

Generation Challenges 2011 Preface

Generation Challenges 2011 (GenChal'11) was the fifth round of shared-task evaluation competitions (STECs) involving the generation of natural language. It followed four previous events: the Pilot Attribute Selection for Generating Referring Expressions (ASGRE) Challenge in 2007 which had its results meeting at UCNLG+MT in Copenhagen, Denmark; Referring Expression Generation (REG) Challenges in 2008, with a results meeting at INLG'08 in Ohio, US; Generation Challenges 2009 with a results meeting at ENLG'09 in Athens, Greece; and most recently Generation Challenges 2010 with a results meeting at INLG'10 in Trim, Ireland. More information about all these NLG STEC events can be found via the links on the Generation Challenges homepage (<http://www.nltg.brighton.ac.uk/research/genchal11>).

GenChal'11 brought together three STECs: the first Surface Realisation Challenge (SR'11) organised by Anja Belz, Deirdre Hogan, Michael White and Amanda Stent; the Challenge on Generating Instructions in Virtual Environments (GIVE) organised by Kristina Striegnitz, Alexandre Denis, Andrew Gargett, Konstantina Garoufi, Alexander Koller, and Mariët Theune; and the new Helping Our Own Challenge (HOO) organised by Robert Dale and Adam Kilgarriff.

In addition, GenChal'11 had a Future Task Proposals Track where researchers were invited to submit papers describing ideas for STECs to be run in the future. The proposals that were submitted to this track are the first two papers in this part of the proceedings: Janarthanam and Lemon's paper on the proposed GRUVE Challenge which can be seen as taking up where the GIVE Challenge is now leaving off; and Gervas and Ballesteros's paper on a Spanish version of the Surface Realisation Challenge.

For the first time this year, GenChal did not have an Open Track or Evaluation Methodologies Track, as these attracted very few submissions in the past.

The SR Task was based on Penn Treebank data and the organisers created two different input representations, one shallow, one deep, mainly from the annotations used in the CoNLL'08 Shared Task. The task for participating teams was to automatically generate surface realisations from the input representations. Five teams submitted six systems to the shallow and deep tracks. The submitted systems were evaluated using four automatic metrics and three human-assessed criteria. This volume includes the SR Task results report and the system reports by the participating teams.

In the GIVE Challenge, participating teams developed systems which generate natural-language instructions that help a human user solve a task in a 3D virtual world. The eight participating systems were evaluated by measuring how accurately and efficiently users were able to perform the task with a given system's instructions, and by collecting subjective ratings of the instruction quality from users. This year's GIVE Challenge maintained the same task as in GIVE-2 (with new evaluation worlds, of course), so that the participating teams could learn from the results of last year's edition and additional teams would be able to participate. The evaluation report for the GIVE Challenge as well as descriptions of the participating systems can be found in this volume. The software infrastructure (and at a later stage the collected data) is available on the GIVE website (<http://www.give-challenge.org/research>).

The first HOO Challenge used a corpus of 1,000-word excerpts of text from papers in the ACL anthology that have been donated by their authors. Each excerpt was copy-edited by professional copy-editors and marked up with the resulting corrections. The task for participants was to produce such corrections automatically. Despite a relatively short turn-around time, six teams were able to participate in HOO. Their system reports and the results report by Dale and Kilgarriff are included in this volume.

The Question Generation Challenge did not run this year. However, the organisers have contributed a report outlining recent and future developments.

Once again, 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 developing and running the SR Task and organising Generation Challenges 2011. It enabled us to create the SR Task data and to carry out human evaluations, as well as to pay for Deirdre Hogan and Eric Kow's time spent working on the SR Task.

Preparations are already underway for a sixth NLG shared-task evaluation event next year, Generation Challenges 2012, which is likely to include a first run of the GRUVE Challenge, a second run of the SR Task, hopefully as a multilingual task, including the Spanish version, and a second run of the HOO task. Results are likely to be presented at INLG'12.

Just like our previous STECs, Generation Challenges 2011 would not have been possible without the contributions of many different people. We would like to thank the students of Oxford University, KCL, UCL and Sussex Universities who participated in the SR Task evaluations; the ENLG'11 organisers, Claire Gardent and Kristina Striegnitz; 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 themselves.

*Anja Belz, Albert Gatt, Alexander Koller and Kristina Striegnitz
September 2011*

Generation Challenges Steering Committee:

Anja Belz, University of Brighton, UK
Robert Dale, Macquarie University, Australia
Albert Gatt, University of Malta and University of Aberdeen, UK
Kevin Knight, ISI, University of Southern California, USA
Alexander Koller, Saarland University, Germany
Chris Mellish, Aberdeen University, UK
Johanna Moore, Edinburgh University, UK
Amanda Stent, Stony Brook University, USA
Kristina Striegnitz, Union College, USA