University of Glasgow

VIA VERITAS VITA

Foster, M. E. (2019) Natural language generation for social robotics: Opportunities and challenges. Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1771), 20180027.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/178908/

Deposited on: 29 January 2019

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk

Natural Language Generation for Social Robotics: Opportunities and Challenges

Mary Ellen Foster

School of Computing Science, University of Glasgow, 18 Lilybank Gardens, Glasgow G12 8RZ, UK
MaryEllen.Foster@glasgow.ac.uk

ABSTRACT

In the increasingly popular and diverse research area of social robotics, the primary goal is to develop robot agents that exhibit socially intelligent behaviour while interacting in a face-to-face context with human partners. An important aspect of face-to-face social conversation is fluent, flexible linguistic interaction: as Bavelas et al. [1] point out, face-to-face dialogue is both the basic form of human communication and the richest and most flexible, combining unrestricted verbal expression with meaningful non-verbal acts such as gestures and facial displays, along with instantaneous, continuous collaboration between the speaker and the listener. In practice, however, most developers of social robots tend not to use the full possibilities of the unrestricted verbal expression afforded by face-to-face conversation; instead, they generally tend to employ relatively simplistic processes for choosing the words for their robots to say. This contrasts with the work carried out Natural Language Generation (NLG), the field of computational linguistics devoted to the automated production of high-quality linguistic content: while this research area is also an active one, in general most effort in NLG is focussed on producing high-quality written text. This article summarises the state-of-the-art in the two individual research areas of social robotics and natural language generation. It then discusses the reasons why so few current social robots make use of more sophisticated generation techniques. Finally, an approach is proposed to bringing some aspects of NLG into social robotics, concentrating on techniques and tools that are most appropriate to the needs of socially interactive robots.

1 Introduction

In popular culture, the prototypical "robot" is a shiny, metallic, human-shaped being that is able to engage fully in all aspects of face-to-face conversation—a canonical version of this would be the protocol droid C-3PO from the Star Wars movies (Figure 1). As modern robot hardware becomes safer, more sophisticated, and more generally available, there is a clear and significant consumer demand for this sort of *socially intelligent* (if not necessarily humanoid) robot [2] that is able to interact flexibly in everyday human environments. For example, SoftBank's humanoid robot "Pepper" (Figure 2a) has so far been deployed in

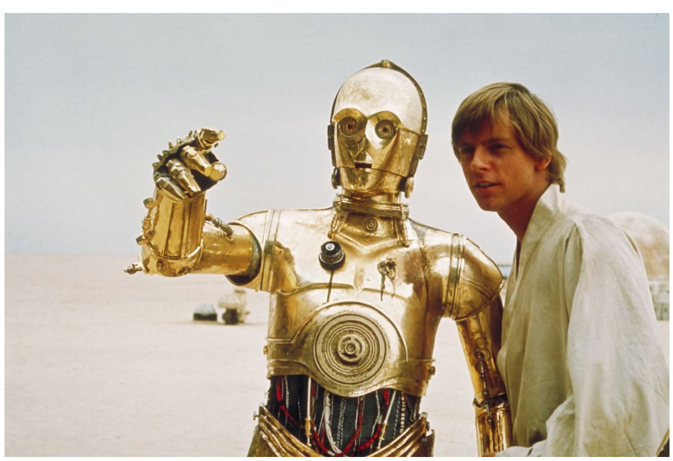

Figure 1. A fictional social robot. Image © 1977 Lucasfilm Ltd.

(a) Pepper **(b)** Cozmo and Vector

Figure 2. Commercially-available social robot platforms

well over 10,000 locations worldwide, while Anki's Cozmo and Vector robots (Figure 2b) have recently become best-sellers in the educational robotics market. Social robots are now being deployed in a wide range of contexts: in public spaces [3–5], in educational settings [6, 7], and in private homes [8, 9].

Developing and deploying a socially interactive robot presents a number of significant technical challenges. Some robots have been able to provide social interaction with minimal functionality—for example, the seal robot Paro, which responds very simply to user touch, has been used to encourage social interaction in elderly populations [10]. However, if a robot is to engage in more sophisticated social interactions, it must incorporate state-of-the-art components for audiovisual processing, social signal processing, action selection, and robot navigation and motion planning—and without these systems, the robot will not be able to carry out even its basic functions. Largely due to the complexity of these necessary tasks, generation of verbal output does not tend to be a priority for social robot developers: since a template-based approach is often sufficient in the short term, most robot system designers choose such a language-generation solution and focus their effort elsewhere.[1] Indeed, of the 51 papers in the proceedings of the 2017 ACM Human-Robot Interaction conference (the top-tier conference in the field), only one reported using any form of NLG beyond the basics [11]. Similarly, at HRI 2018, systems that involved linguistic interaction generally used either "Wizard-of-Oz" techniques (where a human operator manually chooses the system output), or else used template-based techniques or prerecorded speech.

In this paper, we argue that the time is right for developers of social robots to explore the output possibilities provided by incorporating state-of-the-art NLG. This paper begins with a summary of the state-of-the-art in both individual research areas, and then discusses how insights from NLG can improve interactions with social robots, as well as how applications in social robotics can benefit researchers in natural language generation.

2 Background: Natural Language Generation[2]

Natural Language Generation—usually abbreviated as NLG—is the sub-area of computational linguistics that deals with the automated production of high-quality spoken or written content in human languages [12–14]. While the output of an NLG system is text, the input can take various forms: in some cases, the system might generate text based on other, generally human-written text: applications for this include automated machine translation, summarisation, and generating simplified or paraphrased versions of its input. In other cases, the input to the NLG system is non-linguistic, for example football reports, data from environmental sensors, or weather, financial, or medical data. More recent applications have included automatically generating text based on visual input such as images or video. Finally, NLG may be used not only to generate standalone texts, but also to generate linguistic output to be used in an interactive system, either in a text-based chatbot, a spoken dialogue system, or an interactive robot.

Traditionally, the task of NLG has been divided into the following six sub-tasks [13]:

- Content determination: deciding which information to include

[1] A recent example of this is "Olly", a social robot designed for the home market, which uses state-of-the-art techniques for all other aspects of its interaction, but notes that its NLG system is "currently template-based." https://www.thedatalab.com/case-studies/emotech-olly-the-robot

[2] This section is partially based on [12].

- Text struturing: deciding what order to present the information in
- Sentence aggregation: deciding what information to present in individual sentences
- Lexicalisation: choosing the words and phrases to express the information
- Referring expression generation: choosing the words and phrases to express domain objects
- Linguistic realisation: combining words and phrases into sentences

While most NLG systems address the above tasks in some way, the organisation within the system can vary. For example, some systems use a modular approach, with clear divisions among tasks, while others use an approach based on AI planning, which provides a more integrated perspective. However, the most common approach in current NLG systems is to apply a more integrated approach, combining some or all tasks into a single decision process.

The techniques that may be used to create the linguistic output also vary across applications [15]. Simple systems may be developed using **templates**—that is, by slotting the content into pre-built linguistic structures. At its simplest, templates amount to slot-filling as in a "mail merge" context; templates can also be enhanced with conditional behaviour through scripting languages which might, for example, choose among different words to use depending on the context. Template-based output generation is relatively common in practice, particularly in systems that are deployed in real-world contexts, and can provide high-quality output; however, developing and maintaining the templates can require significant effort, particularly as new utterances are added to the system or if the system needs to support other interaction contexts or languages. Also, such a system cannot easily incorporate flexibility such as tailoring the output to different styles depending on the audience. A somewhat more complex approach is to use linguistically-motivated **rules**, or even full-fledged **grammars**, which are able to produce output in a more flexible and extensible way. For example, such a system might include rules for choosing the appropriate plural form of words ("children" rather than "childs") or for ensuring that grammatical gender and case are respected in languages that use them. However, developing an appropriate set of rules can still require effort similar to that needed for developing and maintaining templates, and also requires a level of linguistic knowledge in the developer to write rules at an appropriate level of generality. Note that the distinction between rules and templates is quite fluid in practice; often, linguistically sophisticated templates can be indistinguishable from "real" NLG [16].

As in several other areas of artificial intelligence, much of the recent progress in NLG has been driven by the increasing use of end-to-end, **data-driven techniques**, in particular various forms of machine learning. using such techniques replaces some or all of the manual effort involved in writing templates or rules by allowing the system to be trained on a large set of example target outputs [e.g., 17–20]: the system then directly learns the mapping between inputs and outputs without any need for explicit intermediate representations such as templates, rules, or grammars. A large proportion of the papers at the recent International Language Generation Conference [21]—which provides a snapshot of the state-of-the-art in this area—use such statistical data-driven techniques. A significant challenge in developing a data-driven NLG system is acquiring an appropriate amount of training data for creating the necessary statistical models. Various apporaches have been taken to this problem: for example, using existing weather reports written by experts [22], using a crowdsourcing approach where humans create sample outputs tailored to a particular domain [23], or—in the case of deep-learning approaches such as those currently used for tasks such as the generation of image captions—simply providing a large corpus of more loosely coupled inputs and outputs [24].

In summary, the main task of NLG is to generate high-quality linguistic output. Early work in NLG focussed on largely hand-coded, rule-based systems; however, more recent work in the area has begun to apply end-to-end, data-driven techniques. While such data-driven techniques have been successful in a range of contexts, they require significant amounts of training data, meaning that the most common applications for data-driven NLG are those where the input can be relatively clearly specified, such as data-to-text systems, caption-generation systems, or systems designed to generate descriptions of domain objects such as restaurants. Other techniques are still widely used, particularly in contexts (such as interactive systems) where the input is less well defined.

3 Background: Social Robotics

In 2003, Fong et al. [25] prepared a comprehensive survey of the then-emerging area of socially interactive robots, which they defined as "robots for which social interaction plays a key role." They focussed particularly on peer-to-peer human-robot interaction; that is, where the robot exhibits human-like social skills such as expressing and/or perceiving emotions, using natural non-verbal cues like gaze and gestures, and possi/bly learning or developing various social competencies. They discussed a number of issues relevant to the then-developing field: approaches and challenges in the design of such robots;

issues related to the embodiment of a robot (physical shape, caricatured vs. realistic appearance, etc.); the role of emotion, personality, and user modelling in social robotics; along with approaches to human-oriented perception and learning.

In the decade since that survey was published, this active and growing area of social robotics—and HRI generally—has advanced considerably: the field now includes two dedicated journals [26, 27] and three conference series [28–30], along with numerous other discussion venues, confirming that it is an active and growing area of research. Indeed, interactive robots have been deployed and evaluated in a wide range of situations; recent examples include robots designed to be a socially-aware robot bartender [31], an empathy-invoking conversational companion [32], a helper in a cooperative physical task [33], a mediator for remote negotiation teams [34], and an autonomous mobile helper in a shopping mall [35].

Techniques for evaluating interactive robots have also moved forwards considerably: for example, the Godspeed questionnaire series [36] has been widely used and validated as an instrument for user studies in HRI [37]. More recently, a novel evaluation instrument, the Robotic Social Attributes Scale (ROSaS) has been proposed to address some of the limitations found in the Godspeed questions [38]. Both of these questionnaires focus on measuring user subjective reactions to interacting with a social robot, focussing on the users' perceptions of features such as warmth/friendliness, competence/intelligence, human-likeness, as well as comfort or discomfort in the interaction.

Humans have a strong tendency to anthropomorphise robots and to want to engage in social interaction with them [39]. Developing a social robot that is to able to interact with humans in a real-word setting presents a large set of technical challenges. The robot must be able to navigate a populated space [40]; it must sense and respond to the non-verbal social signals of its human partners [41]; it must recognise and understand the users' speech; and, while not all social robots engage in verbal interaction, a large number do, meaning that supporting flexible natural-language interaction is a crucial task to increase the acceptability of social robots in the wider population. In the following section, we will discuss the role of NLG in current social robots, and will also outline ways that the two fields could come closer together in future.

4 NLG and Social Robotics

In their recent survey of the state-of-the-art in NLG research, Gatt and Krahmer [12] identify **situated language generation** as one of the main growth areas for the field, where situated language is defined as "language use in physical or virtual environments where production choices explicitly take into account perceptual and physical properties." Also, in response to the increased industrial interest in NLG for dialogue systems, a special session at the SIGDIAL 2017 conference was dedicated to this emerging topic [42], with a lively and rich panel discussion. Social robotics is a prime example of a situated context: when the robot and its human partner share a physical space, in some sense *all* language use is unavoidably situated by definition.

Previous research [e.g., 17, 43–45] has found that a spoken dialogue system that incorporates state-of-the-art NLG has the potential to greatly improve the quality of user-system interactions. However, largely due to the numerous other significant technical challenges involved in developing a socially intelligent robot, the majority of such robots tend to make use of extremely simple rule-based or template-based approaches to language generation, and do not benefit from the improved interactions and increased flexibility that are possible with more sophisticated techniques.

One significant area where modern NLG and social robotics research intersect is in the generation of referring expressions—for example, generating navigation instructions or referring to a particular object or location in the world. Referring expression generation is possibly the most thoroughly studied topic in NLG, perhaps primarily because it is one where both the input and the output are relatively straightforward to define (unlike many other, more open-ended tasks such as generating text with a particular style or generating persuasive text). The generation of situated references was the topic of the series of successful NLG shared tasks collectively called the GIVE challenges [46], where researchers competed to develop systems that were able to generate instructions to move around and interact with objects in a virtual, maze-like world. In fact, this is one area where social robotics researchers do tend to adopt similar techniques for output generation—e.g., [11, 47].

However, most social robot interactions involve much more than referring to world objects and filling slots for database queries [48]: they involve interactions situated in the real world, where the output of the robot should include both coverage of diverse topics as well as appropriate situated multimodal behaviour—obviously including linguistic content, but also incorporating non-verbal behaviours such as prosody and gesture. For example, in the MuMMER project [3], the goal is to place a Pepper robot in a shopping mall where it should provide help and guidance, carry out marketing activities, and also provide entertainment to mall visitors—and, crucially, it must do all of this while being socially intelligent and engaging. Supporting this range of behaviours requires generating a wide range of output types. In the current MuMMER prototype, a hybrid chatbot/task-based system is used for interaction management [49]—however, even in that context, the actual textual output is ultimately produced by templates, where information such as a particular movie or current news item is filled into a linguistic structure such as "My favourite is *X*" or "Have you heard about *Y*'.

It must be noted that templates are not always an invalid implementation decision for language generation [16]—indeed, skilfully written templates can provide a high degree of flexibility and expressiveness. In fact, the field of social robotics can draw some benefit from incorporating "traditional" NLG techniques such as rule-based or grammar-based processing: moving beyond the current solutions which mainly involve canned text or very simple templates is still likely to permit more socially intelligent interactions, particularly if the robot is deployed in new contexts or must interact in a different language. For example, open-source text realisers such as SimpleNLG [50] or OpenCCG [51] could be used to provide advantages such as flexibility and cross-lingual support.

In addition, just as in many other fields of natural language processing (and indeed AI), the NLG field is currently moving more and more towards data-driven techniques and sophisticated machine learning methods. The challenge for social robotics is to determine how to incorporate such methods into the sort of interactions needed for a particular deployment context.

For example, the current E2E ("End-to-End") NLG challenge [52] uses data drawn from a restaurant domain, where the input consists of a series of facts about a restaurant. Any system trained on this sort of data would perform well in a context where the robot must convey similar facts to its partner; however, in contexts such as a home companion robot, the robot utterances might be very different. In general, as with all data-intensive techniques, the issue is obtaining sufficient training data. It is not likely that current data-driven NLG techniques can solve all social robotics generation tasks—although given the success of such techniques, it would be still be instructive to learn how far they could be pushed.

A recent workshop at the INLG 2018 conference [53] brought together researchers from HRI and NLG to discuss areas of common interest. This workshop has confirmed that there is strong interest in the NLG community in applying NLG to a social robotics setting; it also confirmed that one of the main challenges in this area is defining a task within HRI where NLG can be shown to make a clear difference, and also making known to the developers of social robots the potential benefits of using a more principled approach to the generation of linguistic output. As part of this, a novel shared NLG task is currently under development that will be similar to the GIVE challenge, but will incorporate aspects of situated human-robot interaction.

5 Summary and Conclusions

In the context of social robotics, most developers tend to employ quite simplistic techniques for language generation, despite the advantages in flexibility and adaptability provided by NLG, as well as significant research from the related area of spoken dialogue systems that using NLG on the output side can also have a significant effect on users' subjective opinions of the system. While this choice tends to be made for pragmatic technical reasons, it is still the case that social roboticists are currently missing out on an important aspect of social interaction.

Social robotics—and HRI more generally—presents a particularly challenging and rich testbed for situated NLG, which is one of the identified growth areas for NLG as a whole. It is to be hoped that in future, the two research communities of social robotics and NLG can find a broader common ground, ideally resulting in mutually beneficial progress on both sides.

The key research and practical challenges in bringing these two areas together can be summarised as follows:

- What HRI application areas are particularly suitable for the application of NLG techniques?
- Can neural NLG techniques be applied in HRI? Are there areas where sufficient training data can be found to make this practical?
- How does the embodied and multimodal nature of HRI change the requirements for an NLG system?
- How can the NLG community convince the developers of interactive robots to consider approaches beyond templates when designing their systems?
- Is it possible to develop an NLG shared task inspired by the needs of an HRI system to push research in this area?

6 Acknowledgements

I would like to acknowledge my fellow organisers Hendrik Buschmeier and Dimitra Gkatzia and all of the participants in the INLG 2018 workshop on NLG for Human-Robot Interaction for the insightful discussions on this topic.

References

1. Bavelas, J. B., Hutchinson, S., Kenwood, C. & Matheson, D. H. Using face-to-face dialogue as a standard for other communication systems. *Can. J. Commun.* **22** (1997).
2. Dautenhahn, K. Socially intelligent robots: dimensions of human-robot interaction. *Philos. Transactions Royal Soc. B: Biol. Sci.* **362**, 679–704 (2007). DOI 10.1098/rstb.2006.2004.
3. Foster, M. E. *et al.* The MuMMER project: Engaging human-robot interaction in real-world public spaces. In *Proceedings of the 8th International Conference on Social Robotics (ICSR 2016)*, 753–763 (Kansas City, MO, USA, 2016). DOI 10.1007/978-3-319-47437-3_74.
4. Hawes, N. *et al.* The STRANDS project: Long-term autonomy in everyday environments. *IEEE Robotics Autom. Mag.* **24**, 146–156 (2017). DOI 10.1109/MRA.2016.2636359.
5. Sasaki, Y. & Nitta, J. Long-term demonstration experiment of autonomous mobile robot in a science museum. In *2017 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, 304–310 (2017). DOI 10.1109/IRIS.2017.8250139.
6. Belpaeme, T. *et al.* L2TOR: Second language tutoring using social robots. In *Proceedings of the ICSR 2015 WONDER Workshop* (2015).
7. DE-ENIGMA Project. DE-ENIGMA: Playfully empowering autistic children. In *Poster presented at the Autism-Europe International Congress* (2016).
8. Sidner, C. *et al.* A robotic companion for social support of isolated older adults. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 289 (2015). DOI 10.1145/2701973.2702103.
9. de Graaf, M., Ben Allouch, S. & van Dijk, J. Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 224–233 (2017). DOI 10.1145/2909824.3020236.
10. Kidd, C. D., Taggart, W. & Turkle, S. A sociable robot to encourage social interaction among the elderly. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, 3972–3976 (IEEE, 2006).
11. Daniele, A. F., Bansal, M. & Walter, M. R. Navigational instruction generation as inverse reinforcement learning with neural machine translation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 109–118 (2017). DOI 10.1145/2909824.3020241.
12. Gatt, A. & Krahmer, E. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *J. Artif. Intell. Res.* **61**, 65–170 (2018).
13. Reiter, E. & Dale, R. *Building Natural Language Generation Systems* (Cambridge University Press, 2000). DOI 10.2277/052102451X.
14. Evans, R., Piwek, P. & Cahill, L. What is nlg? In *Proceedings of INLG 2002*, 144–151 (2002).
15. Mairesse, F. Natural language generation (2005). http://farm2.user.srcf.net/research/papers/ART-NLG.pdf.
16. van Deemter, K., Theune, M. & Krahmer, E. Real versus template-based natural language generation: A false opposition? *Comput. Linguist.* **31**, 15–24 (2005). DOI 10.1162/0891201053630291.
17. Rieser, V., Lemon, O. & Keizer, S. Natural language generation as incremental planning under uncertainty. *IEEE/ACM Transactions on Audio, Speech, Lang. Process.* **22**, 979–994 (2014). DOI 10.1109/TASL.2014.2315271.
18. Li, X., van Deemter, K. & Lin, C. Statistics-based lexical choice for nlg from quantitative information. In *Proceedings of INLG 2016*, 104–108 (Edinburgh, UK, 2016).
19. Narayan, S. & Gardent, C. Unsupervised sentence simplification using deep semantics. In *Proceedings of INLG 2016*, 111–120 (Edinburgh, UK, 2016).
20. Dušek, O. & Jurcicek, F. Sequence-to-sequence generation for spoken dialogue via deep syntax trees and strings. In *Proceedings of ACL 2016*, 45–51 (Berlin, Germany, 2016).
21. Krahmer, E., Goudbeek, M. & Gatt, A. (eds.). *INLG 2018: 10th International Natural Language Generation Conference* (2018).
22. Reiter, E., Sripada, S., Hunter, J., Yu, J. & Davy, I. Choosing words in computer-generated weather forecasts. *Artif. Intell.* **167**, 137–169 (2005).
23. Novikova, J., Lemon, O. & Rieser, V. Crowd-sourcing nlg data: Pictures elicit better data. In *Proceedings of the 9th International Natural Language Generation conference*, 265–273 (2016).
24. Bernardi, R. *et al.* Automatic description generation from images: A survey of models, datasets, and evaluation measures. *J. Artif. Intell. Res.* **55**, 409–442 (2016).
25. Fong, T., Nourbakhsh, I. & Dautenhahn, K. A survey of socially interactive robots. *Robotics Auton. Syst.* **42**, 143–166 (2003). DOI 10.1016/S0921-8890(02)00372-X.

26. Kiesler, S. (ed.) *Journal of Human-Robot Interaction*.
27. Ge, S. S. & Khatib, O. (eds.) *International Journal of Social Robotics* (Springer).
28. Kanda, T. & Sabanovic, S. (eds.). *HRI 2018: 13th IEEE/HRI Conference on Human-Robot Interaction*.
29. Chellai, R. & Chen, W. (eds.). *IEEE RO-MAN 2018: The 27th International Symposium on Robot and Human Interactive Communication*.
30. Ge, S. S. & Cabibhan, J.-J. (eds.). *ICSR 2018: Tenth International Conference on Social Robotics*.
31. Foster, M. E. *et al.* Two people walk into a bar: Dynamic multi-party social interaction with a robot agent. In *Proceedings of ICMI 2012* (Santa Monica, 2012). DOI 10.1145/2388676.2388680.
32. Hoffman, G., Zuckerman, O., Hirschberger, G., Luria, M. & Shani Sherman, T. Design and evaluation of a peripheral robotic conversation companion. In *Proceedings of HRI 2015*, 3–10 (2015). DOI 10.1145/2696454.2696495.
33. Nikolaidis, S., Ramakrishnan, R., Gu, K. & Shah, J. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In *Proceedings of HRI 2015*, 189–196 (2015). DOI 10.1145/2696454.2696455.
34. Bevan, C. & Stanton Fraser, D. Shaking hands and cooperation in tele-present human-robot negotiation. In *Proceedings of HRI 2015*, 247–254 (2015). DOI 10.1145/2696454.2696490.
35. Brscić, D., Kidokoro, H., Suehiro, Y. & Kanda, T. Escaping from children's abuse of social robots. In *Proceedings of HRI 2015*, 59–66 (2015). DOI 10.1145/2696454.2696468.
36. Bartneck, C., Kulić, D., Croft, E. & Zoghbi, S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robotics* **1**, 71–81 (2009). DOI 10.1007/s12369-008-0001-3.
37. Weiss, A. & Bartneck, C. Meta analysis of the usage of the Godspeed questionnaire series. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2015)* (Kobe, Japan, 2015).
38. Carpinella, C. M., Wyman, A. B., Perez, M. A. & Stroessner, S. J. The robotic social attributes scale (RoSAS): Development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, 254–262 (ACM, New York, NY, USA, 2017). DOI 10.1145/2909824.3020208.
39. Young, J. E. *et al.* Evaluating human-robot interaction. *Int. J. Soc. Robotics* **3**, 53–67 (2011).
40. Kruse, T., Pandey, A. K., Alami, R. & Kirsch, A. Human-aware robot navigation: A survey. *Robotics Auton. Syst.* **61**, 1726–1743 (2013).
41. Vinciarelli, A. & Esposito, A. Multimodal analysis of social signals. *The Handb. Multimodal-Multisensor Interfaces, Vol. 2: Signal Process. Archit. Detect. Emot. Cogn.* 203 (2018).
42. Walker, M. *et al.* (eds.). *NLG4DS: SIGDIAL 2017 Special Session on Natural Language Generation for Dialogue Systems*. https://sites.google.com/view/nlg4ds2017.
43. Foster, M. E., Giuliani, M. & Isard, A. Task-based evaluation of context-sensitive referring expressions in human-robot dialogue. *Lang. Cogn. Neurosci.* **29**, 1018–1034 (2014). DOI 10.1080/01690965.2013.855802.
44. Wen, T.-H. *et al.* Stochastic language generation in dialogue using recurrent neural networks with convolutional sentence reranking. In *Proceedings of SIGdial 2015*, 275–284 (Prague, Czech Republic, 2015).
45. Hastie, H., Cuayáhuitl, H., Dethlefs, N., Keizer, S. & Liu, X. *Extrinsic Versus Intrinsic Evaluation of Natural Language Generation for Spoken Dialogue Systems and Social Robotics*, 303–311 (Springer Singapore, 2017). DOI 10.1007/978-981-10-2585-3_24.
46. Striegnitz, K. *et al.* Report on the second second challenge on generating instructions in virtual environments (GIVE-2.5). In *Proceedings of the 13th European Workshop on Natural Language Generation*, 270–279 (2011).
47. Li, S., Scalise, R., Admoni, H., Rosenthal, S. & Srinivasa, S. S. Spatial references and perspective in natural language instructions for collaborative manipulation. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 44–51 (2016). DOI 10.1109/ROMAN.2016.7745089.
48. Papaioannou, I., Dondrup, C. & Lemon, O. Human-robot interaction requires more than slot filling – multi-threaded dialogue for collaborative task and social conversation. In *Proceedings of AI-MHRI 2018* (2018).
49. Papaioannou, I. & Lemon, O. Combining chat and task-based multimodal dialogue for more engaging HRI: A scalable method using reinforcement learning. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 365–366 (Vienna, Austria, 2017). DOI 10.1145/3029798.3034820.
50. Gatt, A. & Reiter, E. SimpleNLG: A realisation engine for practical applications. In *Proceedings of ENLG 2009*, 90–93 (Athens, Greece, 2009).
51. White, M. Efficient realization of coordinate structures in combinatory categorial grammar. *Res. on Lang. Comput.* **4**, 39–75 (2006). DOI 10.1007/s11168-006-9010-2.
52. Novikova, J., Dušek, O. & Rieser, V. The E2E NLG shared task (2017). http://www.macs.hw.ac.uk/InteractionLab/E2E/.
53. Foster, M. E., Buschmeier, H. & Gkatzia, D. (eds.). *INLG 2018 Workshop on NLG for HRI* (2018).