Abstract
This work presents a new approach for modelling “unknown unknowns” in Natural Language Generation (NLG) systems. We first address the question of how to adapt to unknown users, using a combination of cluster-based user modelling and Pareto optimal Multi-Objective Optimisation. Next, we present a corpus study on generating referring expressions (RE) for instruction giving in the real-world, where the visual scene complexity is (currently) unknown. We find significant differences between REs generated in the virtual and the real world and draw general conclusions for NLG systems.

1 Introduction
This line of research investigates Natural Language Generation (NLG) under uncertainty. There are two general types of uncertainty: Aleatoric and epistemic uncertainty. The former addresses variation observed in the real world, and is currently accounted for by, e.g. modelling probability distributions over likely user responses (Rieser et al., 2014). The latter are “things we cannot know”, i.e. true unknowns which cannot be estimated from data. These type of uncertainties are currently not taken into account when generating system outputs. In this paper, we summarise our work on modelling “unknown unknowns” for NLG systems. In particular, we address two problems:

1. How to adapt output to unknown user types?
2. How to generate referring expressions for objects in unknown visual scenes?

2 Talking to Unknown Users
Handling first time users is a common problem for NLG and interactive systems in general: the system cannot adapt to user preferences without prior knowledge. Previous work has shown that it is important for NLG systems to adapt their output to specific users or user groups, such as nurses and patients (Gatt et al., 2009), or lecturers and students (Gkatzia et al., 2014a). However, adaptation becomes impossible when no prior information about this user exists, as is often the case for first time users. Furthermore, user preferences often vary significantly ($p < 0.001$) for the same utterance (Walker et al., 2007), which will naturally affect the performance of models based on population average. We propose a novel model for first time users, which is based on clusters of potential user types.

We apply this framework to medical first aid decision support systems, where we automatically generate short reports of sensor data data, including Breathing Rate ($BR$), the Blood Oxygen Saturation ($SpO_2$) and the Heart Rate ($HR$). In a medical emergency a patient’s survival often depends upon the prompt response and appropriate first aid given by the first person on scene, also known as “bystander”, who typically is a first-time user.

NLG user modelling approaches assume that (1) the system is only used by a single user, and (2) the type of user is known in advance. In order to remedy the first assumption, Gkatzia et al. (2014a) suggest a multi-objective approach to NLG. In this framework, the preferences of lecturers and students are modelled as objective functions that need to be optimized simultaneously. A Reinforcement Learning agent is then trained by using the weighted sum of the modelled preferences as a reward function. However, their experiments show that this approach is unsuccessful – possibly because their linearised reward function cancels out the preferences of each user group. We propose a different MOO approach and we actually demonstrate a Pareto optimal approach to Multi-Objective Optimisation (MOO) for NLG, which
can actually find the middle ground between conflicting objectives.

In order to address the second assumption, earlier work used questionnaires to derive information about the user in order to personalise the output to each individual (Reiter et al., 1999). In this work, we follow a recent cluster-based approach, proposed by (Dethlefs et al., 2014). This approach does not rely on pre-defined user types, but discovers groups with similar preferences based on a limited number of initial ratings. This approach suits our needs since previous work on first aid provision showed that there are no clear pre-defined user groups (Gkatzia et al., 2014b). However, Dethlefs et al.’s (2014) system is only able to adapt to users after some initial interaction (reaching its best performance after 9 user ratings). Yet, in the domain of first aid provision, we deal with first-time users with no prior ratings. As such, we extend Dethlefs et al.’s (2014) approach to deal with truly unknown users, by first clustering user groups based on existing data. We evaluate this model with previously unknown users, i.e. different from the population in the initial data set. The results show that joint optimisation significantly improves over models optimised for one user group only, and hence show the effectiveness of our approach.

3 Instruction Giving Under Uncertainty in Situated Environments

Generating effective referring expressions (RE) is vital to the success of instruction giving assistants in situated real-world settings. Traditionally, research has focused on studying Referring Expression Generation (REG) in virtual, controlled environments, e.g. (Byron et al., 2009; Janarthanam and Lemon, 2011). These virtual environments are constructed so that there is a limited, well defined set of distractor objects, leading to instructions with referring expressions such as “Press the second blue button from the left”. When generating instructions for virtual personal assistants (VPAs), such as Apple’s Siri or Google Now, however, we need to consider real-world settings, rather than a controlled environment. In particular, real-world scenes are likely to be more complex and contain a high degree of uncertainty. For example, we cannot readily assume to know the distractor set due to the (currently) limited object recognition capabilities of VPAs.

In this work, we present first steps towards formulating a framework for REG under uncertainty. In particular, we collect a novel, real-world corpus REAL – “Referring Expression Anchored Language”, and compare our findings to those reported in virtual worlds (Gargett et al., 2010). We then provide a detailed analysis of how syntactic and semantic features contribute to the success of REG, accounting for unobservable latent variables, such as the complexity of the visual scene. To our knowledge, this is the first corpus study to evaluate referring expressions in the real world.

We use these findings to draw conclusions for NLG systems for real-world instruction giving systems. First, we find that semantic features have a bigger impact on the success rate of REs than syntactic features. This implies that content selection is more important than surface realisation for REG.

Second, we find that semantic features such as taxonomic and absolute properties significantly contribute to RE success. Taxonomic properties refer to the type of target object, and in general depend on the local knowledge of the information giver, e.g. the Scottish parliament vs. the new modern building. Similarly, the success of the RE will depend on the expertise of the information follower, see e.g. (Janarthanam and Lemon, 2014). As such, modelling the user’s level of expertise/knowledge is crucial. Absolute properties refer to object attributes, such as colour. Attribute selection for REG has attracted a considerable amount of attention over the past 20 years (Krahmer and van Deemter, 2011). As such, it would be interesting to investigate how these automatic attribute selection algorithms perform in real-world, interactive environments.

Third, we find that more complex scenes seem to justify longer and more complex descriptions. However, user attention is limited in interactive systems. As such, there is an underlying trade-off which needs to be optimised, e.g. following the generation framework described in (Rieser et al., 2014).

Acknowledgments

This research received funding from the EPSRC GUI project – “Generation for Uncertain Information” (EP/L026775/1).
References


Xiwu Han, Somayajulu Sripada, Kit (CJA) Macleod, and Antonio A. R. Ioris. 2014. Latent user models for online river information tailoring. In 8th International Natural Language Generation Conference (INLG).


