



October 28 - November 1, 2016



The Twelfth Conference of
The Association for Machine Translation
in the Americas

<http://www.amtaweb.org/amta-2016-in-austin-tx>

**VOLUME 1:
MT Researchers' Track**

**Editors:
Spence Green & Lane Schwartz**

AMTA 2016

October 28 – November 1, 2016 -- Austin, TX, USA

Proceedings of
AMTA 2016,
Vol. 1: MT Researchers' Track

Spence Green & Lane Schwartz, Eds.



Association for Machine Translation in the Americas

<http://www.amtaweb.org>

©2016 The Authors. These articles are licensed under a Creative Commons 3.0 license, no derivative works, attribution, CC-BY-ND.

Introduction

AMTA occupies a unique place among MT research conferences, with one eye trained on the continuing development of novel research techniques and another on the practical application of those techniques in the commercial and government arenas. This year's research program reflects that diversity. This year we accepted 15 papers for oral presentation out of a total of 29 submissions. The accepted papers cover a nice variety of topics including low resource and interactive MT, post-editing, and domain adaptation. Over the past two years, the machine translation research landscape has undergone a rapid transition away from the phrase-based techniques that have been dominant for the past 10+ years, and towards new methods using artificial neural networks. This year's program reflects that trend, with a number of papers applying neural techniques across a broad range of MT tasks.

As chairs of the research track, we would like to thank all of the authors as well as the many reviewers whose hard work enabled this conference. It has also been a great pleasure to work with George Foster and the other members of the AMTA 2016 organizing committee.

Please enjoy this year's AMTA research track.

Lane Schwartz

Spence Green

Research Program Committee

Yaser	Al-Onaizan	Philipp	Koehn
Tim	Anderson	Roland	Kuhn
Amittai	Axelrod	Shankar	Kumar
Wilker	Aziz	Alon	Lavie
Graeme	Blackwood	Gregor	Leusch
Marine	Carpuat	Lemao	Liu
Daniel	Cer	Qun	Liu
Boxing	Chen	Saab	Mansour
Colin	Cherry	Daniel	Marcu
David	Chiang	Arne	Mauser
Steve	DeNeefe	Arul	Menezes
Loic	Dugast	Haitao	Mi
Nadir	Durrani	Michael	Pust
Marcello	Federico	Baskaran	Sankaran
Minwei	Feng	Rico	Sennrich
Mikel	Forcada	Michel	Simard
George	Foster	Raymond	Slyh
Kevin	Gimpel	Jörg	Tiedemann
Yvette	Graham	Christoph	Tillmann
Spence	Green	Taro	Watanabe
Eva	Hasler	Andy	Way
Yifan	He	Philip	Williams
Ulf	Hermjakob	Dekai	Wu
Hieu	Hoang	François	Yvon
Fei	Huang	Bing	Zhao
Matthias	Huck		
Marcin	Junczys-Dowmunt		

Contents

- 1 Instance Selection for Online Automatic Post-Editing in a Multi-domain Scenario
Rajen Chatterjee, Mihael Arcan, Matteo Negri and Marco Turchi
- 16 Machine Translation Quality and Post-Editor Productivity
Marina Sanchez-Torron and Philipp Koehn
- 27 Fuzzy-match repair using black-box machine translation systems: what can be expected?
John Ortega, Felipe Sánchez-Martínez and Mikel L. Forcada
- 40 Fast, Scalable Phrase-Based SMT Decoding
Hieu Hoang, Nikolay Bogoychev, Lane Schwartz and Marcin Junczys-Dowmunt
- 53 An Effective Diverse Decoding Scheme for Robust Synonymous Sentence Translation
Youngki Park, Hwidong Na, Hodong Lee, Jihyun Lee and Inchul Song
- 65 Ranking suggestions for black-box interactive translation prediction systems with multilayer perceptrons
Daniel Torregrosa, Juan Antonio Pérez-Ortiz and Mikel L. Forcada
- 79 Multi-domain Adaptation for Statistical Machine Translation Based on Feature Augmentation
Kenji Imamura and Eiichiro Sumita
- 93 Bilingual Methods for Adaptive Training Data Selection for Machine Translation
Boxing Chen, Roland Kuhn, George Foster, Colin Cherry and Fei Huang
- 107 Neural Interactive Translation Prediction
Rebecca Knowles and Philipp Koehn
- 121 Guided Alignment Training for Topic-Aware Neural Machine Translation
Wenhu Chen, Evgeny Matusov, Shahram Khadivi and Jan-Thorsten Peter

- 135 **Improving Neural Machine Translation on resource-limited pairs using auxiliary data of a third language**
Ander Martínez and Yuji Matsumoto
- 149 **Which Words Matter in Defining Phrase Reorderings in Statistical Machine Translation?**
Hamidreza Ghader and Christof Monz
- 163 **Translation of Unknown Words in Low Resource Languages**
Biman Gujral, Huda Khayrallah and Philipp Koehn
- 177 **Automatic Construction of Morphologically Motivated Translation Models for Highly Inflected, Low-Resource Languages**
John Hewitt, Matt Post and David Yarowsky
- 191 **Investigating the Impact of Various Partial Diacritization Schemes on Arabic-English Statistical Machine Translation**
Sawsan Alqahtani, Mahmoud Ghoneim and Mona Diab