Preface

Referring Expression Generation Challenges 2008 (REG'08) was the second shared-task evaluation challenge (STEC) in the field of Natural Language Generation (NLG) and took place between September 2007, when REG'08 was first announced, and May 2008, when the evaluation of the participating systems was completed.

REG'08 follows on from the Attribute Selection for Referring Expression Generation Challenge in 2007 (ASGRE'07) which was conceived as a pilot event for shared-task evaluation in NLG. Shared tasks in NLG are themselves a natural continuation of a growing interest in more comparative forms of evaluation among NLG researchers, and mirror trends in other HLT fields.

Since the foundational work of authors such as Appelt, Kronfeld, Grosz, Joshi, Dale and Reiter, Referring Expression Generation (REG) has been the subject of intensive research in NLG, and has — unusually for NLG — led to significant consensus on the REG problem definition, as well as the nature of the inputs and outputs of REG algorithms. This is particularly true of the attribute selection subtask, perhaps the most widely researched NLG subtask, which was the shared task in ASGRE'07. REG'08 included the same task (TUNA-AS), and added two more tasks based on the TUNA corpus: realisation (TUNA-R) and the complete referring expression generation task (TUNA-REG). REG'08 also introduced a new data set of short introductory sections from Wikipedia articles on geographic entities and people (the GREC corpus) and a new task based on it: generation of referring expressions for named entities in the context of a discourse longer than a sentence. The intended application context for this task is improvement of referential clarity in extractive summaries.

In addition to the four shared tasks, REG'08 offered, for each of the two datasets, (i) an open submission track in which participants could submit any work involving the data while opting out of the competetive element, and (ii) an evaluation track, in which proposals for new evaluation methods for the shared task could be submitted. We believe that these two types of open-access tracks are important because they allow the wider research community to shape the focus and methodologies of STECs directly.

We successfully applied (with the help of support letters from many of last year's participants and other HLT colleagues) for funding from the Engineering and Physical Sciences Research Council (EPSRC), the main funding body for HLT in the UK. This support enabled us to double the size of the GREC corpus and to carry out extensive human task performance evaluations, as well as employ a dedicated research fellow (Eric Kow) to help with all aspects of REG'08. It also enabled us to enlist the help of Jette Viethen from Macquarie University with the GREC evaluations.

REG'08 got underway with a first announcement in September 2007. We released samples for both datasets and invited preliminary registrations in January 2008, and released the full Participants' Pack including instructions and training/development data on 22nd February to registered participants. Twelve teams registered for one or more of the TUNA tasks, and five for the GREC Task. Among the participants were teams from Australia, Spain, Ireland, UK, USA, Brazil, Belgium, Netherlands, India and Germany.

By the deadline of 7th April (which was extended slightly), 8 teams submitted 13 systems for TUNA-AS, 4 teams submitted 5 systems for TUNA-R, and 6 teams submitted 15 systems for TUNA-REG. Three teams submitted 6 systems for the brand new GREC task, to which we added four baseline systems. We also received one submission in the TUNA Open Track which applies Portuguese surface realisation to TUNA attribute sets.

The submission process required participants to first submit a report describing their approach and reporting results on the development data for the task(s) they had participated in. For this purpose, they were supplied with programs to compute the relevant evaluation metrics. On submission of the report, they could download the test data, and had 48 hours to submit their outputs on the test set.

All submissions are described in the participants' reports in this volume. We are pleased to say that several of the contributions came from students, as well as from researchers entirely new to the field of NLG.

We had a total of 33 TUNA systems and 10 GREC systems to evaluate. We conducted task-performance experiments for all TUNA-REG systems and the 6 submitted GREC systems. We computed a range of automatic intrinsic measures for all task tracks. For the GREC task, we also tried out a new evaluation idea: automatic *ex*trinsic evaluation using coreference resolution tools. All these evaluation methods are described in detail in the two evaluation reports in this volume.

Preparations are already underway for a third NLG shared-task evaluation event in 2009, Generation Challenges 2009, which will include a TUNA-REG progress check task, a GREC task and a new task on instruction giving in virtual environments (the GIVE Challenge). Results will be presented at ENLG'09.

Like all STECs, REG'08 would not have been possible without the contributions of many different people. We would like to thank the faculty, staff and students of Brighton University and the friends who participated in the evaluation experiments; the INLG'08 organisers, in particular Mike White; the research support team at Brighton University and the EPSRC for help with obtaining funding; Jason Baldridge and Pascal Denis for assistance with the coreference resolvers; and last but not least, the REG'08 participants for making the most of the short available time to build their systems (and for never once complaining, even when they had good reason to). Special thanks are due to Eric Kow and Jette Viethen who worked extremely hard especially during the four weeks of the evaluation period.

Anja Belz and Albert Gatt