2012 AMTA 20 Years

The Tenth Biennial Conference of the
Association for Machine Translation in the Americas

Post-Editing Technology and Practice

Sharon O'Brien
CNGL / DCU

Michel Simard
National Research Council Canada

Lucia Specia
University of Sheffield

SAN DIEGO, CA
OCTOBER 28– NOVEMBER 1, 2012

Proceedings of

WPTP 2012

AMTA 2012 Workshop on

Post-Editing Technology and Practice

Sharon O'Brien, Michel Simard and Lucia Specia
(editors)

28 October 2012
San Diego, California, USA

Foreword

Post-Editing has been around for just about as long as operational machine translation (MT) systems; as such, it is possibly the oldest form of human-machine cooperation for translation. Recently, however, there seems to be a surge of interest for post-editing among the wider user community, partly due to the increasing quality of MT output, but also to the availability of free, high-quality MT software.

Yet, the success of a post-editing operation depends on more than just software, and for every post-editing success story, probably many more failures go unreported. This workshop is an opportunity for post-editing researchers and practitioners to get together and openly discuss the weaknesses and strengths of existing technology, to properly and objectively assess post-editing effectiveness, to establish better practices, and propose tools and technological post-editing solutions that are built around the real needs of users.

The program consists of a mix of oral presentations, posters and software demonstrations. It is a snapshot of the wide variety of scientific and technological work currently taking place.

A number of researchers are tackling the difficult task of understanding the post-editing process itself, for example by studying the relationship between cognitive effort and post-editing time (Koponen et al.), or the relationship between cognitive effort and pauses (Lacruz et al.); others are examining the potential of crowdsourcing post-editing (Tatsumi et al.).

For these sorts of investigation to be effectively carried out, tools are required, specifically those designed for the purpose of observing post-editors and evaluating their work. This workshop features demonstrations and presentations of many such tools: the CASMACAT Workbench (Elming and Bonk), Transcenter (Denkowski and Lavie), PET (Aziz and Specia), and Ruqual (Melby et al.). New technology beyond tools for post-editing *per se* is also taking shape: tools for detecting MT errors (Valotkaite and Asadullah), tools for correcting them (Mundt et al.), or complete online post-editing frameworks with integrated MT functionalities (Penkale and Way).

Post-editing experiments are complex and costly, and it is critical that the experimental evidence that results is preserved and shared between researchers. This is the motivation behind the CRITT TPR database (Carl).

Finally, a special session on *Post-editing experiments in operational settings* will feature accounts of "real-life" experiments, such as recently took

place at Autodesk (Zhechev; Beregovaya and Moran) and various EU institutions (Poulis and Kolovratnik), as well as a report on GALA's ongoing "Post-editing Experiment" (Canek).

We wish to thank the AMTA people for making this event possible, providing logistical and moral support at all times. We must also thank the program committee for delivering high-quality reviews on a very tight schedule: you guys are the best.

Sharon O'Brien, Michel Simard and Lucia Specia

Oral Presentations and Posters

The CRITT TPR-DB 1.0: A Database For Empirical Human Translation Process Research

Michael Carl

Post-editing Time as a Measure of Cognitive Effort

Maarit Koponen, Wilker Aziz, Luciana Ramos and Lucia Specia

Average Pause Ratio as an Indicator of Cognitive Effort in Post-editing: A Case Study

Isabel Lacruz, Gregory M. Shreve and Erik Angelone

Reliably Assessing the Quality of Post-edited Translation Based on Formalized Structured Translation Specifications

Alan K Melby, Jason Housley, Paul J Fields and Emily Tuioti

Learning to Automatically Post-edit Dropped Words in MT

Jacob Mundt, Kristen Parton and Kathleen McKeown

SmartMATE: An Online End-To-End MT Post-editing Framework

Sergio Penkale and Andy Way

To post-edit or not to post-edit? Estimating the benefits of MT post-editing for a European organization

Alexandros Poulis and David Kolovratnik

How Good is Crowd Post-editing? Its Potential and Limitations

Midori Tatsumi, Takako Aikawa, Kentaro Yamamoto and Hitoshi Isahara

Error Detection for Post-editing Rule-based Machine Translation

Justina Valotkaite and Munshi Asadullah

Machine Translation Infrastructure and Post-editing Performance at Autodesk

Ventsislav Zhechev

Demos

PET: A Tool for Assessing Translation Quality Through Post-editing

Wilker Aziz and Lucia Specia

An Analysis of Machine Translation Post-Editing Productivity in Large-Scale Enterprise Deployments

Olga Beregovaya and John Moran

LSPs Experiment with MT Post-editing: Preliminary Results

David Canek

TransCenter: Web-Based Translation Research Suite

Michael Denkowski and Alon Lavie

The CASMACAT Workbench: a Tool for Investigating the Integration of Technology in Translation

Jakob Elming and Ragnar Bonk

SmartMATE: A Post-Editing Framework for Self-Serve Machine Translation

Sergio Penkale and Andy Way

Organizers

Sharon O'Brien – CNGL / DCU
Michel Simard – National Research Council Canada
Lucia Specia – University of Sheffield

Program Committee

Nora Aranberri – TAUS
Diego Bartolome – tauyou
Louise Brunette – Université du Québec en Outaouais
Michael Carl – Copenhagen Business School
Francisco Casacuberta – Universitat Politècnica de València
Mike Dillinger – Translation Optimization Partners
Stephen Doherty – Dublin City University
Andreas Eisele – European Commission
Jakob Elming – Copenhagen Business School
Atefeh Farzindar – NLP Technologies
Marcello Federico – FBK-IRST
Mikel L. Forcada – Universitat dAlacant
Ana Guerberof – Logoscript
Nizar Habash – Columbia University
Daniel Hardt – Copenhagen Business School
Kristian Tangsgaard Hvelplund – Copenhagen Business School
Pierre Isabelle – National Research Council Canada
Maxim Khalilov – TAUS
Philipp Koehn – University of Edinburgh
Roland Kuhn – National Research Council Canada
Philippe Langlais – RALI / Université de Montréal
Alon Lavie – Carnegie Mellon University
Elliott Macklovitch – Translation Bureau Canada
Daniel Marcu – SDL-LW / USC / ISI
John Moran – Transpiral Translation Services
Kristen Parton – Columbia University
Maja Popovic – DFKI
Alexandros Poulis – European Parliament/Intrasoft
Johann Roturier – Symantec

Jean Senellart – SYSTRAN
Myriam Siftar – MTM LinguaSoft
Roberto Silva – Celer Soluciones
Radu Soricut – SDL International
Midori Tatsumi – Dublin City University
Jörg Tiedemann – Uppsala University
Andy Way – Applied Language Solutions
Chris Wendt – Microsoft Research