Generation Challenges 2009

Preface

Generation Challenges 2009 was the third round of shared-task evaluation competitions (STECs) that involve the generation of natural language, and followed the Pilot Attribute Selection for Generating Referring Expressions Challenge in 2007 (ASGRE'07) and Referring Expression Generation Challenges in 2008 (REG'08). More information about all these NLG STEC activities can be found via the links on the Generation Challenges homepage: http://www.nltg.brighton.ac.uk/research/genchal09

Generation Challenges 2009 brought together four STECs: the TUNA Referring Expression Generation Task (TUNA-REG) organised by Albert Gatt, Anja Belz and Eric Kow; the two GREC Challenges, GREC Main Subject Reference Generation (GREC-MSR) and GREC Named Entity Generation (GREC-NEG), organised by Anja Belz, Eric Kow, Jette Viethen and Albert Gatt; and the Giving Instructions in Virtual Environments Challenge (GIVE) organised by Donna Byron, Justine Cassell, Robert Dale, Alexander Koller, Johanna Moore, Jon Oberlander, and Kristina Striegnitz.

In the GIVE Challenge, participating teams developed systems which generate natural-language instructions to users navigating a virtual 3D environment and performing computer-game-like tasks. The four participating systems were evaluated by measuring how quickly, accurately and efficiently users were able to perform tasks with a given system's instructions. The evaluation report for the GIVE Challenge can be found in this volume; the participants' reports will be made publicly available at a later stage.

The TUNA-REG Task was the end-to-end referring expression generation task (combining the attribute selection and realisation subtasks) which was first introduced in REG'08, and which used the TUNA corpus of paired descriptions and pictures of entities. This year's TUNA-REG Task had an open call for participation, but it was also organised in the spirit of a progress check which would give participants from TUNA-REG'08 an opportunity to submit improved systems, the results for which could be compared to last year's results. Of five registered teams from five countries, four teams submitted a total of 6 systems to TUNA-REG. These, along with two sets of human outputs, were evaluated by automatic intrinsic and human-based intrinsic and extrinsic evaluations. The results report and the participants' reports can be found in this volume.

The GREC-MSR Task was the same as in REG'08 and used a corpus of introductory sections from Wikipedia articles on geographic entities and people. The task was to generate referring expressions for mentions of the main subject of the article in the context of the full text of the article. The new GREC-NEG Task used a separate corpus of introductory sections from Wikipedia articles on people, and the task was to generate referring expressions for all mentions of all people in an article.

Eight teams from seven countries registered for each of the GREC-MSR and GREC-NEG tasks. As the system submission deadline approached, it became clear that just two teams were certain that they were going to complete their systems in time. For this reason, and also because of a moving camera-ready deadline, we decided, after careful consideration and consultation with participants, to extend the system development period for the GREC Tasks and to hold the GREC'09 results

meeting at the ACL-IJCNLP'09 Workshop on Language Generation and Summarisation in Singapore on 6 August 2009, and to publish all GREC'09 reports in the proceedings of that workshop.

In addition to the four shared tasks, Generation Challenges 2009 offered (i) an open submission track in which participants could submit any work involving the data from any of the shared tasks, while opting out of the competetive element, (ii) an evaluation track, in which proposals for new evaluation methods for the shared task could be submitted, and (iii) a task proposal track in which proposals for new shared tasks could be submitted. We believe that these types of open-access tracks are important because they allow the wider research community to shape the focus and methodologies of STECs directly. We received one submission in the open submission track, involving the TUNA data, and none in the other tracks.

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 helped with all aspects of organising Generation Challenges 2009, and enabled us to create the new GREC-People corpus and to carry out extensive human evaluations, as well as to employ a dedicated research fellow (Eric Kow) to help with all aspects of Generation Challenges 2009.

Preparations are already underway for a fourth NLG shared-task evaluation event next year, Generation Challenges 2010, which is likely to include a further run of the GREC-NEG Task with an extended training/development corpus, a new task which links GREC-NEG to a named-entity recognition preprocessing stage, and a second run of the GIVE Challenge. We are hoping that results will be presented at INLG'10.

Like our previous STECs, Generation Challenges 2009 would not have been possible without the contributions of many different people. We would like to thank the faculty and staff of Brighton University, and the students of UCL, Brighton and Sussex Universities who participated in the evaluation experiments as well as all other participants in our online data elicitation and evaluation exercises; the ENLG'09 organisers, Mariet Theune and Emiel Krahmer; the research support team at Brighton University and the EPSRC for help with obtaining funding; and last but not least, the participants in the shared tasks for making the most of the short available time to build some very successful systems.

February 2009   *Anja Belz and Albert Gatt*