Natural Language Generation for Public-Space Social Robots

Mary Ellen Foster
School of Computing Science, University of Glasgow
Glasgow, UK
MaryEllen.Foster@glasgow.ac.uk

Figure 1: C-3PO. Image © 1977 Lucasfilm Ltd.

1 SOCIAL ROBOTICS

In popular culture, the prototypical “robot” is an artificial human that is able to engage fully in all aspects of face-to-face conversation (e.g., fig. 1). As modern robot hardware becomes safer, more sophisticated, and more generally available, there is a clear and significant user demand for socially intelligent robots [2] that are able to interact flexibly in everyday human environments. In particular, social robots are now present in a wide range of public spaces including shopping malls [3], airports [9], offices [6], and museums [8].

Developing and deploying a socially interactive robot presents a number of significant technical challenges. A successful social robot must incorporate state-of-the-art components for audiovisual processing, social signal processing, action selection, and robot 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, natural language generation (NLG) 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. 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 [1]. We believe that the time is right for developers of social robots to explore the output possibilities provided by incorporating state-of-the-art NLG.

2 NLG FOR SOCIAL ROBOTICS

In their recent survey of the state-of-the-art in NLG research, Gatt and Krahmer [4] 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.” 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.

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). However, most social robot interactions involve much more than referring to world objects. 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 behaviors requires generating a wide range of output types. In the current MuMMER prototype, a hybrid chatbot/task-based system is used for interaction management [7]—however, even in that context, the actual textual output is ultimately produced by templates.

It must be noted that templates are not always an invalid choice for language generation [10]—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 model-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 [5] or OpenCCG [11] 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.

### 3 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. Public-space social robotics 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


