The cognitive plausibility of algorithms for referring expression generation

Jette Viethen and Robert Dale

Department of Computing
Macquarie University
Sydney, Australia
[jviethen,rdale]@science.mq.edu.au

Over the last two decades, referring expression generation has become one of the most popular research topics within Natural Language Generation. Arguably, describing objects is a central task for any computational system engaged in producing natural language. There are two distinct exercises we might be undertaking when we develop algorithms to determine the content for referring expression. One would be an exercise in engineering, where we attempt to develop algorithms that are able to effectively identify intended referents for hearers, for use in sophisticated applications of natural language generation. Alternatively, we might be engaged in an exercise in computational psycholinguistics, where we attempt to model what it is that speakers do when they refer to an entity. Our interest in the present paper is in the second of these exercises: how might we develop algorithms that emulate the referring behaviour of humans, and in so doing, perhaps begin to explain how humans carry out this task?

We present an analysis of a corpus of experimentally collected referring expressions in which humans describe simple target objects surrounded by a few distractor objects with the aim of shedding some light on the processes that might be at play. We show that popular algorithms for the generation of referring expressions are not able to produce the full range of variation displayed in the human-produced data and are therefore unlikely candidates for cognitive models. This shortcoming is in large parts due to the serial fashion in which these algorithms consider properties for inclusion in a referring expression, where each decision is dependent on what has been included so far. One important finding from our analysis is that the data vastly underspecifies what might be involved algorithmically in the generation of referring expressions in physical scenes such as the ones from our corpus. However, it does allow a number of conclusions to be drawn that lead us to propose to replace serial dependency with a parallel gestalt model that considers visually salient properties independently of each other. We conclude with a list of questions that need to be addressed experimentally before a full algorithmic account of human referring expression generation is possible.