications of robot platform choices on the dialogue would be discussed more extensively. Especially due to the situatedness in the environment, the required multimodal inputs need to be taken into account in advance. As a start, we would suggest to include only those that are necessary for the interaction so as to not overly complicate the DM system. Using already used evaluation metrics, especially those that are easy to record, would help to compare the performance of the systems used.

Even though dialogue is common in HRI, it is rarely the main focus of the interaction design. In current research, dialogue is often only seen as a tool to achieve a task with a robot. Moreover, in papers that do take a DM perspective, the dialogue manager is typically evaluated in non-embodied agents, which neglects robot-specific challenges that need to be addressed. In order to more structurally address these challenges, it is important that the best of both fields is combined to develop a more standardized approach for DM that can be drawn upon in the diverse interaction scenarios that arise in HRI studies.

ACKNOWLEDGMENTS

This research was (partially) funded by the Hybrid Intelligence Center, a 10-year programme funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research, https://hybridintelligence-centre.nl, grant number 024.004.022. We thank our anonymous reviewers for their constructive and helpful feedback, which enabled us to improve our paper.

REFERENCES

[1] Wendy Aguilar and Luis A. Pineda. 2009. Integrating graph-based vision perception to spoken conversation in human-robot interaction. In *Bio-Inspired Systems: Computational and Ambient Intelligence*, Joan Cabestany, Francisco Sandoval, Alberto Prieto, and Juan M. Corchado (Eds.). Vol. 5517. Springer, Berlin, 789–796. https://doi.org/10.1007/978-3-642-02478-8_99 Series Title: Lecture Notes in Computer Science.

[2] Fernando Alonso-Martín, Aívaro Castro-González, Francisco Javier Fernandez de Gorostiza Luengo, and Miguel Ángel Salichs. 2015. Augmented robotics dialog system for enhancing human–robot interaction. *Sensors* 15, 7 (2015), 15799–15829.

[3] Fernando Alonso-Martín, Javier F. Gorostiza, María Malfaz, and Miguel A. Salichs. 2013. Multimodal fusion AS communicative acts during human-robot interaction. *Cybernetics and Systems* 44, 8 (Nov. 2013), 681–703. https://doi.org/10.1080/01969722.2013.832096

[4] Amir Aly and Adriana Tapus. 2013. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI'13)*. 325–332. https://doi.org/10.1109/HRI.2013.6483606 ISSN: 2167-2148.

[5] Saeid Amiri, Sujay Bajracharya, Cihangir Goktolgal, Jesse Thomason, and Shiqi Zhang. 2019. Augmenting knowledge through statistical, goal-oriented human-robot dialog. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'19)*. IEEE, 744–750.

[6] Agnese Augello, Giuseppe Cittá, Manuel Gentile, and Antonio Lieto. 2023. A storytelling robot managing persuasive and ethical stances via ACT-R: An exploratory study. *International Journal of Social Robotics* 15, 12 (Dec. 2023), 2115–2131. https://doi.org/10.1007/s12369-021-00847-w

[7] Héctor H. Avilés, Iván V. Meza, Wendy Aguilar, and Luis Alberto Pineda. 2010. Integrating pointing gestures into a Spanish-spoken dialog system for conversational service robots. *International Conference on Agents and Artificial Intelligence* Vol. 2. SCITEPRESS, 2010, 585–588.

[8] Kim Baraka, Patrícia Alves-Oliveira, and Tiago Ribeiro. 2020. An extended framework for characterizing social robots. In *Human-robot Interaction*. Springer, 21–64.

[9] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81.

[10] Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018. Social robots for education: A review. *Science Robotics* 3, 21 (2018), eaat5954.

[11] Erik Billing, Julia Rosén, and Maurice Lamb. 2023. Language models for human-robot interaction. In *ACM/IEEE International Conference on Human-robot Interaction*. ACM Digital Library, 905–906.

[12] Daniel G. Bobrow, Ronald M. Kaplan, Martin Kay, Donald A. Norman, Henry Thompson, and Terry Winograd. 1977. GUS, a frame-driven dialog system. *Artificial Intelligence* 8, 2 (1977), 155–173.

[13] Dan Bohus and Eric Horvitz. 2014. Managing human-robot engagement with forecasts and... um... hesitations. In *Proceedings of the 16th International Conference on Multimodal Interaction*. 2–9.

[14] Adriana Bono, Agnese Augello, Giovanni Pilato, Filippo Vella, and Salvatore Gaglio. 2020. An ACT-R based humanoid social robot to manage storytelling activities. *Robotics* 9, 2 (April 2020), 25. https://doi.org/10.3390/robotics9020025

[15] Hayet Brabra, Marcos Báez, Boualem Benatallah, Walid Gaaloul, Sara Bouguelia, and Shayan Zamanirad. 2021. Dialogue management in conversational systems: A review of approaches, challenges, and opportunities. *IEEE Transactions on Cognitive and Developmental Systems* 14, 3 (2021), 783–798.

[16] Cynthia Breazeal, Kerstin Dautenhahn, and Takayuki Kanda. 2016. Social robotics. In *Springer Handbook of Robotics*, 1935–1972.

[17] Cynthia Breazeal, Nick DePalma, Jeff Orkin, Sonia Chernova, and Malte Jung. 2013. Crowdsourcing human-robot interaction: New methods and system evaluation in a public environment. *Journal of Human-Robot Interaction Steering Committee* 2, 1 (March 2013), 82–111. https://doi.org/10.5898/JHRI.2.1.Breazeal

[18] Joost Broekens, Marcel Heerink, and Henk Rosendal. 2009. Assistive social robots in elderly care: A review. *Gerontechnology* 8, 2 (2009), 94–103.

[19] Joana Campos, James Kennedy, and Jill F. Lehman. 2018. Challenges in exploiting conversational memory in human-agent interaction. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 1649–1657.

[20] Birte Carlmeyer, David Schlangen, and Britta Wrede. 2014. Towards closed feedback loops in HRI: Integrating Inprotk and Pamini. In *Proceedings of the 2014 Workshop on Multimodal, Multi-party, Real-world Human-robot Interaction*. 1–6.

[21] Joyce Y. Chai, Lanbo She, Rui Fang, Spencer Ottarson, Cody Littley, Changsong Liu, and Kenneth Hanson. 2014. Collaborative effort towards common ground in situated human-robot dialogue. In *2014 9th ACM/IEEE International Conference on Human-robot Interaction (HRI'14)*. IEEE, 33–40.

[22] Crystal Chao and Andrea Thomaz. 2016. Timed Petri nets for fluent turn-taking over multimodal interaction resources in human-robot collaboration. *International Journal of Robotics Research* 35, 11 (Sept. 2016), 1330–1353. https://doi.org/10.1177/0278364915627291

[23] Xiaoping Chen, Jianmin Ji, Jiehui Jiang, Guoqiang Jin, Feng Wang, and Jiongkun Xie. 2010. Developing high-level cognitive functions for service robots. In *AAMAS*, Vol. 10. 989–996.

[24] Heriberto Cuayáhuitl, Ivana Kruijff-Korbayová, and Nina Dethlefs. 2014. Nonstrict hierarchical reinforcement learning for interactive systems and robots. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 4, 3 (2014), 1–30.

[25] Jan Deriu, Alvaro Rodrigo, Arantxa Otegi, Guillermo Echegoyen, Sophie Rosset, Eneko Agirre, and Mark Cieliebak. 2021. Survey on evaluation methods for dialogue systems. *Artificial Intelligence Review* 54, 1 (2021), 755–810.

[26] Laurence Devillers, Sophie Rosset, Guillaume Dubuisson Duplessis, Mohamed A. Sehili, Lucile Béchade, Agnés Delaborde, Clement Gossart, Vincent Letard, Fan Yang, Yücel Yemez, Bekir B. Türker, Metin Sezgin, Kevin El Haddad, Stéphane Dupont, Daniel Luzzati, Yannick Esteve, Emer Gilmartin, and Nick Campbell. 2015. Multimodal data collection of human-robot humorous interactions in the Joker project. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII'15)*. 348–354. https://doi.org/10.1109/ACII.2015.7344594 ISSN: 2156-8111.

[27] Francesca Dino, Rohola Zandie, Hojjat Abdollahi, Sarah Schoeder, and Mohammad H. Mahoor. 2019. Delivering cognitive behavioral therapy using a conversational social robot. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'19)*. IEEE, 2089–2095.

[28] Malcolm Doering, Dylan F. Glas, and Hiroshi Ishiguro. 2019. Modeling interaction structure for robot imitation learning of human social behavior. *IEEE Transactions on Human-machine Systems* 49, 3 (June 2019), 219–231. https://doi.org/10.1109/THMS.2019.2895753

[29] Finale Doshi and Nicholas Roy. 2007. Efficient model learning for dialog management. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*. 65–72.

[30] M. M. S. N. Edirisinghe, M. A. V. J. Muthugala, and A. G. B. P. Jayasekara. 2018. Application of robot autobiographical memory in long-term human-robot social interactions. In *2018 2nd International Conference On Electrical Engineering (EECon'18)*. IEEE, 138–143.

[31] Rui Fang, Malcolm Doering, and Joyce Y. Chai. 2015. Embodied collaborative referring expression generation in situated human-robot interaction. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-robot Interaction*. ACM, 271–278. https://doi.org/10.1145/2696454.2696467

[32] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2002. A survey of socially interactive robots: Concepts, design, and applications. *Robotics and Autonomous Systems* 42, 3–4 (2002).

[33] Mary Ellen Foster, Andre Gaschler, Manuel Giuliani, Amy Isard, Maria Pateraki, and Ronald P. A. Petrick. 2012. Two people walk into a bar: Dynamic multi-party social interaction with a robot agent. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*. 3–10.

[34] Kotaro Funakoshi, Mikio Nakano, Toyotaka Torii, Yuji Hasegawa, Hiroshi Tsujino, Noriyuki Kimura, and Naoto Iwahashi. 2007. Robust acquisition and recognition of spoken location names by domestic robots. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1435–1440.

[35] Felix Gervits, Anton Leuski, Claire Bonial, Carla Gordon, and David Traum. 2021. A classification-based approach to automating human-robot dialogue. In *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems*. Springer Singapore, 115–127.

[36] Felix Gervits and Matthias Scheutz. 2018. Pardon the interruption: Managing turn-taking through overlap resolution in embodied artificial agents. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 99–109.

[37] Felix Gervits, Ravenna Thielstrom, Antonio Roque, and Matthias Scheutz. 2020. It's about time: Turn-entry timing for situated human-robot dialogue. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 86–96.

[38] Petra Gieselmann and Mari Ostendorf. 2007. Problem-sensitive response generation in human-robot dialogs. In *Proceedings of the 8th SIGDial Workshop on Discourse and Dialogue*. 219–222.

[39] Dylan Glas, Satoru Satake, Takayuki Kanda, and Norihiro Hagita. 2012. An interaction design framework for social robots. In *Robotics: Science and Systems*, Vol. 7. 89.

[40] Dylan F. Glas, Takayuki Kanda, and Hiroshi Ishiguro. 2016. Human-robot interaction design using interaction composer: Eight years of lessons learned. In *The 11th ACM/IEEE International Conference on Human Robot Interaction (HRI'16)*. IEEE Press, 303–310.

[41] Javi F. Gorostiza and Miguel A. Salichs. 2010. Natural programming of a social robot by dialogs. In *2010 AAAI Fall Symposium Series*.

[42] JinGuang Han, Emer Gilmartin, and Nick Campbell. 2013. Herme, yet another interactive conversational robot. In *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. 711–712. https://doi.org/10.1109/ACII.2013.127 ISSN: 2156-8111.

[43] Jan-Gerrit Harms, Pavel Kucherbaev, Alessandro Bozzon, and Geert-Jan Houben. 2018. Approaches for dialog management in conversational agents. *IEEE Internet Computing* 23, 2 (2018), 13–22.

[44] Sachithra Hemachandra and Matthew R. Walter. 2015. Information-theoretic dialog to improve spatial-semantic representations. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*. IEEE, 5115–5121.

[45] Sachithra Hemachandra, Matthew R. Walter, and Seth Teller. 2014. Information theoretic question asking to improve spatial semantic representations. In *2014 AAAI Fall Symposium Series*.

[46] Hartwig Holzapfel. 2005. Towards development of multilingual spoken dialogue systems. In *Proceedings of the 2nd Language and Technology Conference*.

[47] Hartwig Holzapfel. 2008. A dialogue manager for multimodal human-robot interaction and learning of a humanoid robot. *Industrial Robot: An International Journal* 35, 6 (Oct. 2008), 528–535. https://doi.org/10.1108/01439910810909529

[48] Takamasa Iio, Yuichiro Yoshikawa, Mariko Chiba, Taichi Asami, Yoshinori Isoda, and Hiroshi Ishiguro. 2020. Twin-robot dialogue system with robustness against speech recognition failure in human-robot dialogue with elderly people. *Applied Sciences* 10, 4 (2020), 1522.

[49] Koji Inoue, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara. 2020. An attentive listening system with android ERICA: Comparison of autonomous and WOZ interactions. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 118–127.

[50] Radu Ion, Valentin Gabriel Badea, George Cioroiu, Verginica Barbu Mititelu, Elena Irimia, Maria Mitrofan, and Dan Tufis. 2020. A dialog manager for micro-worlds. *Studies in Informatics and Control* 29, 4 (2020), 411–420.

[51] Ryan Blake Jackson and Tom Williams. 2022. Enabling morally sensitive robotic clarification requests. *ACM Transactions on Human-robot Interaction* 11, 2 (June 2022), 1–18. https://doi.org/10.1145/3503795

[52] Ridong Jiang, Yeow Kee Tan, Dilip Kumar Limbu, Tran Anh Tung, and Haizhou Li. 2011. A configurable dialogue platform for ASORO robots. In *Proceedings of APSIPA ASC*.

[53] Martin Johansson and Gabriel Skantze. 2015. Opportunities and obligations to take turns in collaborative multiparty human-robot interaction. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 305–314.

[54] Andreas Kipp and Franz Kummert. 2014. Dynamic dialog system for human robot collaboration: Playing a game of pairs. In *Proceedings of the 2nd International Conference on Human-agent Interaction*. 225–228.

[55] Laura Kunold, Nikolai Bock, and Astrid Rosenthal-von der Pütten. 2023. Not all robots are evaluated equally: The impact of morphological features on robots' assessment through capability attributions. *ACM Transactions on Human-Robot Interaction* 12, 1 (2023), 1–31.

[56] Divesh Lala, Koji Inoue, and Tatsuya Kawahara. 2018. Evaluation of real-time deep learning turn-taking models for multiple dialogue scenarios. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. 78–86.

[57] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. 2017. Attentive listening system with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*. 127–136.

[58] Changyoon Lee, You-Sung Cha, and Tae-Yong Kuc. 2008. Implementation of dialogue system for intelligent service robots. In *2008 International Conference on Control, Automation and Systems*. IEEE, 2038–2042.

[59] Cheongjae Lee, Sangkeun Jung, Kyungduk Kim, and Gary Geunbae Lee. 2010. Hybrid approach to robust dialog management using agenda and dialog examples. *Computer Speech & Language* 24, 4 (2010), 609–631.

[60] Sau-lai Lee, Ivy Yee-man Lau, Sara Kiesler, and Chi-Yue Chiu. 2005. Human mental models of humanoid robots. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE, 2767–2772.

[61] Jamy Li. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-computer Studies* 77 (2015), 23–37.

[62] Shuyin Li, Britta Wrede, and Gerhard Sagerer. 2006. A computational model of multi-modal grounding for human robot interaction. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue (SigDIAL'06)*. Association for Computational Linguistics, 153. https://doi.org/10.3115/1654595.1654626

[63] Shuyin Li, Britta Wrede, and Gerhard Sagerer. 2006. A dialog system for comparative user studies on robot verbal behavior. In *The 15th IEEE International Symposium on Robot and Human Interactive Communication (ROMAN'06)*. IEEE, 129–134.

[64] Pierre Lison. 2012. Probabilistic dialogue models with prior domain knowledge. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 179–188.

[65] Pierre Lison. 2013. Model-based Bayesian reinforcement learning for dialogue management. *arXiv preprint arXiv:1304.1819* (2013).

[66] Pierre Lison. 2015. A hybrid approach to dialogue management based on probabilistic rules. *Computer Speech & Language* 34, 1 (2015), 232–255.

[67] Pierre Lison and Casey Kennington. 2023. Who's in charge? Roles and responsibilities of decision-making components in conversational robots. arXiv:2303.08470 (March 2023). http://arxiv.org/abs/2303.08470 arXiv:2303.08470 [cs].

[68] Phoebe Liu, Dylan F. Glas, Takayuki Kanda, and Hiroshi Ishiguro. 2016. Data-driven HRI: Learning social behaviors by example from human-human interaction. *IEEE Transactions on Robotics* 32, 4 (Aug. 2016), 988–1008. https://doi.org/10.1109/TRO.2016.2588880

[69] Rui Liu and Xiaoli Zhang. 2019. A review of methodologies for natural-language-facilitated human–robot cooperation. *International Journal of Advanced Robotic Systems* 16, 3 (2019), 1729881419851402.

[70] Dongcai Lu, Shiqi Zhang, Peter Stone, and Xiaoping Chen. 2017. Leveraging commonsense reasoning and multimodal perception for robot spoken dialog systems. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'17)*. 6582–6588. https://doi.org/10.1109/IROS.2017.8206570 ISSN: 2153-0866.

[71] Lorenzo Lucignano, Francesco Cutugno, Silvia Rossi, and Alberto Finzi. 2013. A dialogue system for multimodal human-robot interaction. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*. ACM, 197–204. https://doi.org/10.1145/2522848.2522873

[72] Matthew Marge, Carol Espy-Wilson, Nigel G. Ward, Abeer Alwan, Yoav Artzi, Mohit Bansal, Gil Blankenship, Joyce Chai, Hal Daumé III, Debadeepta Dey, et al. 2022. Spoken language interaction with robots: Recommendations for future research. *Computer Speech & Language* 71 (2022), 101255.

[73] Yoichi Matsuyama, Hikaru Taniyama, Shinya Fujie, and Tetsunori Kobayashi. 2010. Framework of communication activation robot participating in multiparty conversation. In *2010 AAAI Fall Symposium Series*.

[74] Nikolaos Mavridis. 2015. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems* 63 (2015), 22–35.

[75] Pierrick Milhorat, Divesh Lala, Koji Inoue, Tianyu Zhao, Masanari Ishida, Katsuya Takanashi, Shizuka Nakamura, and Tatsuya Kawahara. 2019. A conversational dialogue manager for the humanoid robot ERICA. In *Advanced Social Interaction with Agents*. Springer, 119–131.

[76] David Moher, Larissa Shamseer, Mike Clarke, Davina Ghersi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, and Lesley A. Stewart. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 4, 1 (2015), 1–9.

[77] Mikio Nakano, Naoto Iwahashi, Takayuki Nagai, Taisuke Sumii, Xiang Zuo, Ryo Taguchi, Takashi Nose, Akira Mizutani, Tomoaki Nakamura, Muhanmad Attamim, et al. 2010. Grounding new words on the physical world in multi-domain human-robot dialogues. In *2010 AAAI Fall Symposium Series*.

[78] Carlo Nuccio, Agnese Augello, Salvatore Gaglio, and Giovanni Pilato. 2018. Interaction capabilities of a robotic receptionist. In *Intelligent Interactive Multimedia Systems and Services 2017 (Smart Innovation, Systems and Technologies)*, Giuseppe De Pietro, Luigi Gallo, Robert J. Howlett, and Lakhmi C. Jain (Eds.). Springer International Publishing, Cham, 171–180. https://doi.org/10.1007/978-3-319-59480-4_18

[79] Stanislav Ondas, Jozef Juhar, Matus Pleva, Peter Fercak, and Rastislav Husovsky. 2017. Multimodal dialogue system with NAO and VoiceXML dialogue manager. In *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom'17)*. IEEE, 000439–000444. https://doi.org/10.1109/CogInfoCom.2017.8268286

[80] Nurziya Oralbayeva, Amir Aly, Anara Sandygulova, and Tony Belpaeme. 2023. Data-driven communicative behaviour generation: A survey. *ACM Transactions on Human-Robot Interaction* 13, 1 (2023), 1–39.

[81] Mourad Ouzzani, Hossam Hammady, Zbys Fedorowicz, and Ahmed Elmagarmid. 2016. Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews* 5, 1 (2016), 210. https://doi.org/10.1186/s13643-016-0384-4

[82] Julia Peltason and Britta Wrede. 2010. Pamini: A framework for assembling mixed-initiative human-robot interaction from generic interaction patterns. In *Proceedings of the SIGDIAL 2010 Conference*. 229–232.

[83] Julia Peltason and Britta Wrede. 2012. Structuring human-robot-interaction in tutoring scenarios. In *Towards Service Robots for Everyday Environments*. Springer, 471–482.

[84] Ronald P. A. Petrick, Mary Ellen Foster, and Amy Isard. 2012. Social state recognition and knowledge-level planning for human-robot interaction in a bartender domain. In *Workshops at the 26th AAAI Conference on Artificial Intelligence*.

[85] David Porfirio, Allison Sauppé, Aws Albarghouthi, and Bilge Mutlu. 2018. Authoring and verifying human-robot interactions. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, 75–86. https://doi.org/10.1145/3242587.3242634

[86] David Porfirio, Allison Sauppé, Aws Albarghouthi, and Bilge Mutlu. 2020. Transforming robot programs based on social context. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 1–12. https://doi.org/10.1145/3313831.3376355

[87] Thomas Prommer, Hartwig Holzapfel, and Alex Waibel. 2006. Rapid simulation-driven reinforcement learning of multimodal dialog strategies in human-robot interaction. *Ninth International Conference on Spoken Language Processing* (2006).

[88] Marcelo Quinderé, L. Seabra Lopes, and António J. S. Teixeira. 2007. A dialogue manager for an intelligent mobile robot. *The 4th International Workshop on Natural Language Understanding and Cognitive Science* Vol. 2. SCITEPRESS, 2007.

[89] Marcelo Quinderé, Luís Seabra Lopes, and António J. S. Teixeira. 2007. An information state based dialogue manager for a mobile robot. In *8th Annual Conference of the International Speech Communication Association*. Citeseer.

[90] Marcelo Quinderé, Luís Seabra Lopes, and António J. S. Teixeira. 2013. Evaluation of a dialogue manager for a mobile robot. In *2013 IEEE RO-MAN*. IEEE, 126–132.

[91] Stephanie Rosenthal and Manuela Veloso. 2010. Mixed-initiative long-term interactions with an all-day-companion robot. In *2010 AAAI Fall Symposium Series*.

[92] Alessandro Russo, Grazia D'Onofrio, Aldo Gangemi, Francesco Giuliani, Misael Mongiovi, Francesco Ricciardi, Francesca Greco, Filippo Cavallo, Paolo Dario, Daniele Sancarlo, et al. 2019. Dialogue systems and conversational agents for patients with dementia: The human–robot interaction. *Rejuvenation Research* 22, 2 (2019), 109–120.

[93] Paul E. Rybski, Kevin Yoon, Jeremy Stolarz, and Manuela M. Veloso. 2007. Interactive robot task training through dialog and demonstration. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*. 49–56.

[94] Satoru Satake, Kotaro Hayashi, Keita Nakatani, and Takayuki Kanda. 2015. Field trial of an information-providing robot in a shopping mall. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'15)*. IEEE, 1832–1839.

[95] Lanbo She, Shaohua Yang, Yu Cheng, Yunyi Jia, Joyce Chai, and Ning Xi. 2014. Back to the blocks world: Learning new actions through situated human-robot dialogue. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL'14)*. 89–97.

[96] Borui Shen and Diana Inkpen. 2016. Speech intent recognition for robots. In *2016 3d International Conference on Mathematics and Computers in Sciences and in Industry (MCSI'16)*. IEEE, 185–190.

[97] Masahiro Shiomi, Daisuke Sakamoto, Takayuki Kanda, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. 2008. A semi-autonomous communication robot: A field trial at a train station. In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*. 303–310.

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

[99] Gabriel Skantze, Anna Hjalmarsson, and Catharine Oertel. 2013. Exploring the effects of gaze and pauses in situated human-robot interaction. In *Proceedings of the SIGDIAL 2013 Conference*. 163–172.

[100] Robert Stalnaker. 2002. Common ground. *Linguistics and Philosophy* 25, 5/6 (2002), 701–721.

[101] Komei Sugiura, Naoto Iwahashi, Hideki Kashioka, and Satoshi Nakamura. 2010. Active learning of confidence measure function in robot language acquisition framework. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1774–1779.

[102] Jesse Thomason, Shiqi Zhang, Raymond J. Mooney, and Peter Stone. 2015. Learning to interpret natural language commands through human-robot dialog. In *24th International Joint Conference on Artificial Intelligence*.

[103] Cristen Torrey, Aaron Powers, Matthew Marge, Susan R. Fussell, and Sara Kiesler. 2006. Effects of adaptive robot dialogue on information exchange and social relations. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*. 126–133.

[104] David R. Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In *Current and New Directions in Discourse and Dialogue*. Springer, 325–353.

[105] H. Trung. 2006. *Multimodal Dialogue Management-State of the Art*. Human Media Interaction Department, University of Twente.

[106] Daniel Couto Vale and Vivien Mast. 2013. Tacit social contracts for wheelchairs. In *Proceedings of the SIGDIAL 2013 Conference*. 294–303.

[107] Dimitrios Vogiatzis and Vangelis Karkaletsis. 2008. A framework for human-robot interaction. In *Proceedings of the 1st International Conference on PErvasive Technologies Related to Assistive Environments*. 1–7.

[108] Tom Williams, Gordon Briggs, Bradley Oosterveld, and Matthias Scheutz. 2015. Going beyond literal command-based instructions: Extending robotic natural language interaction capabilities. In *29th AAAI Conference on Artificial Intelligence*.

[109] Jie Zeng, Yukiko I. Nakano, Takeshi Morita, Ichiro Kobayashi, and Takahira Yamaguchi. 2018. Eliciting user food preferences in terms of taste and texture in spoken dialogue systems. In *Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-food Interaction*. 1–5.

[110] Bowen Zhang and Harold Soh. 2023. Large language models as zero-shot human models for human-robot interaction. *arXiv preprint arXiv:2303.03548* (2023).

[111] Shiqi Zhang and Peter Stone. 2015. CORPP: Commonsense reasoning and probabilistic planning, as applied to dialog with a mobile robot. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29.

[112] Yin Jiang Zhao, Yan Ling Li, and Min Lin. 2019. A review of the research on dialogue management of task-oriented systems. *Journal of Physics: Conference Series* 1267 (2019), 012025.

Received 18 October 2022; revised 5 February 2024; accepted 8 February 2024