

# **15th Conference of the Association for Machine Translation in the Americas (AMTA 2022)**

Volume 1: MT Research Track

Orlando, Florida, USA  
12 – 16 September 2022

## **Editors:**

**Kevin Duh  
Francisco Guzman**

**Stephen Richardson**

ISBN: 978-1-7138-6757-9

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2022) by the Association for Computational Linguistics  
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-ND. (<https://creativecommons.org/licenses/by-nd/4.0/>)

Printed with permission by Curran Associates, Inc. (2023)

For permission requests, please contact the Association for Computational Linguistics at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Contents

- 1 Building Machine Translation System for Software Product Descriptions Using Domain-specific Sub-corpora Extraction  
Pintu Lohar, Sinead Madden, Edmond O'Connor, Maja Popovic, Tanya Habruseva
- 14 Domain-Specific Text Generation for Machine Translation  
Yasmin Moslem, Rejwanul Haque, John Kelleher, Andy Way
- 31 Strategies for Adapting Multilingual Pre-training for Domain-Specific Machine Translation  
Neha Verma, Kenton W Murray, Kevin Duh
- 45 Prefix Embeddings for In-context Machine Translation  
Suzanna Sia, Kevin Duh
- 58 Fast Vocabulary Projection Method via Clustering for Multilingual Machine Translation on GPU  
Hossam Amer, Mohamed Afify, Young Jin Kim, Hitokazu Matsushita, Hany Hassan
- 70 Language Tokens: Simply Improving Zero-Shot Multi-Aligned Translation in Encoder-Decoder Models  
Muhammad N ElNokrashy, Amr Hendy, Mohamed Maher, Mohamed Afify, Hany Hassan
- 83 Low Resource Chat Translation: A Benchmark for Hindi--English Language Pair  
Baban Gain, Ramakrishna Appicharla, Asif Ekbal, muthusamy chelliah, Soumya Chennabasavraj, Nikesh Garera

- 97      How Robust is Neural Machine Translation to Language Imbalance in Multilingual Tokenizer Training?  
  
Shiyue Zhang, Vishrav Chaudhary, Naman Goyal, James Cross, Guillaume Wenzek, Mohit Bansal, Francisco Guzman
- 117     How Effective is Byte Pair Encoding for Out-Of-Vocabulary Words in Neural Machine Translation?  
  
Ali Araabi, Christof Monz, Vlad Niculae
- 131     On the Effectiveness of Quasi Character-Level Models for Machine Translation  
  
Salvador Carrión-Ponz, Francisco Casacuberta
- 144     Improving Translation of Out Of Vocabulary Words using Bilingual Lexicon Induction in Low-Resource Machine Translation  
  
Jonas Waldendorf, Alexandra Birch, Barry Hadow, Antonio Valerio Micele Barone
- 157     Doubly-Trained Adversarial Data Augmentation for Neural Machine Translation  
  
Weiting Tan, Shuoyang Ding, Huda Khayrallah, Philipp Koehn
- 175     Limitations and Challenges of Unsupervised Cross-lingual Pre-training  
  
Martín Quesada Zaragoza, Francisco Casacuberta
- 188     Few-Shot Regularization to Tackle Catastrophic Forgetting in Multilingual Machine Translation  
  
Salvador Carrión-Ponz , Francisco Casacuberta
- 200     Quantized Wasserstein Procrustes Alignment of Word Embedding Spaces  
  
Prince O Aboagye, Yan Zheng, Michael Yeh, Junpeng Wang, Zhongfang Zhuang, Huiyuan Chen, Liang Wang, Wei Zhang, Jeff Phillips

- 215 Refining an Almost Clean Translation Memory Helps Machine Translation  
Shivendra Bhardwa, David Alfonso-Hermelo, Philippe Langlais, Gabriel Bernier-Colborne, Cyril Goutte, Michel Simard
- 227 Practical Attacks on Machine Translation using Paraphrase  
Elizabeth M Merkhofer, John Henderson, Abigail Gertner, Michael Doyle, Lily Wong
- 240 Sign Language Machine Translation and the Sign Language Lexicon: A Linguistically Informed Approach  
Irene Murtagh, Víctor Ubieta Nogales, Josep Blat
- 252 A Neural Machine Translation Approach to Translate Text to Pictographs in a Medical Speech Translation System - The BabelDr Use Case  
Jonathan Mutal, Pierrette Bouillon, Magali Norré, Johanna Gerlach, Lucia Ormaechea Grijalba
- 264 Embedding-Enhanced GIZA++: Improving Low-Resource Word Alignment Using Embeddings  
Kelly Marchisio, Conghao Xiong, Philipp Koehn
- 274 Gender bias Evaluation in Luganda-English Machine Translation  
Eric Peter Wairagala, Jonathan Mukiibi, Jeremy Francis Tsubira, Claire Babirye, Joyce Nakatumba-Nabende, Andrew Katumba, Ivan Ssenkungu
- 287 Adapting Large Multilingual Machine Translation Models to Unseen Low Resource Languages via Vocabulary Substitution and Neuron Selection  
Mohamed A Abdelghaffar, Amr El Mogy, Nada Ahmed Sharaf
- 298 Measuring the Effects of Human and Machine Translation on Website Engagement  
Geza Kovacs, John DeNero

- 309 Consistent Human Evaluation of Machine Translation across Language Pairs  
Daniel Licht, Cynthia Gao, Janice Lam, Francisco Guzman, Mona Diab, Philipp Koehn
- 322 Evaluating Machine Translation in Cross-lingual E-Commerce Search  
Hang Zhang, Liling Tan, Amita Misra