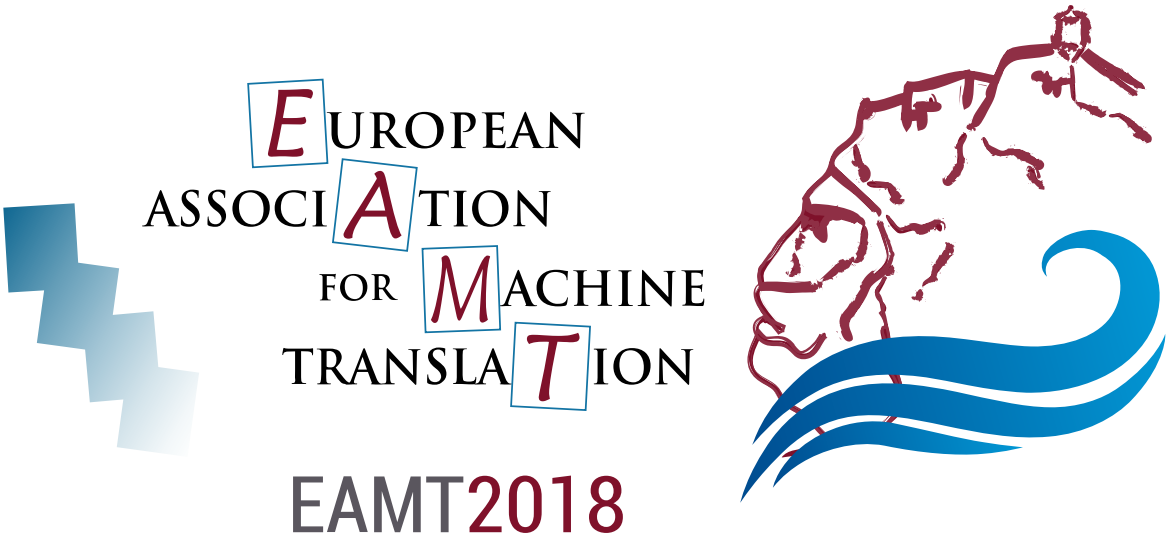

Proceedings of the

21st Annual Conference of the European Association for Machine Translation

28–30 May 2018

Universitat d'Alacant

Alacant, Spain

Edited by

Juan Antonio Pérez-Ortiz

Felipe Sánchez-Martínez

Miquel Esplà-Gomis

Maja Popović

Celia Rico

André Martins

Joachim Van den Bogaert

Mikel L. Forcada

Organised by

Universitat d'Alacant
Universidad de Alicante

transducens
research group

The papers published in this proceedings are —unless indicated otherwise— covered by the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 International (CC-BY-ND 3.0). You may copy, distribute, and transmit the work, provided that you attribute it (authorship, proceedings, publisher) in the manner specified by the author(s) or licensor(s), and that you do not use it for commercial purposes. The full text of the licence may be found at https://creativecommons.org/licenses/by-nc-nd/3.0/deed.en.

© 2018 The authors
ISBN: 978-84-09-01901-4

The WMT Shared Tasks

Barry Haddow
School of Informatics
University of Edinburgh
Edinburgh
Scotland
bhaddow@staffmail.ed.ac.uk

Abstract

The annual WMT Conference in Machine Translation has been running shared tasks since 2006. It started with a translation task based on Europarl, and has grown to include tasks on all aspects of MT corpus preparation, training and evaluation, including the flagship task on news translation. I will review the history of the task, lessons learnt, and plans for future tasks.

1 Introduction

We began organising shared tasks in machine translation at the Workshop in Machine Translation (WMT) in 2006, initially with a translation task based on Europarl. In later years, funding from the EU projects EuroMatrixPlus and MosesCore (FP7) and QT21 and Cracker (H2020), plus commercial sponsorship, enabled us to increase the number of tasks and to produce professionally translated, unseen test sets drawn from news texts for the translation task. In 2016 WMT became a conference (retaining the acronym) and in the last three years the number of shared tasks has varied between 7 and 10.

The shared tasks have covered translation (mainly news, but also other domains such as IT and biomedical and also more specialized tasks such as pronoun and multimodal), training (both tuning of SMT and training of NMT), reference-based evaluation, quality estimation, corpus preparation (document alignment and corpus cleaning) as well as automatic post-editing. The quality estimation task has included different subtasks on estimating the quality of MT output at word, sentence and document level, as well as trying to predict the post-editing effort required for a given MT output. The data from all the WMT tasks, including the training data, test data and task submissions is made available for future research and has been heavily used in academic publications.

In the news translation task we have tried to innovate in MT evaluation, whilst still providing for comparison with previous years. After several years using a *relative ranking* approach, where evaluators compare output from different systems, we switched to *direct assessment* (DA) in 2017. In DA, evaluators provide an assessment of adequacy on a scale from 0 to 100, which we find offers a reliable system ranking and a more interpretable and comparable final score. The news task covers a variety of languages, mainly European, with English–German and English–Czech as our "core" languages. We have included both low-resource (e.g. Estonian–English and Hindi–English) and high-resource (e.g. French–English) pairs, and we release our own parallel and monolingual data sets, as well as using standard sets like Europarl.

In this talk I will review the history of the tasks, the lessons learnt and plans for future tasks, focusing on the news translation task. I will explain how this task provides a common benchmark for comparing different MT systems, which helps to drive MT research. I will also show how running the task reveals difficulties and pitfalls in comparative evaluation of MT systems.

2 Website

The URL for the latest conference/task is `www.statmt.org/wmt18`, where you will find links to all previous conferences/workshops, tasks and papers.

© 2018 The authors. This article is licensed under a Creative Commons 3.0 licence, no derivative works, attribution, CC-BY-ND.

Pérez-Ortiz, Sánchez-Martínez, Esplà-Gomis, Popović, Rico, Martins, Van den Bogaert, Forcada (eds.)
Proceedings of the 21st Annual Conference of the European Association for Machine Translation, p. 9
Alacant, Spain, May 2018.