The 15th Conference of the Association for Machine Translation in the Americas

2022.amtaweb.org

PROCEEDINGS

Volume 1: MT Research Track

Editor:
Kevin Duh, Francisco Guzman (Research Track Co-chairs)
Stephen Richardson (General Conference Chair)

Welcome to the 15th biennial conference of the Association for Machine Translation in the Americas – AMTA 2022!

Dear MT Colleagues and Friends,

For this year's conference of the Association for Machine Translation in the Americas – AMTA 2022 – we are finally able to come together in person at the venue we had intended to enjoy two years ago, the spectacular Sheraton Orlando Lake Buena Vista Resort in Orlando, Florida! We are very grateful that the COVID pandemic is now sufficiently controlled (albeit still with us) that we can once again meet, network, and enjoy one another's company while expanding our knowledge of the ever-accelerating field of machine translation. At the same time, we will be joined by likely more than twice the number of remote attendees, as the last two years of virtual conferences and ongoing health concerns will forever more require us to adopt a hybrid conference format. While this format certainly creates complexity for organizers, and it can feel a little less personal as we interact with remote speakers and attendees, it nevertheless provides significantly greater accessibility and opportunities to learn from colleagues around the globe. We are grateful for their very positive contributions to our conference!

Since the MT Summit we hosted last year, we have continued to witness amazing progress in MT technology and tremendous growth in the adoption of this technology by individual translators, language services providers, small businesses, large enterprises, non-profits, governments, and NGOs. Indeed, a unique aspect of AMTA conferences is that it brings together users and practitioners from across the MT spectrum of academia, industry, and government so that R&D personnel can learn from those who are using the technology and vice versa.

We are pleased once again with the number of submissions to our conference. As MT has become more mainstream than ever, we have had to be more selective in the presentations included in our conference tracks. This is unfortunate on the one hand, but on the other, it demonstrates the growth of our field and the increasing quality and relevance of the work performed by so many people. Of special note this year is the emphasis on speech translation and dubbing, MT quality evaluation, and massively multilingual MT systems. These topics are reflected by the topics of our keynote speakers and panels in the conference schedule, and we trust you will find them most enlightening.

As with all our conferences, AMTA 2022 would simply not have been possible without the selfless work of so many people on the AMTA board and organizing committee, all of whom are volunteers. I express my deepest thanks, respect, and admiration to each one of them. They include:

Patti O'Neill-Brown, AMTA VP, Local Arrangements, Networking
Natalia Levitina, AMTA Secretary, Sponsorships
Jen Doyon, AMTA Treasurer, Local Arrangements
Kevin Duh, Research Track
Paco Guzman, Research Track
Janice Campbell, Users and Providers Track, Networking
Jay Marciano, Users and Providers Track, Workshops and Tutorials
Konstantin Savenkov, Users and Providers Track

Alex Yanishevsky, Users and Providers Track, Conference Online Platform
Steve La Rocca, Government Track
Kenton Murray, Student Mentoring,
Konstantin Dranch, Communications
Lara Daly, Marketing
Alon Lavie, AMTA Consultant
Elaine O'Curran, AMTA Counselor, Publications
Elliott Macklovitch, Publications
Derick Fajardo, Exhibitions

Finally, I express my gratitude to our amazing sponsors, whose tremendous financial support has enabled us to handle the added complexity and cost of the hybrid format. Once again, greatly discounted student registrations have been provided by Microsoft, our Visionary++ sponsor, as well as an included conference banquet for in-person attendees. Systran has also contributed significantly to our online platforms as a Visionary sponsor. Our Leader-level sponsors are Pangeanic, Meta, Acclaro, AppTek, and Intento, and our Patron-level sponsors are AWS, Google RWS, Star, and Welocalize. Additional exhibitors are ModelFront and Unbabel, and our Media and Marketing sponsors are Slator, Multilingual, and Akorbi. Many of these sponsors and exhibitors will provide demonstrations of their systems and software during our Technology Exhibition sessions, and we hope that all our attendees will take advantage of this great opportunity to see the very latest commercial offerings and advancements in the world of MT.

Again, welcome to AMTA 2022! I look forward to finally being with many of you in person in Orlando and to interacting with many others online.

Steve Richardson
AMTA President and AMTA 2022 General Conference Chair

Introduction

The research track at AMTA 2022 continues the tradition of bringing MT practitioners together from academia, industry and government from around the world.

This year we have a very rich program with 25 papers from a variety of topics. The most popular subject this year is low-resource machine translation (32%), spanning from pre-training and adaptation to unseen languages, to gender bias evaluation for low-resource languages. In addition, we have many works discussing pre-processing and data adaptation (e.g. analyses on subword tokenization); applications of MT (e.g. website engagements, e-commerce search); and even papers discussing sign language translation. We are also excited about our invited keynote speakers for the research track: Angela Fan (Meta AI) will talk about Massively Multilingual MT.

We hope that this conference brings many productive exchanges of ideas and sparks future collaborations.

We would like to thank the hard work of individuals that made this happen: the authors, the reviewers, the timely emergency reviewers, the AMTA organizing committee; and Akiko Eriguchi for her help in preparing the proceedings and organizing session chairs.

Sincerely,

Kevin Duh and Francisco Guzmán (Research Track Co-Chairs)

Program committee

Abraham Glasser (Rochester Institute of Technology)
Akiko Eriguchi (Microsoft)
Alina Karakanta (Fondazione Bruno Kessler)
M. Amin Farajian (Unbabel)
Asif Ekbal (IIT Patna)
Atsushi Fujita (NICT, Japan)
Beatrice Savoldi (UNINT)
Bing Zhao (SRI International)
Boxing Chen (Alibaba Group)
Chao-Hong Liu (ADAPT Centre, Dublin City University)
Chenhui Chu (Kyoto University)
Christian Federmann (Microsoft)
Claudio Russello (UNINT)
Colin Cherry (Google)
Constantin Orasan (University of Surrey)
David Chiang (University of Notre Dame)
Derek F. Wong (University of Macau)
Dimitar Shterionov (Tilburg University)
Duygu Ataman (UZH)
Elijah Rippeth (University of Maryland)
Evgeny Matusov (AppTek)
Federico Gaspari (ADAPT Centre, Dublin City University)
Flammie Pirinen (UiT–Norgga Árktalaš Universitehta)
François YVON (CNRS)
Hailong Cao (Harbin Institute of Technology)
Haitao Mi (Tencent America)
Huda Khayrallah (Johns Hopkins University)
Jampierre Rocha (Lenovo)
Jan Niehues (KIT)
Jasper Kyle Catapang (University of Birmingham)
Jeremy Gwinnup (Air Force Research Laboratory)
John McCrae (National University of Ireland Galway)
Jörg Tiedemann (University of Helsinki)
Josep Maria Crego (Systran)
Juan Pino (Meta AI)
Katharina Kann (University of Colorado Boulder)
Katsuhito Sudoh (NAIST)
Kelly Marchisio (Johns Hopkins University)

Kenji Imamura (NICT)
Kevin Duh (Johns Hopkins University)
Koichiro Watanabe (The University of Tokyo)
Loic Barrault (Meta AI)
Marco Gaido (Fondazione Bruno Kessler)
Marco Turchi (Zoom)
Maria Antonette Clariño (University of the Philippines)
Marianna Martindale (University of Maryland)
Mathias Müller (University of Zurich)
Matthias Huck (SAP SE)
Mehdi Rezagholizadeh (Huawei Noah's Ark Lab)
Nathaniel Oco (De La Salle University, Philippines)
Neha Verma (Johns Hopkins University)
Ohnmar Htun (Rakuten Asia Pte.Ltd.)
Patrick Simianer (Lilt)
Philipp Koehn (Johns Hopkins University)
Priya Rani (National University of Ireland Galway)
Raj Dabre (NICT)
Rebecca Knowles (National Research Council Canada)
Rico Sennrich (University of Zurich)
Rosalee Wolfe (ILSP / Athena RC)
Sangjie Duanzhu (Qinghai Normal University)
Santanu Pal (Saarland University)
Sara Papi (FBK)
Shankar Kumar (Google)
Shinji Watanabe (Carnegie Mellon University)
Sunit Bhattacharya (Charles University)
Takashi Ninomiya (Ehime University)
Taro Watanabe (NAIST)
Tetsuji Nakagawa (Google Japan G.K.)
Thepchai Supnithi (NECTEC, National Science and Technology Development Agency)
Tomek Korybski (University of Surrey)
Toshiaki Nakazawa (The University of Tokyo)
Tsz Kin Lam (Heidelberg University)
Valentin Malykh (Huawei Noah's Ark lab)
Vedanuj Goswami (Meta AI)
Vishrav Chaudhary (Microsoft)
Xinyi Wang (Carnegie Mellon University)
Xuan Zhang (Johns Hopkins University)

Contents

345 Evaluating Machine Translation in Cross-lingual E-Commerce Search

Hang Zhang, Liling Tan, Amita Misra