Natural Language Generation for Social Robotics Applications*

Mary Ellen Foster
School of Computing Science, University of Glasgow
MaryEllen.Foster@glasgow.ac.uk
http://maryellenfoster.uk/

Abstract

A crucial aspect of face-to-face social conversation is fluent, flexible linguistic interaction. However, most developers of social robots 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.

Introduction

Developing and deploying a socially interactive robot presents a number of significant technical challenges, requiring audiovisual processing, social signal processing, action selection, and robot navigation and motion planning. 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 simple, 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. This paper begins surveys 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 NLG.

Natural Language Generation

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 (Gatt and Krahmer 2018). 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, while in other cases, the input to the NLG system is non-linguistic; 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.

*This is an abridged version of (Foster, 2019).

The techniques that may be used to create the linguistic output also vary across applications. Simple systems may be developed using **templates**—that is, by slotting the content into pre-built linguistic structures. A 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. 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 (Narayan and Gardent 2016; Dušek and Jurcicek 2016): the system then directly learns the mapping between inputs and outputs without any need for explicit intermediate representations such as templates, rules, or grammars.

Social Robotics

Humans have a strong tendency to anthropomorphise robots and to want to engage in social interaction with them. 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; it must sense and respond to the non-verbal social signals of its human partners; 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.

NLG and Social Robotics

In their recent survey of the state-of-the-art in NLG research, (Gatt and Krahmer 2018) 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." Previous research (Foster, Giuliani, and Isard 2014;

Hastie et al. 2017) has found that an interactive 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. However, most social robot interactions involve much more than referring to world objects and filling slots for database queries (Papaioannou, Dondrup, and Lemon 2018): 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. Supporting this range of behaviours requires generating a wide range of output types.

It must be noted that templates are not always an invalid implementation decision for language generation (van Deemter, Theune, and Krahmer 2005)—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 (Gatt and Reiter 2009) or OpenCCG (White 2006) could be used to provide advantages such as flexibility and cross-lingual support.

A recent workshop at the INLG 2018 conference (Foster, Buschmeier, and Gkatzia 2018) 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 (Striegnitz et al. 2011), but will incorporate aspects of situated human-robot interaction.

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.

References

[Dušek and Jurcicek 2016] Dušek, O., and Jurcicek, F. 2016. Sequence-to-sequence generation for spoken dialogue via deep syntax trees and strings. In *Proceedings of ACL 2016*, 45–51.

[Foster, Buschmeier, and Gkatzia 2018] Foster, M. E.; Buschmeier, H.; and Gkatzia, D., eds. 2018. *INLG 2018 Workshop on NLG for HRI*.

[Foster, Giuliani, and Isard 2014] Foster, M. E.; Giuliani, M.; and Isard, A. 2014. Task-based evaluation of context-sensitive referring expressions in human-robot dialogue. *Language, Cognition, and Neuroscience* 29(8):1018–1034.

[Foster 2019] Foster, M. E. 2019. Natural language generation for social robotics: Opportunities and challenges. *Philosophical Transactions of the Royal Society B: Biological Sciences*.

[Gatt and Krahmer 2018] Gatt, A., and Krahmer, E. 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research* 61:65–170.

[Gatt and Reiter 2009] Gatt, A., and Reiter, E. 2009. SimpleNLG: A realisation engine for practical applications. In *Proceedings of ENLG 2009*, 90–93.

[Hastie et al. 2017] Hastie, H.; Cuayáhuitl, H.; Dethlefs, N.; Keizer, S.; and Liu, X. 2017. *Extrinsic Versus Intrinsic Evaluation of Natural Language Generation for Spoken Dialogue Systems and Social Robotics*. Springer Singapore. 303–311.

[Narayan and Gardent 2016] Narayan, S., and Gardent, C. 2016. Unsupervised sentence simplification using deep semantics. In *Proceedings of INLG 2016*, 111–120.

[Papaioannou, Dondrup, and Lemon 2018] Papaioannou, I.; Dondrup, C.; and Lemon, O. 2018. Human-robot interaction requires more than slot filling multi-threaded dialogue for collaborative task and social conversation. In *Proceedings of AI-MHRI 2018*.

[Striegnitz et al. 2011] Striegnitz, K.; Denis, A.; Gargett, A.; Garoufi, K.; Koller, A.; and Theune, M. 2011. 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.

[van Deemter, Theune, and Krahmer 2005] van Deemter, K.; Theune, M.; and Krahmer, E. 2005. Real versus template-based natural language generation: A false opposition? *Computational Linguistics* 31(1):15–24.

[White 2006] White, M. 2006. Efficient realization of coordinate structures in combinatory categorial grammar. *Research on Language and Computation* 4(1):39–75.