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

2022.amtaweb.org

PROCEEDINGS

Volume 2:
MT Users & Providers Track
and
Government Track

Editors

Government Track Chair
Stephen Larocca

Users & Providers Track Co-chairs
Janice Campbell, Jay Marciano,
Konstantin Savenkov and
Alex Yanishevsky

General Conference Chair
Stephen Richardson

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

User/Provider Track: Introduction

The User/Provider Track at AMTA 2022 features twenty-six presentations by and for machine translation experts and practitioners, language service providers, technology service providers, universities, linguists, and commercial enterprises.

We are privileged to have Marco Trombetti, a renowned computer scientist, entrepreneur and investor as well as Co-Founder and CEO of Translated, one of the first companies to utilize AI in translation, as the keynote speaker for the track.

The latest State of Machine Translation report will present MT engine performance results across industries and language pairs and provide additional details about scoring methodologies.

New this year is a presentation on a machine translation non-profit organization whose goals are to provide access to open resources as well as build a community of contributors.

As would be expected in a commercial track, there are presentations which focus on making business cases showing the financial and market benefits of incorporating MT in the translation workflow. Case studies carried out jointly by a technology and/or language service provider and a client, showcase real world use cases.

Recurring themes at this conference continue to be data, engine training, AI applications, low resource languages and PEMT.

Quality of MT output is a matter of concernment in the industry and there are several presentations addressing it from various perspectives. A range of topics are presented, such as monitoring, assessing, predicting quality outcomes and applying risk modeling. Source-based Quality Estimation against TMs is offered as a new approach. Automatic Post Editing is improved by leveraging GPT-3 features. Setting customer quality expectations can be achieved by defining Business Critical Errors. Finally, commonly applied auto scores are compared to the ATA grading framework.

Translating speech is rapidly growing in importance. Presentations on this topic include methods to connect subtitles to the correct speakers; STT/TTS for audio visual translation using neural voices; voice synthesis for e-learning content; and real-time simultaneous interpreting with automatic dubbing and STS translation.

As far as engine training and model fine-tuning, presentation topics focus primarily on data used as input for training. Data augmentation, quality vs quantity, deep learning to achieve better segmentation and alignment, and advanced filtering techniques are discussed. Customizing NMT for limited support language pairs and regional language variants are also discussed. One presentation challenges the sustainability of the engine training process by promoting knowledge distillation to decrease power consumption.

Finally, there are presentations that focus on challenges in very specific domains: MT for video gaming, where in-domain data is quite limited; patent translations which must hold up to intense legal and scientific scrutiny.

We would like to thank the AMTA organizing committee for the intense planning that went into hosting a hybrid conference. We also thank the session and keynote speakers for their excellent presentations. We are especially grateful to the volunteer moderators for supporting the speakers, fielding the questions and keeping the presentations on schedule.

Sincerely,

Janice Campbell, Jay Marciano, Konstantin Savenkov, Alex Yanishevsky

The User/Provider Track Co-Chairs

Government Track: Introduction

The Government Track at AMTA 2022 features eleven presentations. North American government issues in machine translation figure prominently, with the Government of Canada's Translation Bureau and the United States' government efforts sharing eight of the eleven presentations.

Contributions from colleagues overseas are of course most welcome, including those from the Dalian University of Foreign Languages in the People's Republic of China and Singapore's Ministry of Communications and Information. Likewise, SYSTRAN, a multinational corporation with a very long history of providing translation technology to governments, is a welcome contribution to the government track program at AMTA 2022.

The government track is proud to be associated with Dr. Alex Waibel of Carnegie Mellon University and Karlsruhe Institute of Technology who is our Keynote Speaker. Dr. Waibel also anchors the special panel on Advances in Spoken Language MT, an area of translation technology in which Alex's contributions are unmatched and where interest by government entities is on the rise.

Cordially,

Steve LaRocca

Government Track Chair, "standing on the shoulders" of those who precede me

Contents

Users and Providers Track

Government Track