Proceedings of the
Second Joint EM+/CNGL Workshop
"Bringing MT to the User: Research on Integrating MT in the Translation Industry"

JEC 2010

Held in conjunction with
The Ninth Conference of the Association for Machine Translation in the Americas

Edited by
Ventsislav Zhechev

November 4th, 2010
Denver

Preface to the Proceedings of the Second Joint EM+/CNGL Workshop

Welcome to the Second Joint EM+/CNGL Workshop "Bringing MT to the User: Research on Integrating MT in the Translation Industry". This AMTA 2010 Workshop is organised jointly by the EuroMatrix+ Project (`http://www.euromatrixplus.eu`) and the Centre for Next Generation Localisation (`http://cngl.ie`) and took place in Denver, CO on 4 November 2010.

Premise: Recent years have seen a revolution in MT triggered by the emergence of statistical approaches to MT and improvements in translation quality. MT (rule-based, statistical and hybrid) is now available for many languages for free on the Web and is making strong inroads into the corporate localisation and translation industries. Open-source MT solutions are competing with proprietary products. Increasing numbers of translators are post-editing TM/MT output. At the same time, there has been some disconnect between academic research on MT, which (rightly so) focuses on algorithms to increase translation quality, and many of the practical issues that need to be addressed to make MT maximally useful in real translation and localisation scenarios.

Objectives: This workshop brings together MT researchers, developers, industrial users and translators to discuss issues that are most important in real world industrial settings involving MT, but currently not very popular in research circles.

We would like to thank the members of the Program Committee for their time and effort in reviewing the submissions to the workshop as well as for their valuable comments to the authors. The workshop has a mixed Industry–Academia program committee to promote collaboration.

This year we are presenting six papers by authors from eight countries. Three papers have mixed industry–academic background, two have academic background only and one has only industrial background.

Industry members of the Program Committee: Manuel Tomás Carrasco Benítez (DGT of the EC), Marc Dymetman (XRCE), Daniel Grasmick (Lucy Software), Fred Hollowood (Symantec), Johann Roturier (Symantec), Dag Schmidtke (Microsoft), Jean Senellart (Systran), Nicolas Stroppa (Google)

Academic members of the Program Committee: Julien Bourdaillet (University of Montreal), Michael Carl (CBS, Denmark), Mikel Forcada (Universitat d'Alacant), Josef van Genabith (CNGL, EM+), Eiichiro Sumita (NICT, Japan), Philipp Koehn (EM+), Harold Somers (CNGL), Ventsislav Zhechev (EM+, CNGL)

Finally, we would like to thank the organisers of the main AMTA 2010 conference for their help, in particular Alon Lavie, George Foster, Michelle Vanni and Pricilla Rasmussen.

Second Joint EM+/CNGL Workshop Co-Chairs:
Ventsislav Zhechev, EuroMatrix+, CNGL, Dublin City University, Ireland
Philipp Koehn, EuroMatrix+, University of Edinburgh, UK
Josef van Genabith, CNGL, EuroMatrix+, Dublin City University, Ireland

`http://web.me.com/emcnglworkshop/JEC2010`

Second Joint EM+/CNGL Workshop Program

November 4th, 2010

09:00 – 09:05	Opening Remarks
09:05 – 10:00	**Invited Talk:** Creating Value at the Boundary Between Humans and Machines *Daniel Marcu, SDL Language Weaver*
10:00 – 10:30	Shared Resources, Shared Values? Ethical Implications of Sharing Translation Resources *Jo Drugan and Bogdan Babych*
10:30 – 11:00	Coffee Break
11:00 – 11:30	Machine Translation of TV Subtitles for Large Scale Production *Martin Volk, Rico Sennrich, Christian Hardmeier and Frida Tidström*
11:30 – 12:00	Source Text Characteristics and Technical and Temporal Post-Editing Effort: What is Their Relationship? *Midori Tatsumi and Johann Roturier*
12:00 – 12:30	Estimating Machine Translation Post-Editing Effort with HTER *Lucia Specia and Atefeh Farzindar*
12:30 – 02:00	Lunch Break
02:00 – 02:55	**Invited Talk:** t.b.a.
02:55 – 03:25	Convergence of Translation Memory and Statistical Machine Translation *Philipp Koehn and Jean Senellart*
03:25 – 03:55	Coffee Break
03:55 – 04:25	Integrating Machine Translation with Translation Memory: A Practical Approach *Panagiotis Kanavos and Dimitrios Kartsaklis*
04:25 – 05:25	**Round Table/Panel:** t.b.a.
05:25 – 05:30	Closing Remarks

Table of Contents

Author Index