



Proceedings of the  
**23rd Annual Conference of  
the European Association  
for Machine Translation**

1 – 3 June 2022  
Ghent, Belgium

*Edited by*

Lieve Macken, Andrew Rufener, Joachim Van den Bogaert, Joke Daems, Arda Tezcan, Bram Vanroy, Margot Fonteyne, Loïc Barrault, Marta R. Costa-jussà, Ellie Kemp, Spyridon Pilos, Christophe Declercq, Maarit Koponen, Mikel L. Forcada, Carolina Scarton, Helena Moniz

*Organised by*







The papers published in this proceedings are —unless indicated otherwise— covered by the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 International (CC-BY-ND 3.0). You may copy, distribute, and transmit the work, provided that you attribute it (authorship, proceedings, publisher) in the manner specified by the author(s) or licensor(s), and that you do not use it for commercial purposes. The full text of the licence may be found at <https://creativecommons.org/licenses/by-nc-nd/3.0/deed.en>.

© 2022 The authors

ISBN: 9789464597622





# Contents

|  |             |
|--|-------------|
| <b>Foreword from the General Chair</b> . . . . .   | <b>vii</b>  |
| <b>Message from the Organising Committee Chairs</b> . . . . .  | <b>ix</b>   |
| <b>Preface by the Programme Chairs</b> . . . . .   | <b>xi</b>   |
| <b>EAMT 2022 Committees</b> . . . . .  | <b>xiii</b> |
| <b>Invited Speeches</b> . . . . .  | <b>1</b>    |
| <b>EAMT 2020 and EAMT 2021 Best Thesis Award — Anthony C Clarke Award</b>  | <b>3</b>    |
| Maha Elbayad. <i>Rethinking the Design of Sequence-to-Sequence Models for Efficient Machine Translation</i> . . . . .  | 5           |
| Mattia Antonino Di Gangi. <i>Neural Speech Translation: From Neural Machine Translation to Direct Speech Translation</i> . . . . .   | 7           |
| Danielle Saunders. <i>Domain Adaptation for Neural Machine Translation</i> . . . . .   | 9           |
| <b>Research papers</b> . . . . .   | <b>11</b>   |
| Minh-Quang Pham, Josep Crego and François Yvon. <i>Multi-Domain Adaptation in Neural Machine Translation with Dynamic Sampling Strategies</i> . . . . .  | 13          |
| Rudy Loock, Sophie Léchauguette and Benjamin Holt. <i>The use of online translators by students not enrolled in a professional translation program: beyond copying and pasting for a professional use</i> . . . . .          | 23          |
| Xabier Soto, Olatz Perez-De-Viñaspre, Gorka Labaka and Maite Oronoz. <i>Comparing and combining tagging with different decoding algorithms for back-translation in NMT: learnings from a low resource scenario</i> . . . . . | 31          |
| Dongqi Pu and Khalil Sima’an. <i>Passing Parser Uncertainty to the Transformer. Labeled Dependency Distributions for Neural Machine Translation.</i> . . . . .   | 41          |
| Mark Pluymaekers. <i>How well do real-time machine translation apps perform in practice? Insights from a literature review</i> . . . . .   | 51          |
| Ricardo Rei, Ana C Farinha, José G.C. de Souza, Pedro G. Ramos, André F.T. Martins, Luisa Coheur and Alon Lavie. <i>Searching for COMETINHO: The Little Metric That Could</i> . . . . .                                      | 61          |
| Lise Volkart and Pierrette Bouillon. <i>Studying Post-Editese in a Professional Context: A Pilot Study</i> . . . . .   | 71          |
| Minghan Wang, Jiaxin Guo, Yuxia Wang, Daimeng Wei, Hengchao Shang, Yinglu Li, Chang Su, Yimeng Chen, Min Zhang, Shimin Tao and Hao Yang. <i>Diformer: Directional Transformer for Neural Machine Translation</i> . . . . .   | 81          |
| Taido Purason and Andre Tättar. <i>Multilingual Neural Machine Translation With the Right Amount of Sharing</i> . . . . .  | 91          |

|  |            |
|--|------------|
| Lieve Macken, Bram Vanroy, Luca Desmet and Arda Tezcan. <i>Literary translation as a three-stage process: machine translation, post-editing and revision</i> . . . . .   | 101        |
| Àlex R. Atrio and Andrei Popescu-Belis. <i>On the Interaction of Regularization Factors in Low-resource Neural Machine Translation</i> . . . . .   | 111        |
| Sebastian T. Vincent, Loïc Barrault and Carolina Scarton. <i>Controlling Extra-Textual Attributes about Dialogue Participants: A Case Study of English-to-Polish Neural Machine Translation</i> . . . . .                                | 121        |
| Nishant Kambhatla, Logan Born and Anoop Sarkar. <i>Auxiliary Subword Segmentations as Related Languages for Low Resource Multilingual Translation</i> . . . . .  | 131        |
| Pedro Mota, Vera Cabarrão and Eduardo Farah. <i>Fast-Paced Improvements to Named Entity Handling for Neural Machine Translation</i> . . . . .  | 141        |
| Alina Kramchaninova and Arne Defauw. <i>Synthetic Data Generation for Multilingual Domain-Adaptable Question Answering Systems</i> . . . . .   | 151        |
| Tobias van der Werff, Rik van Noord and Antonio Toral. <i>Automatic Discrimination of Human and Neural Machine Translation: A Study with Multiple Pre-Trained Models and Longer Context</i> . . . . .                                    | 161        |
| Khetam Al Sharou and Lucia Specia. <i>A Taxonomy and Study of Critical Errors in Machine Translation</i> . . . . .   | 171        |
| <b>User papers</b> . . . . .   | <b>181</b> |
| Artur Nowakowski, Krzysztof Jassem, Maciej Lison, Rafał Jaworski, Tomasz Dwojak, Karolina Wiater and Olga Posesor. <i>nEYron: Implementation and Deployment of an MT System for a Large Audit &amp; Consulting Corporation</i> . . . . . | 183        |
| Bianka Buschbeck, Jennifer Mell, Miriam Exel and Matthias Huck. <i>“Hi, how can I help you?” Improving Machine Translation of Conversational Content in a Business Context</i> . . . . .   | 191        |
| Madalena Gonçalves, Marianna Buchicchio, Craig Stewart, Helena Moniz and Alon Lavie. <i>Agent and User-Generated Content and its Impact on Customer Support MT</i> . . . . .   | 201        |
| Miguel Menezes, Vera Cabarrão, Pedro Mota, Helena Moniz, and Alon Lavie. <i>A Case Study on the Importance of Named Entities in a Machine Translation Pipeline for Customer Support Content</i> . . . . .                                | 211        |
| Maria Afara, Randy Scansani and Loïc Dugast. <i>Investigating automatic and manual filtering methods to produce MT-ready glossaries from existing ones</i> . . . . .   | 221        |
| Celia Soler Uguet, Fred Bane, Anna Zaretskaya and Tània Blanch Miró. <i>Comparing Multilingual NMT Models and Pivoting</i> . . . . .   | 231        |
| Kamal Kumar Gupta, Soumya Chennabasavraj, Nikesh Garera and Asif Ekbal. <i>Pre-training Synthetic Cross-lingual Decoder for Multilingual Samples Adaptation in E-Commerce Neural Machine Translation</i> . . . . .                       | 241        |
| <b>Translators’ papers</b> . . . . .   | <b>249</b> |
| Justus Brockmann, Claudia Wiesinger and Dragoş Ciobanu. <i>Error Annotation in Post-Editing Machine Translation: Investigating the Impact of Text-to-Speech Technology</i>   | 251        |
| Alina Karakanta, Luisa Bentivogli, Mauro Cettolo, Matteo Negri and Marco Turchi. <i>Post-editing in Automatic Subtitling: A Subtitlers’ perspective</i> . . . . .  | 261        |
| Sabrina Girletti. <i>Working with Pre-translated Texts: Preliminary Findings from a Survey on Post-editing and Revision Practices in Swiss Corporate In-house Language Services</i> . . . . .  | 271        |
| <b>Project/product descriptions</b> . . . . .  | <b>281</b> |

|  |     |
|--|-----|
| Arda Tezcan. <i>Dynamic Adaptation of Neural Machine-Translation Systems Through Translation Exemplars</i> . . . . .   | 283 |
| Diego Bartolome and Chris Jacob. <i>Language I/O Solution for Multilingual Customer Support</i> . . . . .  | 285 |
| Fernando Alva-Manchego and Matthew Shardlow. <i>Towards Readability-Controlled Machine Translation of COVID-19 Texts</i> . . . . .   | 287 |
| Joke Daems and Janiça Hackenbuchner. <i>DeBiasByUs: Raising Awareness and Creating a Database of MT Bias</i> . . . . .   | 289 |
| Mikel L. Forcada, Pilar Sánchez-Gijón, Dorothy Kenny, Felipe Sánchez-Martínez, Juan Antonio Pérez Ortiz, Riccardo Superbo, Gema Ramírez Sánchez, Olga Torres-Hostench and Caroline Rossi. <i>MultitraiNMT Erasmus+ project: Machine Translation Training for multilingual citizens (multitrainmt.eu)</i> . . . . .   | 291 |
| Senja Pollak and Andraž Pelicon. <i>EMBEDDIA project: Cross-Lingual Embeddings for Less- Represented Languages in European News Media</i> . . . . .  | 293 |
| Masaru Yamada, Takanori Mizowaki, Longhui Zou and Michael Carl. <i>Trados-to-Translog-II: Adding Gaze and Quality data to the CRITT TPR-DB</i> . . . . .   | 295 |
| Margot Fonteyne, Maribel Montero Perez, Joke Daems and Lieve Macken. <i>Writing in a second Language with Machine translation (WiLMa)</i> . . . . .  | 297 |
| Eirini Kaldeli, Mercedes García-Martínez, Antoine Isaac, Paolo Sebastiano Scalia, Arne Stabenau, Iván Lena Almor, Carmen Grau Lacal, Martín Barroso Ordóñez, Amando Estela and Manuel Herranz. <i>Europeana Translate: Providing multilingual access to digital cultural heritage</i> . . . . .  | 299 |
| Pierrette Bouillon, Johanna Gerlach, Jonathan Mutal and Marianne Starlander. <i>The PASSAGE project : Standard German Subtitling of Swiss German TV content</i> . . . . .  | 301 |
| Marta Bañón, Miquel Esplà-Gomis, Mikel L. Forcada, Cristian García-Romero, Taja Kuzman, Nikola Ljubešić, Rik van Noord, Leopoldo Pla Sempere, Gema Ramírez-Sánchez, Peter Rupnik, Vít Suchomel, Antonio Toral, Tobias van der Werff and Jaume Zaragoza. <i>MaCoCu: Massive collection and curation of monolingual and bilingual data: focus on under-resourced languages</i> . . . . . | 303 |
| Sheila Castilho and Natália Resende. <i>MT-Pese: Machine Translation and Post-Editese</i> . . . . .  | 305 |
| Elena Murgolo, Javad Pourmostafa Roshan Sharami and Dimitar Shterionov. <i>A Quality Estimation and Quality Evaluation Tool for the Translation Industry</i> . . . . .   | 307 |
| Toms Bergmanis, Marcis Pinnis, Roberts Rozis, Jānis Šlapiņš, Valters Šics, Berta Bernāne, Guntars Pužulis, Endijs Titomers, Andre Tättar, Taido Purason, Hele-Andra Kuulmets, Agnes Luhtaru, Liisa Rätsep, Maali Tars, Annika Laumets-Tättar and Mark Fishel. <i>MTee: Open Machine Translation Platform for Estonian Government</i> . . . . .   | 309 |
| Raúl Vázquez, Michele Boggia, Alessandro Raganato, Niki A. Loppi, Stig-Arne Grønroos and Jörg Tiedemann. <i>Latest Development in the FoTran Project – Scaling Up Language Coverage in Neural Machine Translation Using Distributed Training with Language-Specific Components</i> . . . . .   | 311 |
| Gabriele Sarti and Arianna Bisazza. <i>InDeep × NMT: Empowering Human Translators via Interpretable Neural Machine Translation</i> . . . . .   | 313 |
| José G.C. de Souza, Ricardo Rei, Ana C. Farinha, Helena Moniz and André F. T. Martins. <i>QUARTZ: Quality-Aware Machine Translation</i> . . . . .  | 315 |
| Artur Nowakowski, Krzysztof Jassem, Maciej Lison, Kamil Guttmann and Mikołaj Pokrywka. <i>POLENG MT: An Adaptive MT Platform</i> . . . . .   | 317 |
| Carlos Amaral and Peggy van der Kreeft. <i>plain X - AI Supported Multilingual Video Workflow Platform</i> . . . . .   | 319 |

|   |     |
|---|-----|
| Sheila Castilho. <i>DELA Project: Document-level Machine Translation Evaluation</i> . . .   | 321 |
| Giorgio Bernardinello and Judith Klein. <i>Background Search for Terminology in STAR MT Translate</i> . . . . .   | 323 |
| Dimitar Shterionov, Mirella De Sisto, Vincent Vandeghinste, Aoife Brady, Mathieu De Coster, Lorraine Leeson, Josep Blat, Frankie Picron, Marcello Paolo Scipioni, Aditya Parikh, Louis ten Bosh, John O’Flaherty, Joni Dambre and Jorn Rijckaert. <i>Sign Language Translation: Ongoing Development, Challenges and Innovations in the SignON Project</i> . . . . . | 325 |
| André F. T. Martins, Ben Peters, Chrysoula Zerva, Chunchuan Lyu, Gonçalo Correia, Marcos Treviso, Pedro Martins and Tsvetomila Mihaylova. <i>DeepSPIN: Deep Structured Prediction for Natural Language Processing</i> . . . . .   | 327 |
| Dimitra Anastasiou, Anders Ruge, Radu Ion, Svetlana Segărceanu, George Suciuc, Olivier Pedretti, Patrick Gratz and Hoorieh Afkari. <i>A Machine Translation-Powered Chatbot for Public Administration</i> . . . . .   | 329 |
| Natalia Resende. <i>MTrill: Machine Translation Impact on Language Learning</i> . . . . .   | 331 |
| Jourik Ciesielski and Heidi Van Hiel. <i>Connecting client infrastructure with Yamagata Europe machine translation using JSON-based data exchange</i> . . . . .   | 333 |
| Alina Karakanta, Luisa Bentivogli, Mauro Cettolo, Matteo Negri and Marco Turchi. <i>Towards a methodology for evaluating automatic subtitling</i> . . . . .   | 335 |
| Ekaterina Lapshinova-Koltunski, Maja Popović and Maarit Koponen. <i>DiHuTra: a Parallel Corpus to Analyse Differences between Human Translations</i> . . . . .  | 337 |
| Peggy van der Kreeft, Alexandra Birch, Sevi Sariisik, Felipe Sánchez-Martínez and Wilker Aziz. <i>GoURMET – Machine Translation for Low-Resourced Languages</i> . . .   | 339 |
| Tamás Váradi, Marko Tadić, Svetla Koeva, Maciej Ogrodniczuk, Dan Tufiş, Radovan Garabík, Simon Krek and Andraž Repar. <i>Curated Multilingual Language Resources for CEF AT (CURLICAT): overall view</i> . . . . .  | 341 |
| Joachim Van den Bogaert, Laurens Meeus, Alina Kramchaninova, Arne Defauw, Sara Szoc, Frederic Everaert, Koen Van Winckel, Anna Bardadym and Tom Vanallemeersch. <i>Automatically extracting the semantic network out of public services to support cities becoming Smart Cities</i> . . . . .   | 343 |
| Artūrs Vasīļevskis, Jānis Ziedīņš, Marko Tadić, Željka Motika, Mark Fishel, Hrafn Loftsson, Jón Guðnason, Claudia Borg, Keith Cortis, Judie Attard and Donatienne Spiteri. <i>National Language Technology Platform (NLTP): overall view</i> . . .  | 345 |
| Anabela Barreiro, José GC de Souza, Albert Gatt, Mehul Bhatt, Elena Lloret, Aykut Erdem, Dimitra Gkatzia, Helena Moniz, Irene Russo, Fabio Kepler, Iacer Calixto, Marcin Paprzycki, François Portet, Isabelle Augenstein and Mirela Alhasani. <i>Multi3Generation: Multitask, Multilingual, Multimodal Language Generation</i> . . .                                | 347 |
| Petra Bago, Sheila Castilho, Jane Dunne, Federico Gaspari, Andre Kåsen, Gauti Kristmannsson, Jon Arild Olsen, Natalia Resende, Níels Rúnar Gíslason, Dana D. Sheridan, Páraic Sheridan, John Tinsley and Andy Way. <i>Achievements of the PRINCIPLE Project: Promoting MT for Croatian, Icelandic, Irish and Norwegian</i>  | 349 |
| Mattia Di Gangi, Nick Rossenbach, Alejandro Pérez, Parnia Bahar, Eugen Beck, Patrick Wilken and Evgeny Matusov. <i>Automatic Video Dubbing at AppTek</i> . . .  | 351 |
| Itziar Aldabe, Jane Dunne, Aritz Farwell, Owen Gallagher, Federico Gaspari, Maria Giagkou, Jan Hajic, Jens Peter Kückens, Teresa Lynn, Georg Rehm, German Rigau, Katrin Marheinecke, Stelios Piperidis, Natalia Resende, Tea Vojtěchová and Andy Way. <i>Overview of the ELE Project</i> . . . . .  | 353 |
| Maarit Koponen, Kais Allkivi-Metsoja, Antonio Pareja-Lora, Dave Sayers and Márta Seresi. <i>LITHME: Language in the Human-Machine Era</i> . . . . .   | 355 |

|  |            |
|--|------------|
| Ana Guerberof Arenas and Antonio Toral. <i>CREAMT: Creativity and narrative engagement of literary texts translated by translators and NMT</i> . . . . .   | 357        |
| Pintu Lohar, Guodong Xie and Andy Way. <i>Developing Machine Translation Engines for Multilingual Participatory Spaces</i> . . . . .   | 359        |
| Luisa Bentivogli, Mauro Cettolo, Marco Gaido, Alina Karakanta, Matteo Negri and Marco Turchi. <i>Extending the MuST-C Corpus for a Comparative Evaluation of Speech Translation Technology</i> . . . . . | 361        |
| Carlos Amaral and Sebastião Miranda. <i>Monitio - Large Scale MT for Multilingual Media Monitoring</i> . . . . .   | 363        |
| <b>Sponsors</b> . . . . .  | <b>365</b> |



# Foreword from the General Chair

As president of the European Association for Machine Translation (EAMT) and General Chair of the 23rd Annual Conference of the EAMT, it is with great pleasure that I write these opening words to the Proceedings of EAMT 2022 (a first time for me!). The preparations for EAMT 2022 were initially started by the former President, Mikel Forcada, to whom I am deeply grateful for all the assistance and hand over.

A first note of appreciation and gratitude to the Executive Board Members who have moved to new plans in life, after long and outstanding dedicated service to the EAMT community. Firstly, Tony Clarke, EAMT treasurer for 23 years, in appreciation for his invaluable service as the longest-standing treasurer of our Association. To Andy Way, in appreciation for his years of service as secretary, president, conference organizer, and member of the Executive Board of our Association. To Viggo Hansen, our gratitude for his years of service as secretary, conference organizer, and member of the Executive Board of our Association.

One of the most significant milestones this year was the John Hutchins Machine Translation Archive new domain, an achievement built upon the hard work of our former president, Mikel Forcada, and a group of dedicated members, Barry Haddow, Leopoldo Pla, and Matt Post. The John Hutchins Machine Translation Archive is alive at: <https://mt-archive.net/>. We invite our community to visit John's archive!

A few lines more of gratitude to Matt Post, for being responsible for the import of the MT Archive conference proceedings into the ACL anthology. Our community is very thankful to Matt Post for the massive import work and patience along the way!

Now our EAMT 2022 event! After an online-edition in Lisbon, in 2020 (in which I had no opportunity to welcome you in person as a co-chair) and a cancellation in 2021, we now move forward to a fully and much hoped for live event in Ghent, Belgium! Winds of change in the pandemics are bringing a new hope. Embedded in this spirit, the local organizers are enthusiastic about hosting an in-person event, after the two-years interregnum, anticipating a much needed gathering of the community. Let us hope that these changes are here to stay.

Despite the positive changes in terms of covid, our community reached out to us requesting for support, specifically for freelance translators and/or members from low-income areas and war zones. For the first time, we have opened two calls for grants, encompassing, on one hand, students from Translation Studies and, on another, members from Middle East and African countries. A total of seven grants were given. We hope this initiative may mitigate the hard times we are living and may bring richer discussions into our EAMT 2022, diversifying geographically

our membership.

EAMT 2022 will have a three-day, four-track programme put together by our chairs: Lööc Barrault and Marta Costa-Jussà, research track co-chairs; Ellie Kemp and Spyridon Pilos, user track co-chairs; Maarit Koponen and Christophe Declercq, translator track co-chairs; and Mikel Forcada, as the projects/products track chair. Carolina Scarton, Secretary of EAMT, was the chair of the Best Thesis Award and also the technical coordinator of the reviewing process (our gratitude to Carolina who is always willing to support our community).

This year, the programme will also include two keynote speakers, invited by Lieve Macken and Andrew Rufener (with our full support and enthusiasm), Laura Rossi (Medtronic) and Jörg Tiedemann (University of Helsinki), combining industry and academia visions on the field, a true honour to have them and to be able to discuss their talks in person.

EAMT 2022 brings a new breeze of hope and it is the result of the hard work of our local organizers from the Language and Translation Technology Team (LT3) of Ghent University and CrossLang. Our gratitude and appreciation to the LT3 team, Lieve Macken (co-chair), Joke Daems, Arda Tezcan, and Bram Vanroy; and to the CrossLang team, Andrew Rufener (co-chair), Joachim Van den Bogaert, and especially to Martine Massiera for her outstanding work taking care of our sponsors, registration process, all social events, and smoothly handling the logistics of the new calls for grants.

EAMT has been supported by generous sponsors in its initiatives along the years. This year is no exception. Our gratitude to our sponsors: Microsoft (platinum sponsor, who will also be giving a talk entitled “Microsoft and Translators’ quest to break down language barriers”), Pangeanic and Yamagata (silver sponsors), STAR Group, Unbabel and Welocalize (bronze sponsors), Apertium (collaborator sponsor), Springer (best paper award supporter) and MultiLingual (media sponsor).

A final note to our participants! By the time I am writing these lines, there are already 113 participants and the number continues growing! Even in unstable times, this number is a very positive sign! We can finally meet in person! Let us take this opportunity to revive fruitful discussions, scientific collaborations, and constructive feedback in our community. I’m looking forward to seeing you all, finally!

Lisboa, 2022

**Helena Moniz**

President of the EAMT

General Chair of EAMT 2022

University of Lisbon / INESC-ID, Portugal



# Message from the Organising Committee Chairs

It is with great pleasure that we finally welcome you in Ghent to attend the 23rd Annual Conference of the European Association for Machine Translation.

The idea of organizing the 23rd Annual Conference of EAMT in Ghent jointly by the LT3 research team and CrossLang emerged in 2019 at the Croke Park stadium in Dublin where we enjoyed the Gala Dinner of the MT Summit. Unfortunately, COVID-19 prevented us from organizing an on-site event and EAMT2021 was first postponed and eventually cancelled. We are therefore extremely pleased to be able to welcome you all in our beautiful city for EAMT2022 and to meet you all in person.

A lot can happen in two years. The venue we had originally booked, the Aula Academica of Ghent University, a historic building of 1826, was no longer available due to renovation work. As Ghent is a great mix of old and new, we instead welcome you in the trendy Zebrastraat and keep the historic part for the conference dinner, which will be organized in the church of Monasterium PoortAckere.

People also change jobs. With a new team (and a new EAMT president) we continued the preparations for the conference. We kept the basic format of previous editions, but added a second keynote speaker. This not only allowed us to find the optimal balance between academia and industry but also ensured gender balance. We are really looking forward to the talks of Jörg Tiedemann and Laura Rossi.

We did not opt for a hybrid conference as the advantages did not outweigh the disadvantages. As a compromise, we will record the oral sessions and make the recordings available after the conference.

We would express our sincerest gratitude to everyone who made EAMT2022 possible: Mikel Forcada as former president of EAMT; Helena Moniz as new president of EAMT; Carolina Scarton as EAMT Secretary; L  ic Barrault and Marta Costa-Juss   as research track chairs; Ellie Kemp and Spyridon Pilos as user track co-chairs; Maarit Koponen and Christophe Declercq as translator track chairs.

We extend our thanks to our sponsors for their invaluable support: Microsoft (Platinum sponsor), Pangeanic and Yamagata (Silver sponsors), STAR Group, Unbabel and Welocalize (Bronze sponsors), Apertium (Collaborator sponsor), Springer (Supporter sponsor), and Multi-Lingual (media sponsor).

This conference would not have been possible without the hard work of all members of the

joint organizing team: Andrew Rufener, Martine Massiera and Joachim Van den Bogaert of CrossLang; Lieve Macken, Joke Daems, Arda Tezcan, Bram Vanroy and Margot Fonteyne of the Language and Translation Technology Team (LT<sup>3</sup>) of Ghent University. Sincere thanks as well to Sam Delmotte of Ghent University for recording the oral sessions.

**Lieve Macken**  
Ghent University, LT<sup>3</sup>

**Andrew Rufener**  
CrossLang

On behalf of the local organizers

# Preface by the Programme Chairs

On behalf of the programme chairs, a warm welcome to the 23rd annual conference of the European Association for Machine Translation in Ghent, Belgium. After all the restrictions, rescheduling and cancellations of events in the past couple of years, and after a prolonged period with almost all meetings online, we are delighted to finally be meeting our colleagues face-to-face again!

Following the approach which has proven so successful in the previous editions of EAMT, the conference programme consists of papers and posters divided into four tracks. These relate to research, users, translators and projects/products.

The research track this year was one of the most competitive tracks ever in the history of EAMT. Only 17 out of 39 papers were accepted (an acceptance rate of 44%), based on three-peer reviews. The papers describe state-of-the-art work being conducted and, therefore, are highly relevant to our community. Eight papers will be presented orally and nine as posters, as you may already find in the programme. We invite our community to reach out to the authors and discuss the relevant work conducted in such a demanding track.

The submissions for the user track in this edition mostly tackle the customer support domain - a particular focus of the oral sessions of the programme - and industry usage of MT. This track will discuss a number of practical issues for users. These range from the notion of “users” in a very challenging domain, to conversational data with strict time constraints, and the quality of the MT produced.

The translator’s track, as is evident from the name, emphasises the perspective of translators on MT. This year, the track features three peer-reviewed papers, each of which addresses aspects of machine translation and post-editing carried out by translators in different settings. The diverging uses of post-editing and machine translation cover a survey of corporate use of post-editing and revision in the NMT era, post-editing practices for automatically generated subtitles, and annotation of post-editing and machine translation errors using speech-to-text technology.

Forty-four papers were submitted to the largest ever project/product track in the history of EAMT conferences. Of them, 41 were eventually accepted, some of them after an additional round of improvements with the general audience of EAMT in sight. As these lines are written, authors are preparing their posters, and also their poster booster slides, in anticipation of their (strictly-timed) two minutes of glory before the poster session.

In addition to the papers and posters relating to the different tracks, the programme also features two fascinating invited talks: Laura Rossi with her talk titled “I once said to my boss

‘SMT will never work...’ ” and Jörg Tiedemann with “Democratizing machine translation with OPUS-MT”.

We wish to thank the members of the scientific programme committee for their time and support, and for their invaluable expertise in peer-reviewing the submissions. Our thanks naturally go to all the authors, without whom the programme would not exist, the local organisers for all their hard work, as well as Carol Scarton, Helena Moniz and Mikel Forcada for their unfailing advice and support.

**Löic Barrault**  
META AI Research

**Marta Costa-jussà**  
META AI Research

**Ellie Kemp**  
CLEAR Global

**Spyridon Pilos**  
European Court of Auditors

**Christophe Declercq**  
Univ. of Utrecht & Univ. College London

**Maarit Koponen**  
University of Eastern Finland

**Mikel Forcada**  
Universitat d’Alacant

# EAMT 2022 Committees

## General Chair

Helena Moniz, University of Lisbon, INESC-ID

## Programme Chairs

### Research track

Loïc Barrault, META AI Research  
Marta R. Costa-jussà, META AI Research

### User track

Ellie Kemp, CLEAR Global  
Spyridon Pilos, European Court of Auditors

### Translators' track

Christophe Declercq, Univ. of Utrecht & Univ. College London  
Maarit Koponen, Univ. of Eastern Finland

### Project/Product track

Mikel L. Forcada, Univ. d'Alacant

### Thesis award

Carolina Scarton, Univ. of Sheffield

## Organising committee

Lieve Macken, LT3, Ghent University (co-chair)  
Andrew Rufener, CrossLang (co-chair)  
Joachim Van den Bogaert, CrossLang  
Joke Daems, LT3, Ghent University  
Arda Tezcan, LT3, Ghent University  
Bram Vanroy, LT3, Ghent University  
Margot Fonteyne, LT3, Ghent University

## Programme Committee

### Research track

Duygu Ataman, New York University  
Parnia Bahar, RWTH Aachen University  
Anabela Barreiro, INESC-ID  
Luisa Bentivogli, Fondazione Bruno Kessler  
Magdalena Biesialska, Universitat Politècnica de Catalunya  
José G. C. de Souza, Unbabel  
Michael Carl, Kent State University  
Sheila Castilho, Dublin City University  
Mauro Cettolo, FBK - Fondazione Bruno Kessler  
Colin Cherry, Google Research  
Vishal Chowdhary, Microsoft  
Chenhui Chu, Kyoto University  
Raj Dabre, IIT Bombay  
Mattia Antonino Di Gangi, AppTek  
Miguel Domingo, Universitat Politècnica de València  
Miquel Espla-Gomis, Universitat d'Alacant  
Catarina Farinha, Unbabel  
Mireia Farrús, Universitat de Barcelona  
George Foster, Google  
Mercedes García-Martínez, Pangeanic SL  
Jesús González-Rubio, WebInterpret  
Barry Haddow, The University of Edinburgh  
Felix Hieber, Amazon  
Matthias Huck, SAP SE  
Shankar Kumar, Google  
Ekaterina Lapshinova-Koltunski, Saarland University  
Samuel Läubli, Zurich University of Applied Sciences  
Lieve Macken, Ghent University  
Andreas Maletti, Universität Leipzig  
Joss Moorkens, Dublin City University  
Mathias Müller, University of Zurich  
Masaaki Nagata, NTT  
Toshiaki Nakazawa, The University of Tokyo  
Jan Niehues, Maastricht University  
André Niyongabo, Polytechnic University of Catalonia  
Constantin Orasan, University of Surrey  
Pavel Pecina, Charles University  
Stephan Peitz, Apple  
Sergio Penkale, Unbabel  
Andrei Popescu-Belis, HEIG-VD / HES-SO  
Maja Popovic, ADAPT Centre @ DCU  
Celia Rico, Universidad Complutense de Madrid, Spain  
Matïss Rikters, Tilde  
Rudolf Rosa, Charles University  
Felipe Sánchez-Martínez, Universitat d'Alacant  
Germán Sanchis-Trilles, Sciling S.L.

Yves Scherrer, University of Helsinki  
Rico Sennrich, University of Zurich  
Patrick Simianer, Lilt, Inc.  
Katsuhito Sudoh, Nara Institute of Science and Technology  
Aleš Tamchyna, Memsource a. s.  
Jörg Tiedemann, University of Helsinki  
Antonio Toral, University of Groningen  
Marco Turchi, Fondazione Bruno Kessler  
Vincent Vandeghinste, Instituut voor de Nederlandse Taal, Leiden & Centre for Computational Linguistics, KU Leuven  
Sebastian Vincent, The University of Sheffield  
Marion Weller-Di Marco, CIS - University of Munich  
Deyi Xiong, Tianjin University  
François Yvon, CNRS  
Jiajun Zhang, Institute of Automation Chinese Academy of Sciences

## User track

Nora Aranberri, University of the Basque Country

Adam Bittlingmayer, ModelFront

Vera Cabarrão, Unbabel

Federico Gaspari, Università per Stranieri “Dante Alighieri” di Reggio Calabria, Italy

Georg Kirchner, Dell Technologies

László János Laki, Hungarian Research Centre for Linguistics

Lena Marg, Welocalize

Mary Nurminen, Tampere University

Morgan O’Brien, Momentive AI

Niko Papula, Multilizer

Daniel Prou, European Commission

Steve Richardson, Brigham Young University

Laura Rossi, Medtronic

Marina Sánchez-Torrón, Unbabel

Yury Sharshov, TBSJ

Dimitar Shterionov, Tilburg University

Masaru Yamada, Rikkyo University



## **Translators' track**

Khetam Al Sharou, Imperial College London

Dorothee Behr, GESIS - Leibniz Institute for the Social Sciences

Patrick Cadwell, Dublin City University

Ruben de La Fuente, PayPal

Gökhan Dogru, Universitat Autònoma de Barcelona

Maria Fernandez-Parra, Swansea University

Dorothy Kenny, Dublin City University

Caroline Lehr, Zurich University of Applied Sciences

Rudy Loock, Université de Lille, France, & CNRS "Savoirs, Textes, Langage" research unit

Antoni Oliver, Universitat Oberta de Catalunya

David Orrego-Carmona, Aston University

Alessandra Rossetti, Universiteit Antwerpen

Caroline Rossi, Université Grenoble Alpes

María Del Mar Sánchez Ramos, Universidad de Alcalá

Vilemini Sosoni, Ionian University

## **Thesis award**

José G. C. de Souza, Unbabel  
Vera Cabarrao, Unbabel Lda.; INESC-ID  
Sheila Castilho, Dublin City University  
Mirella De Sisto, Tilburg University  
Mattia Antonino Di Gangi, AppTek  
Miquel Esplà-Gomis, Universitat d'Alacant  
Federico Gaspari, Dublin City University  
Barry Haddow, University of Edinburgh  
Ekaterina Lapshinova-Koltunski, Saarland University  
Lieve Macken, Ghent University  
Andre Martins, Unbabel  
Maja Popovic, ADAPT Centre @ DCU  
Celia Rico, Universidad Complutense de Madrid  
Víctor M. Sánchez-Cartagena, Universitat d'Alacant  
Marina Sánchez-Torrón, Unbabel  
Arda Tezcan, Ghent University  
Marco Turchi, Fondazione Bruno Kessler

# Invited Speeches

## Democratizing machine translation with OPUS-MT

**Jörg Tiedemann**, University of Helsinki, Finland

The demand for translation is ever growing and this trend will not stop. Being able to access the same kind of information is a fundamental prerequisite for equality in society and translation plays a crucial role when fighting discrimination based on language barriers. Efficient tools and a better coverage of the linguistic diversity in the World are necessary to cope with the amount of material that needs to be handled. Our mission is to support the development of high quality tools for automatic and computer-assisted translation by providing open services and resources that are independent of commercial interests and profit-driven companies. Equal information access is a human right and not only a privilege for people who can pay for it. In this talk I will discuss the current state of OPUS-MT, our project on open neural machine translation and the challenges that we try to tackle with multilingual NLP, transfer learning and data augmentation. I will report about on-going work on knowledge distillation, the creation of compact models for real-time translation and our work on modularization of neural MT.

## **“I once said to my boss ‘SMT will never work...’ ”**

**Laura Rossi**, Medtronic

I once said to my boss: ‘SMT will never work...’, yet here we are: after being statistical, MT became neural and even adaptive, and achieved levels of quality that were unthinkable 20 years ago, covering, in addition, more and more language pairs every day. Customizations of MT systems have turned into a commodity, made available through specialized companies, LSPs and even as a self-service model. MT is very well integrated in human translation workflows to lower prices and shorten turnarounds. So, what are users, and in particular corporate users, looking for next? What creates a differentiative and appealing offer? What makes them choose for one or the other vendor? The race is moving towards automation, integration, well-being and sustainability.

