

# Contents

<b>Foreword from the General Chair</b> . . . . .	<b>v</b>
<b>Message from the Organising Committee Chair</b> . . . . .	<b>ix</b>
<b>Preface by the Programme Chairs</b> . . . . .	<b>xi</b>
<b>EAMT 2023 Organising Committees</b> . . . . .	<b>xv</b>
<b>Keynote Addresses</b> . . . . .	<b>1</b>
<b>EAMT 2023 Best Thesis Award — Anthony C Clarke Award</b> . . . . .	<b>3</b>
Biao Zhang. <i>Towards Efficient Universal Neural Machine Translation</i> . . . . .	5
<b>Research: Technical</b> . . . . .	<b>7</b>
Javad Pourmostafa Roshan Sharami, Dimitar Shterionov, Frédéric Blain, Eva. <i>Tailoring Domain Adaptation for Machine Translation Quality Estimation</i> . . . . .	9
Elise Bertin-Lemée, Annelies Braffort, Camille Challant, Claire Danet and Michael Filhol. <i>Example-Based Machine Translation from Text to a Hierarchical Representation of Sign Language</i> . . . . .	21
Mikko Aulamo, Ona de Gibert, Sami Virpioja and Jörg Tiedemann. <i>Unsupervised Feature Selection for Effective Parallel Corpus Filtering</i> . . . . .	31
Antoni Oliver González and Sergi Álvarez. <i>Filtering and rescoring the CCMatrix corpus for Neural Machine Translation training</i> . . . . .	39
Taisiya Glushkova, Chrysoula Zerva and André F. T. Martins. <i>BLEU Meets COMET: Combining Lexical and Neural Metrics Towards Robust Machine Translation Evaluation</i> . . . . .	47
Aarón Galiano-Jiménez, Felipe Sánchez-Martínez, Víctor M. Sánchez-Cartagena and Juan Antonio Pérez-Ortiz. <i>Exploiting large pre-trained models for low-resource neural machine translation</i> . . . . .	59
Nathaniel Berger, Miriam Exel, Matthias Huck and Stefan Riezler. <i>Enhancing Supervised Learning with Contrastive Markings in Neural Machine Translation Training</i> . . . . .	69
Francesco Fomicola, Silvia Bernardini, Federico Garcea, Adriano Ferraresi and Alberto Barrón-Cedeño. <i>Return to the Source: Assessing Machine Translation Suitability</i> . . . . .	79
Jianfei He, Shichao Sun, Xiaohua Jia and Wenjie Li. <i>Empirical Analysis of Beam Search Curse and Search Errors with Model Errors in Neural Machine Translation</i> . . . . .	91
Varun Gumma, Raj Dabre and Pratyush Kumar. <i>An Empirical Study of Leveraging Knowledge Distillation for Compressing Multilingual Neural Machine Translation Models</i> . . . . .	103
Pedro Henrique Martins, João Alves, Tânia Vaz, Madalena Gonçalves, Beatriz Silva, Marianna Buchicchio, José G. C. de Souza and André F. T. Martins. <i>Empirical Assessment of kNN-MT for Real-World Translation Scenarios</i> . . . . .	115

Shenbin Qian, Constantin Orasan, Felix Do Carmo, Qiuliang Li and Diptesh Kanojia. <i>Evaluation of Chinese-English Machine Translation of Emotion-Loaded Microblog Texts: A Human Annotated Dataset for the Quality Assessment of Emotion Translation</i> . . . . .	125
Benoist Wolleb, Romain Silvestri, Georgios Vernikos, Ljiljana Dolamic and Andrei Popescu-Belis. <i>Assessing the Importance of Frequency versus Compositionality for Subword-based Tokenization in NMT</i> . . . . .	137
Harritxu Gete, Thierry Etchegoyhen and Gorka Labaka. <i>What Works When in Context-aware Neural Machine Translation?</i> . . . . .	147
Rachel Bawden and François Yvon. <i>Investigating the Translation Performance of a Large Multilingual Language Model: the Case of BLOOM</i> . . . . .	157
Flavia De Camillis, Egon W. Stemle, Elena Chiochetti and Francesco Fericola. <i>The MT@BZ corpus: machine translation &amp; legal language</i> . . . . .	171
Sonal Sannigrahi and Rachel Bawden. <i>Investigating Lexical Sharing in Multilingual Machine Translation for Indian Languages</i> . . . . .	181
Tom Kocmi and Christian Federmann. <i>Large Language Models Are State-of-the-Art Evaluators of Translation Quality</i> . . . . .	193
Ali Vardasbi, Telmo Pessoa Pires, Robin Schmidt and Stephan Peitz. <i>State Spaces Aren't Enough: Machine Translation Needs Attention</i> . . . . .	205
Malina Chichirau, Rik van Noord and Antonio Toral. <i>Automatic Discrimination of Human and Neural Machine Translation in Multilingual Scenarios</i> . . . . .	217
Yasmin Moslem, Rejwanul Haque, John D. Kelleher and Andy Way. <i>Adaptive Machine Translation with Large Language Models</i> . . . . .	227
Angel Navarro, Miguel Domingo and Francisco Casacuberta. <i>Segment-based Interactive Machine Translation at a Character Level</i> . . . . .	239
<b>Research: Translators and users</b> . . . . .	<b>249</b>
Manuel Lardelli and Dagmar Gromann. <i>Gender-Fair Post-Editing: A Case Study Beyond the Binary</i> . . . . .	251
Miguel A. Jimenez-Crespo. <i>“Translationese” (and “post-editese”?) no more: on importing fuzzy conceptual tools from Translation Studies in MT research</i> . . . . .	261
María Do Campo Bayón and Pilar Sánchez-Gijón. <i>A social media NMT engine for a low-resource language combination</i> . . . . .	269
Hadeel Saadany, Constantin Orasan, Rocio Caro Quintana, Felix Do Carmo and Leonardo Zilio. <i>Analysing Mistranslation of Emotions in \\Multilingual Tweets by Online MT Tools</i> . . . . .	275
Janiča Hackenbuchner and Ralph Krüger. <i>DataLitMT – Teaching Data Literacy in the Context of Machine Translation Literacy</i> . . . . .	285
Salmi Leena, Aletta G. Dorst, Maarit Koponen and Katinka Zeven. <i>Do Humans Translate like Machines? Students’ Conceptualisations of Human and Machine Translation</i> . . . . .	295
Lieve Macken, Bram Vanroy and Arda Tezcan. <i>Adapting Machine Translation Education to the Neural Era: A Case Study of MT Quality Assessment</i> . . . . .	305
Sergi Alvarez and Antoni Oliver. <i>PE effort and neural-based automatic MT metrics: do they correlate?</i> . . . . .	315
Susana Valdez, Ana Guerberof Arenas and Kars Ligtenberg. <i>Migrant communities living in the Netherlands and their use of MT in healthcare settings</i> . . . . .	325

Vicent Briva-Iglesias and Sharon O'Brien. <i>Measuring Machine Translation User Experience (MTUX): A Comparison between AttrakDiff and User Experience Questionnaire</i> . . . . .	335
Fred Bane, Anna Zaretskaya, Tània Blanch Miró, Celia Soler Uguet and João Torres. <i>Coming to Terms with Glossary Enforcement: A Study of Three Approaches to Enforcing Terminology in NMT</i> . . . . .	345
Miguel Angel Rios Gaona, Raluca-Maria Chereji, Alina Secara and Dragos Ciobanu. <i>Quality Analysis of Multilingual Neural Machine Translation Systems and Reference Test Translations for the English-Romanian language pair in the Medical Domain</i> . . . . .	355
Maja Popovic, Ekaterina Lapshinova-Koltunski and Maarit Koponen. <i>Computational analysis of different translations: by professionals, students and machines</i> . . . .	365
Bettina Hiebl and Dagmar Gromann. <i>Quality in Human and Machine Translation: An Interdisciplinary Survey</i> . . . . .	375
Fadi Al-Ghawanmeh, Alexander Jensenius and Kamel Smaili. <i>How can machine translation help generate Arab melodic improvisation?</i> . . . . .	385
Sheila Castilho, Clodagh Quinn Mallon, Rahel Meister and Shengya Yue. <i>Do online Machine Translation Systems Care for Context? What About a GPT Model?</i> . .	393
<b>Implementations and case studies</b> . . . . .	<b>418</b>
Zeynep Yirmibeşoğlu, Olgun Dursun, Harun Dallı, Mehmet Şahin, Ena Hodzik, Sabri Gürses and Tunga Güngör. <i>Incorporating Human Translator Style into English-Turkish Literary Machine Translation</i> . . . . .	419
Konstantinos Chatzitheodorou, M. <sup>a</sup> Ángeles García Escrivá and Carmen Grau Lacal. <i>Machine translation of anonymized documents with human-in-the-loop</i> . . . . .	429
Marjolene Paulo, Vera Cabarrão, Helena Moniz, Miguel Menezes, Rachel Grewcock and Eduardo Farah. <i>Context-aware and gender-neutral Translation Memories</i> . .	437
Jose-Manuel De-la-Torre-Vilariño, Juan-Luis García-Mendoza and Alessia Petrucci. <i>Improving Machine Translation in the E-commerce Luxury Space. A case study</i> .	445
Mariana Cabeça, Marianna Buchicchio, Madalena Gonçalves, Christine Maroti, João Godinho, Pedro Coelho, Helena Moniz and Alon Lavie. <i>Quality Fit for Purpose: Building Business Critical Errors Test Suites</i> . . . . .	451
Maja Popovic, Vasudevan Nedumpozhimana, Meegan Gower, Sneha Rautmare, Nishtha Jain and John Kelleher. <i>Using MT for multilingual covid-19 case load prediction from social media texts</i> . . . . .	461
Matthias Wirth, Volker D. Hähnke, Franco Mascia, Arnaud Wéry, Konrad Vowinckel, Marco del Rey, Raúl Mohedano del Pozo, Pau Montes and Alexander Klenner-Bajaja. <i>Building Machine Translation Tools for Patent Language: A Data Generation Strategy at the European Patent Office</i> . . . . .	471
Rebecca Knowles, Samuel Larkin, Marc Tessier and Michel Simard. <i>Terminology in Neural Machine Translation: A Case Study of the Canadian Hansard</i> . . . . .	481
<b>Products and projects</b> . . . . .	<b>489</b>
Paola Ruffo, Joke Daems and Lieve Macken. <i>Developing User-centred Approaches to Technological Innovation in Literary Translation (DUAL-T)</i> . . . . .	491
Marie-Aude Lefer, Romane Bodart, Adam Obrusnik and Justine Piette. <i>The Post-Edit Me! project</i> . . . . .	493
Antoni Oliver, Mercè Vázquez, Marta Coll-Florit, Sergi Álvarez, Víctor Suárez, Claudi Aventín-Boya, Cristina Valdés, Mar Font and Alejandro Pardos. <i>TAN-IBE: Neural Machine Translation for the romance languages of the Iberian Peninsula</i> . . .	495

Cristina Toledo Báez. <i>GAMETRAPP: Training app for post-editing neural machine translation using gamification in professional settings</i> . . . . .	497
Bram Vanroy, Arda Tezcan and Lieve Macken. <i>MATEO: MACHine Translation Evaluation Online</i> . . . . .	499
Vincent Vandeghinste, Dimitar Shterionov, Mirella De Sisto, Aoife Brady, Mathieu De Coster, Lorraine Leeson, Josep Blat, Frankie Picron, Marcello Paolo Scipioni, Aditya Parikh, Louis ten Bosch, John O’Flaherty, Joni Dambre, Jorn Rijckaert, Bram Vanroy, Victor Ubieto Nogales, Santiago Egea Gomez, Ineke Schuurman, Gorka Labaka, Adrián Núñez-Marcos, Irene Murtagh, Euan McGill and Horacio Saggion. <i>SignON: Sign Language Translation. Progress and challenges.</i> . . . . .	501
Mirella De Sisto, Vincent Vandeghinste, Lien Soetemans, Caro Brosens and Dimitar Shterionov. <i>GoSt-ParC-Sign: Gold Standard Parallel Corpus of Sign and spoken language</i> . . . . .	503
Marta Bañón, Mălina Chichirău, Miquel Esplà-Gomis, Mikel Forcada, Aarón Galiano-Jiménez, Taja Kuzman, Nikola Ljubešić, Rik van Noord, Leopoldo Pla Sempere, Gema Ramírez-Sánchez, Peter Rupnik, Vit Suchomel, Antonio Toral and Jaume Zaragoza-Bernabeu. <i>MaCoCu: Massive collection and curation of monolingual and bilingual data: focus on under-resourced languages</i> . . . . .	505
Mathias Müller, Sarah Ebling, Eleftherios Avramidis, Alessia Battisti, Michèle Berger, Richard Bowden, Annelies Braffort, Necati Cihan Camgoz, Cristina España-Bonet, Roman Grundkiewicz, Zifan Jiang, Oscar Koller, Amit Moryossef, Regula Perrollaz, Sabine Reinhard, Annette Rios Gonzales, Dimitar Shterionov, Sandra Sidler-Miserez, Katja Tissi and Davy Van Landuyt. <i>First WMT Shared Task on Sign Language Translation (WMT-SLT22)</i> . . . . .	507
Maarit Koponen, Mary Nurminen, Nina Havumetsä and Juha Lång. <i>DECA: Democratic epistemic capacities in the age of algorithms</i> . . . . .	509
Yves Scherrer, Olli Kuparinen and Aleksandra Miletic. <i>CorCoDial - Machine translation techniques for corpus-based computational dialectology</i> . . . . .	511
Julian Hamm and Judith Klein. <i>How STAR Transit NXT can help translators measure and increase their MT post-editing efficiency</i> . . . . .	513
Lucía Ormaechea, Pierrette Bouillon, Maximin Coavoux, Emmanuelle Esperança-Rodier, Johanna Gerlach, Jérôme Goulian, Benjamin Lecouteux, Cécile Macaire, Jonathan Mutal, Magali Norré, Adrien Pupier and Didier Schwab. <i>PROPICTO: Developing Speech-to-Pictograph Translation Systems to Enhance Communication Accessibility</i> . . . . .	515
Mikko Aulamo, Nikolay Bogoychev, Shaoxiong Ji, Graeme Nail, Gema Ramírez-Sánchez, Jörg Tiedemann, Jelmer van der Linde and Jaume Zaragoza. <i>HPLT: High Performance Language Technologies</i> . . . . .	517
<b>Sponsors</b> . . . . .	<b>521</b>

# Foreword from the General Chair

As president of the European Association for Machine Translation (EAMT) and General Chair of the 24th Annual Conference of the EAMT, it is with great pleasure that I write these opening words to the Proceedings of EAMT 2023.

According to tradition, my first note of deep appreciation and gratitude goes to Heidi Depraetare and Khalil Simaan, Executive Board Members, who have moved to new adventures in their lives, after long, outstanding, and dedicated service to the EAMT community.

We have several milestones to celebrate this year, built upon the hard work of our Executive Committee and our community: upgraded grants for low income and war zones and for Translation Studies, a record submission rate for research projects (9 projects), a record for submissions for the best thesis candidates, and one of the highest number of papers ever submitted to our conference! I could not be prouder of the contagious energy from our community.

The EAMT Executive Committee (EC) has been very busy. Luc Meertens (treasurer) and Carolina Scarton (secretary) have been tirelessly supporting all initiatives. André Martins and Celia Rico, our co-chairs for low income areas, war zones and Translation Studies grants, selected 10 grantees. Barry Haddow and Carolina Scarton, our co-chairs for the Research Projects, selected 4 projects with a diverse set of topics. To all our co-chairs, my gratitude! The selection work is never an easy task and this year was particularly hard. The same applied to the best thesis award – co-chairs Carolina Scarton and Helena Moniz had a very difficult time selecting a single candidate, since the submissions were of very high quality.

EAMT, as full sponsor of the MT Marathon, would also like to highlight the outstanding work that the MT Marathon organisers conducted, enriching the vitality of our community with their projects and keynotes. A special thank you to Jindra Helcl, Ondřej Bojar, and Barry Haddow for all the efforts on yet another successful MT Marathon event.

EAMT, in an effort to reach out to our community in Africa, also sponsored three student grants to attend the AfricaNLP workshop at ICLR'23. Thank you, André Martins, for bridging our association with this initiative.

Now to Tampere, Finland! EAMT 2023 will have a three-day, four-track programme put together by our chairs: Eva Vanmassenhove and Tharindu Ranasinghe (research: technical track co-chairs); Nora Aranberri and Sergi Alvarez Vidal (research: translators & users track co-chairs); Carla Parra Escartín and Mara Nunziatini (implementations & case studies track co-chairs); and Mikel Forcada and Helena Moniz (products & projects track chairs). And backing up all the scientific components of our conference and filters of quality for the final selection: our reviewers. Thank you for your work and for the alignment with all the chairs!

This year EAMT 2023 will also have an extra day for workshops and tutorials, organised by our co-chairs Judith Brenner and Maja Popovic. Once more, the submissions for workshops and tutorials largely exceeded our expectations for this inaugural year!

The programme will continue the tradition of including two keynote speakers, Lynne Bowker (Full Professor at the University of Ottawa, Canada) and Marco Turchi (Head of MT at Zoom Video Communications). Our keynote speakers will demonstrate their extensive and impactful work in Translation Studies, technologies and machine translation, speech translation, and automatic post-editing. We bring you a fresh overview of the field, integrating a wide range of topics.

EAMT 2023 will also include a panel on *The Impact of Large Language Models (LLMs) on MT: A European View*, with several guests: Andreas Eisele (Responsible for MT at the European Union), André Martins (Unbabel/University of Lisbon), Christian Federmann (Microsoft), Helena Moniz (EAMT/University of Lisbon), Kenneth Church (Northeastern University), and Mikel Forcada (EAMT/Universitat d’Alacant). This panel is a moment to have a European view on a subject dominated by non-European initiatives.

EAMT 2023 would never be possible without the bright, enthusiastic, and hard working local organising team! What a dream team! Whenever EAMT had a request for a possible new addition, the answer was usually “Why not?” I’m so grateful for being able to work with you! Starting with our chair, Mary Nurminen (Tampere University), Judith Brenner (University of Eastern Finland), Maarit Koponen (University of Eastern Finland), Sirkku Latomaa (Tampere University), Mikhail Mikhailov (Tampere University), and Frederike Schierl, (Tampere University). Thank you, Tampere University and the University of Eastern Finland, for all your support! Tampere University has been such an amazing and flexible host!

EAMT has been supported by generous sponsors in its initiatives along the years. This year is no exception. Our gratitude to our Silver sponsors: Pangeanic, Unbabel and ZOO Digital. To our Bronze sponsors: CrossLang, ModelFront, STAR, TransPerfect, and Welocalize. Also to Apertium, our long standing collaborator sponsor; Springer, our Supporter sponsor for the Best Paper award; and our Media sponsors, MultiLingual and Slator. Your support is vital in our efforts to give back to our community through grants and other initiatives.

A note still to all our EAMT members and our participants! Without you no effort would make sense! Let us take this opportunity to create scientific collaboration and give constructive feedback. To fully enjoy the conference, please check our Code of Conduct at <https://events.tuni.fi/eamt23/ethics/>. I'm looking forward to seeing you all!

It is EAMT's greatest wish to continue giving back to our community and to drive and be driven by our community's energy and enthusiasm. Reach out to us if you have new ideas or suggestions you would like to implement. We will try hard to accomplish it with you. Learn more about us at <https://eamt.org/>.

**Helena Moniz**

President of the EAMT

General Chair of EAMT 2023

University of Lisbon / INESC-ID, Portugal





# Message from the Organising Committee Chair

*Tervetuloa Tampereelle!*

The local organising committee welcomes you all to EAMT 2023! Thank you for choosing to join us, either in person or remotely, for 3 days of talks, posters, chats with colleagues, and evening activities. Plus an extra day for many of you to focus on a particular issue in a workshop or tutorial. Attendance at EAMT conferences continues to grow, and the highest number ever will participate in the Tampere conference.

After 2 conferences in the southern parts of Europe – Alacant in 2018 and (virtually) Lisbon in 2020 – EAMT moved northward to Ghent in 2022 and then farther up to Finland in 2023. Perhaps Tampere, which is on the same latitude as the southern parts of Greenland, will be the farthest north the conference will ever be held. We hope you enjoy our light nights!

A few new things will be introduced in this year’s conference. For the first time, we will have an extra **workshop and tutorial day** adjacent to the main conference. As first-timers, we were unsure about the number and types of proposals we would receive. We were delighted to receive a large number of proposals of very high quality, and it was difficult to make selection decisions. We hope that everyone will have a good experience with this addition to the conference!

A second change this year are the **conference tracks**, which have been reconfigured to show an updated view of happenings in the MT world. Whereas we previously had 1 track for research, we now have 2: one that focuses on technical research and another that focuses on academic research on translators and other types of MT users, a field that has been steadily growing. The track on implementations and case studies highlights cases of actual MT use (‘in the wild’). The fourth track puts focus on various ongoing products and projects in the MT sphere.

The final change is that we are trying out a **hybrid light** conference attendance option to include those who cannot make it to Tampere themselves. We look forward to your feedback on all of these innovations.

A conference like this does not just happen – it is the result of great efforts by a number of people, and we'd like to thank them. First is the EAMT organisation, and especially President Helena Moniz and Secretary Carol Scarton, who went to great lengths to support our efforts. It would not have been possible without you. Next we'd like to thank our program chairs, who managed the vital work of selecting the best proposals and papers for the conference: Sergi Alvarez Vidal, Nora Aranberri, Judith Brenner, Mikel Forcada, Helena Moniz, Mara Nunziatini, Carla Parra Escartín, Maja Popovic, Tharindu Ranasinghe and Eva Vanmassenhove.

We'd also like to thank our Silver sponsors, Pangeanic, Unbabel and ZOO Digital; Bronze sponsors CrossLang, ModelFront, STAR Group, TransPerfect and Welocalize; Collaborator sponsor Apertium; Supporter sponsor Springer, and Media sponsors MultiLingual and Slator. Your support of our conference and activities is greatly appreciated!

Thanks also go out to the Tampere University Congress Office, which made so, so much of our work easier.

Personally, I'd like to thank my colleagues on the local planning committee: Judith Brenner, Maarit Koponen, Sirkku Latomaa, Mikhail Mikhailov and Frederike Schierl. We have been a small but very effective, 6-person powerhouse of activity. Thank you for your enthusiasm, willingness to jump into new things, and professionalism. It has been a pleasure to serve with you.

We look forward to meeting you all and to your active participation in the conference! Let's continue to make EAMT a unique space for a diverse group of researchers, developers, practitioners, leaders, vendors, users, and translators to share experiences and ideas.

**Mary Nurminen**

Tampere University and the University of Eastern Finland

On behalf of the local organising committee

# Preface by the Programme Chairs

On behalf of the programme chairs, a warm welcome to the 24th annual conference of the European Association for Machine Translation in Tampere, Finland. 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. However, the year 2023 sees a change in the structuring of the conference tracks. This year, we are introducing two tracks for research papers: one for more technical papers on MT development and another for research focusing on various types of users of MT. In addition to the two research tracks, two other tracks showcase use cases and implementations as well as projects and products. For the first time, the programme also includes workshops and tutorials. And the programme would not be complete without the two keynote speeches by Lynne Bowker and Marco Turchi.

This year at EAMT, there is a notable change since the traditional research track has been transformed into two distinct research tracks: the Technical Track and the Translators and Users Track. The **Technical Research track** invited and received technical submissions on all aspects of machine translation and related areas, serving as a hub for cutting-edge research and technological advancements, covering topics such as neural machine translation, language models, quality estimation, and more. It garnered significant attention and proved to be the most popular track at EAMT 2023, receiving a total of 51 submissions from 24 countries. Among these, 22 papers were accepted, resulting in an acceptance rate of 43%. Seven of the accepted papers will be presented orally, while the remaining 15 will be presented as posters.

A considerable number of the accepted papers are centered around Neural Machine Translation (NMT) and its diverse facets. Noteworthy topics include, among others, real-word translation (Martins et al., 2023), knowledge distillation (Gumma et al., 2023), and multilingual NMT (Chichirau et al., 2023). Several papers delve into machine translation quality estimation (QE), with a specific focus on domain adaptation of QE (Sharami et al., 2023), evaluating large language models for QE (Kocmi and Federmann, 2023), and emotion translation QE (Qian et al., 2023). It was evident that leveraging large language models such as GPT and BLOOM in machine translation and related fields is a prevailing trend. Given the popularity of tools like ChatGPT, we anticipate this trend to persist in future conferences as well. Additionally, the EAMT 2023 technical research track features several papers dedicated to low-resource languages (Sannigrahi et al., 2023; Galiano-Jiménez et al., 2023).

Last but not least, we would like to sincerely thank all the reviewers who provided feedback and insightful comments for the submissions received. We hope you enjoy reading this year's contributions to the Technical Research Track.

This edition has witnessed the reshaping and renaming of the MT tracks involving users. For the first time, a research track has been assigned to showcase studies carried out from a user perspective (translators, language experts and citizens who avail of the technology without pertaining to the language industry) and properly acknowledge the value and quality of the research in this field of study. As such, it has been the focus of the **Translators and Users Research track** to gather the widest range of topics in order to highlight the breadth of the area, current efforts and concerns regarding the quality and use of the technology.

We would like to thank the response of the community, which has contributed with an extensive selection of research themes. Work spanning MT literacy, concrete use cases and guidelines for their evaluation, assessments of MT output that go beyond sentence-level precision and fluency, translation styles and editing effort were submitted to the track. We believe that their outcomes serve as feedback for MT development but also help to establish targets for researchers in this particular subfield.

Overall, 18 papers were submitted to the track, out of which 16 were accepted (an 89% acceptance rate). 4 papers will be presented orally while the remaining will be exhibited in a dedicated poster session.

The EAMT conference has always sought to be an inclusive venue where researchers, users and MT practitioners could meet, discuss and share knowledge and expertise around machine translation from all possible points of view. With the aim of encouraging more practitioners to share their day-to-day experiences and learn from real use cases of MT, this year a new track was created: the **Implementations & Case Studies track**. This track aims to allow those using MT in their organisations to share their experiences from different angles. The 8 papers that will be presented at the conference showcase the wide variety of topics that this may cover, from building domain-specific MT engines to using MT for epidemiological social media surveillance, among others. They all cover different domains, from e-commerce to patents, demonstrating how MT, now more than ever, is a ubiquitous technology used by very different organisations and for ever-expanding purposes. And while this happens, it still poses challenges along the way that need to be tackled in real-world settings to ensure MT implementations are successful.

De-la-Torre-Vilariño et al. (2023) focus on how to build a domain specific, high-quality MT workflow in the e-commerce luxury space, while Zeynep et al. (2023) experiment with how MT systems can be fine-tuned towards the specific stylistic features of literary translators for the translation of literature. Within the customer support domain, Cabeça et al. (2023) focus

on building test suites to monitor MT and QE systems, paying particular attention to those errors that are critical to customers. Paulo et al. (2023) propose ways to identify context-dependent translation units that require gender agreement, and explain how to minimise such context-dependency through manipulating the translation units to make them gender-neutral and hence minimise gender bias in their MT training data. Also in the area of data preparation for MT, Wirth et al (2023) describe the process used at the European Patent Office for generating MT data to train their patent-specific MT models, and the challenges that this task poses. Chatzitheodorou et al. (2023) tackle the challenge of reconciling the competing needs of data privacy and data quality through post-editing anonymised texts. Another common challenge in MT is how to successfully incorporate terminology and tackle the tradeoffs that this may imply. Knowles et al. (2023) address this challenge in their paper. Finally, the last paper in this track explores how MT can be used for document classification purposes: Popović et al. (2023) showcase how MT can be used for scaling epidemiological social media surveillance.

The **Products and Projects track** has been upgraded with clearer criteria for submission, based on the extensive experience gathered after years of running this track. This year we received 20 submissions and 16 papers were accepted. The selection will provide a plethora of products and projects being developed by our community with a rich set of topics. It will surely be a very lively session with the usual poster boosters (one of our EAMT conferences' favourite moments) and poster sessions.

For the first time, this year's EAMT conference includes a **Workshop and Tutorial Day**. We invited proposals for in-depth sessions on any aspect of machine translation and related fields, and a total of 7 workshop proposals and 4 tutorial proposals were received. Almost all of the submissions were for a full-day event, demonstrating the organisers' eagerness to take the audience on a deep dive into their respective areas of expertise. After a careful review process, taking into account all aspects of the submissions, 4 workshops and 1 tutorial were accepted. The workshop topics range from gender-inclusive translation technology, open-source MT tools and automated translation of sign and spoken languages to language generation, while the tutorial explores the evolving role of the post-editor with speakers from both academia and the industry.

Our special thanks go to our track advisor, Jay Marciano, whose extensive experience in organising and hosting MT-related conferences and events was a great source of inspiration and guidance in the implementation of the first Workshop and Tutorial Day at an EAMT conference.

We also wish to thank Karen Patteri de Souza from the University of Eastern Finland for invaluable help in putting together this proceedings volume.

Nora Aranberri  
University of the Basque Country (UPV/EHU)

Carla Parra Escartín  
RWS Language Weaver

Judith Brenner  
University of Eastern Finland

Mikel Forcada  
Universitat d'Alacant

Maarit Koponen  
University of Eastern Finland

Maja Popovic  
Dublin City University

Helena Moniz  
University of Lisbon, INESC-ID

Tharindu Ranasinghe  
University of Wolverhampton

Mara Nunziatini  
Welocalize

Sergi Alvarez Vidal  
Universitat Oberta de Catalunya

Eva Vanmassenhov  
Tilburg University

# EAMT 2023 Organising Committees

## General Chair

Helena Moniz, EAMT President, University of Lisbon, INESC-ID

## Programme Chairs

### Research: technical track

Tharindu Ranasinghe, University of Wolverhampton  
Eva Vanmassenhove, Tilburg University

### Research: translators & users track

Sergi Alvarez Vidal, Universitat Oberta de Catalunya  
Nora Aranberri, University of the Basque Country (UPV/EHU)

### Implementations & case studies track

Mara Nunziatini, Welocalize  
Carla Parra Escartín, RWS Language Weaver

### Products & projects track

Mikel Forcada, Universitat d'Alacant  
Helena Moniz, University of Lisbon, INESC-ID

### Workshops & tutorials

Judith Brenner, University of Eastern Finland  
Maja Popovic, DCU

## Organising committee

Mary Nurminen, Tampere University (chair)  
Judith Brenner, University of Eastern Finland  
Maarit Koponen, University of Eastern Finland  
Sirkku Latomaa, Tampere University  
Mikhail Mikhailov, Tampere University  
Frederike Schierl, Tampere University

## Programme Committee

### Research: technical track

Fernando Alva-Manchego, Cardiff University  
Mihael Arcan, Insight Centre for Data Analytics, National University of Ireland Galway  
Duygu Ataman, New York University  
Eleftherios Avramidis, German Research Center for Artificial Intelligence (DFKI)  
Parnia Bahar, AppTek  
Loïc Barrault, Meta AI  
Anabela Barreiro, INESC-ID  
Rachel Bawden, Inria  
Luisa Bentivogli, FBK  
Magdalena Biesialska, Universitat Politècnica de Catalunya  
Frederic Blain, Tilburg University  
Michael Carl, Kent State University  
Francisco Casacuberta, Universitat Politècnica de València  
Sheila Castilho, Dublin City University  
Mauro Cettolo, FBK - Fondazione Bruno Kessler  
Colin Cherry, Google  
Vishal Chowdhary, Microsoft  
Chenhui Chu, Kyoto University  
Raj Dabre, IIT Bombay  
Aswarth Abhilash Dara, Amazon Alexa AI  
Mirella De Sisto, Tilburg University  
Mattia Antonino Di Gangi, AppTek  
Miguel Domingo, Universitat Politècnica de València  
Christian Dugast, tech2biz  
Hiroshi Echizenya, Hokkai-Gakuen University  
Cristina España-Bonet, UdS and DFKI  
Miquel Esplà-Gomis, Universitat d'Alacant  
Marcello Federico, AWS AI Labs  
Christian Federmann, Microsoft  
Mark Fishel, University of Tartu  
Markus Freitag, Google AI  
Marco Gaido, Fondazione Bruno Kessler  
Mercedes García-Martínez, Uniphore  
Ulrich Germann, The University of Edinburgh  
Jesús González Rubio, WebInterpret  
Isao Goto, NHK  
Francisco Javier Guzman, Facebook  
Barry Haddow, The University of Edinburgh  
Rejwanul Haque, National College of Ireland  
Chris Hokamp, Aylien  
Matthias Huck, SAP SE  
Diptesh Kanojia, IIT Bombay  
Rebecca Knowles, National Research Council Canada  
Philipp Koehn, Johns Hopkins University  
Shankar Kumar, Google  
Maria Kunilovskaya, University of Saarland



Ekaterina Lapshinova-Koltunski, University of Hildesheim  
Samuel Läubli, Zurich University of Applied Sciences  
Gregor Leusch, eBay  
Andreas Maletti, Universität Leipzig  
Arul Menezes, Microsoft  
Antonio Valerio Miceli Barone, Univeristy of Edinburgh  
Helena Moniz, INESC  
Mathias Müller, University of Zurich  
Kenton Murray, Johns Hopkins University  
Maria Nadejde, Amazon  
Masaaki Nagata, NTT  
Toshiaki Nakazawa, The University of Tokyo  
Jan Niehues, Maastricht University  
André Niyongabo, Princeton University  
Constantin Orasan, University of Surrey  
Daniel Ortiz-Martínez, Universitat de Barcelona  
Pavel Pecina, Charles University  
Stephan Peitz, Apple  
Sergio Penkale, Unbabel  
Alberto Poncelas, Rakuten Institute of Technology  
Andrei Popescu-Belis, HEIG-VD / HES-SO  
Maja Popovic, ADAPT Centre @ DCU  
Preethi Raghavan, Fidelity  
Ayla Rigouts Terry, KU Leuven Kulak, Centre for Computational Linguistics  
Miguel Rios, University of Vienna  
Annette Rios Gonzales, University of Zurich  
Rudolf Rosa, Charles University  
Fatiha Sadat, UQAM  
V́ctor M. Sánchez-Cartagena, Universitat d'Alacant  
Felipe Sánchez-Martínez, Universitat d'Alacant  
Germán Sanchis-Trilles, Sciling S.L.  
Danielle Saunders, RWS Language Weaver  
Beatrice Savoldi, Fondazione Bruno Kessler  
Yves Scherrer, University of Helsinki  
Helmut Schmid, Ludwig Maximilian University of Munich  
Rico Sennrich, University of Zurich  
Dimitar Shterionov, Tilburg University  
Michel Simard, National Research Council Canada (NRC)  
Patrick Simianer, Lilt, Inc.  
Felix Stahlberg, Google Research  
Katsuhito Sudoh, Nara Institute of Science and Technology  
Aleš Tamchyna, Phrase a.s.  
Joël Tang, Imperial College London  
Arda Tezcan, Ghent University  
Jörg Tiedemann, University of Helsinki  
Antonio Toral, University of Groningen  
Masao Utiyama, NICT  
Vincent Vandeghinste, Instituut voor de Nederlandse Taal, Leiden // Centre for Computational Linguistics, KU Leuven

Dušan Variš, Institute of Formal and Applied Linguistics, Charles University in Prague  
David Vilar, Google  
Sebastian Vincent, The University of Sheffield  
Martin Volk, University of Zurich  
Trang Vu, Monash University  
Ekaterina Vylomova, The University of Melbourne  
Longyue Wang, Tencent AI Lab  
Taro Watanabe, Nara Institute of Science and Technology  
Marion Weller-Di Marco, CIS - University of Munich  
Minghao Wu, Monash University  
Yinfei Yang, Redfin Inc.  
François Yvon, CNRS  
Marcos Zampieri, George mason University  
Rabih Zbib, Avature  
Dakun Zhang, Systran SAS

## **Research: translators & users track**

Loubna Bilali, Kent State University  
Lynne Bowker, University of Ottawa  
Vicent Briva-Iglesias, SFI CRT D-REAL, Dublin City University  
Patrick Cadwell, Dublin City University  
Michael Carl, Kent State University  
João Lucas Cavalheiro Camargo, Dublin City University  
Dragos Ciobanu, University of Vienna  
Joke Daems, Ghent University  
Helle Dam Jensen, Aarhus University  
Alice Delorme Benites, Zurich University of Applied Sciences  
Lettie Dorst, Leiden University  
Emmanuelle Esperança-Rodier, LIG - GETALP - UGA  
Maria Fernandez-Parra, Swansea University  
Federico Gaspari, ADAPT Centre, Dublin City University  
Ana Guerberof Arenas, University of Groningen  
Junyan Jiang, New York University Shanghai  
Dorothy Kenny, Dublin City University  
Rudy Loock, Université de Lille, France, & CNRS "Savoirs, Textes, Langage" research unit  
Lieve Macken, Ghent University  
Marianna Martindale, University of Maryland  
Helena Moniz, INESC  
Joss Moorkens, Dublin City University  
Lucas N Vieira, University of Bristol  
Sharon O'Brien, Dublin City University  
Antoni Oliver, Universitat Oberta de Catalunya  
David Orrego-Carmona, University of Warwick  
John Ortega, Northeastern University  
Valentina Ragni, University of Warsaw  
Celia Rico, Universidad Complutense de Madrid  
Alessandra Rossetti, Vrije Universiteit Brussel  
María Del Mar Sánchez Ramos, Universidad de Alcalá  
Vilemini Sosoni, Ionian University  
Susana Valdez, Leiden University Centre for Linguistics

## **Implementations & case studies track**

Eleftherios Avramidis, German Research Center for Artificial Intelligence (DFKI)

Frederick Bane, Transperfect

Adam Bittlingmayer, ModelFront

Marianna Buchicchio, Unbabel

Laura Casanellas, Laura Casanellas

Miriam Exel, SAP SE

Mercedes García-Martínez, Uniphore

László János Laki, Hungarian Research Centre for Linguistics

Helena Moniz, INESC

Raj Patel, Huawei Ireland Research Centre

Spyridon Pilos, European Court of Auditors

Heather Rossi, RWS

Marina Sánchez-Torrón, Unbabel

Konstantin Savenkov, Intento, Inc.

Cecilia Yalangozian, Reviewer

Anna Zaretskaya, TransPerfect

## **Products & projects track**

Mikel Forcada, Universitat d'Alacant

Helena Moniz, University of Lisbon (FLUL), INESC-ID

## Thesis award

Rachel Bawden, Inria  
Daniel Beck, The University of Melbourne  
William Byrne, University of Cambridge  
José G. C. de Souza, Unbabel  
Vera Cabarrão, Unbabel / INESC-ID  
Sheila Castilho, Dublin City University  
Anna Currey, Amazon Web Services  
Mattia Antonino Di Gangi, AppTek  
Maha Elbayad, LIG/ Inria  
Miquel Esplà-Gomis, Universitat d'Alacant  
Marcello Federico, Amazon AI  
Mikel Forcada, DLSI - Universitat d'Alacant  
Barry Haddow, The University of Edinburgh  
Diptesh Kanojia, IIT Bombay  
Philipp Koehn, Johns Hopkins University  
Helena Moniz, INESC/FLUL  
Mary Nurminen, Tampere University  
Constantin Orasan, University of Surrey  
John E. Ortega, Northeastern University  
Santanu Pal, Wipro Limited  
Pavel Pecina, Charles University  
Maja Popovic, ADAPT Centre @ DCU  
Celia Rico, Universidad Complutense de Madrid  
Víctor M. Sánchez-Cartagena, Universitat d'Alacant  
Marina Sánchez-Torrón, Unbabel  
Danielle Saunders, University of Cambridge  
Dimitar Shterionov, Tilburg University  
Felix Stahlberg, Google Research  
Arda Tezcan, Ghent University  
Antonio Toral, University of Groningen  
Ualsher Tukeyev, al-Farabi Kazakh National University  
Bram Vanroy, KU Leuven; Ghent University  
Longyue Wang, Tencent AI Lab

# Keynote Addresses

## Towards an Outward Turn in Translation Technology Research?

Lynne Bowker, University of Ottawa

In 2019, Susan Bassnett and David Johnson guest edited a special issue of *The Translator* (vol. 25, issue 3) with the theme “the Outward Turn”. In the introduction, the editors note that Translation Studies (TS) has witnessed numerous turns in the past decades (e.g. linguistic, cultural, sociological), and is perhaps not really in need of another, not the least because fields do not develop in a neat linear way. Nevertheless, Bassnett and Johnson point to what they see as a potentially worrying trend whereby TS scholars seem increasingly to talk mainly to one another, which puts TS at risk of lurching “into ultimate self-referentiality, especially in the global academic marketplace where reference and citation are perceived as valuable ends in themselves” (p. 185). Of course, the fields of translation technology and TS do not face precisely the same issues, nor will they necessarily benefit from the same specific approaches. Yet at a higher level, we might do well to pay attention to discussions about an Outward Turn in TS and consider how this could benefit the translation technology community. For instance, Bassnett and Johnson suggest that at one level, the idea of an Outward Turn entails the recognition of the need for an increasing plurality of voices from across the globe; yet, this must be coupled with a recognition of the importance of creating space where different traditions can maintain their perspective and assert the value of their own concerns and insights within the homogenizing context of internationalization. In other words, an Outward Turn in TS would see researchers focus on the issues that increasingly surround them and recognize that uniformity can ultimately be damaging for everyone. In what ways might the broad strokes of an Outward Turn be relevant for translation technology research? This presentation will consider how various aspects of this need for expanding horizons within and beyond the contours of the translation technology field could manifest themselves in our collective research agenda.

# Towards Real-time Meeting Translation

Marco Turchi, Zoom Video Communications

Nowadays, machine translation (MT) has become the prominent solution to break language barriers and is used daily to translate emails, chats, technical documents, news articles, etc. At Zoom, we provide users with translation solutions to allow them to better connect, collaborate, and communicate in different languages during meetings. However, different from the classic speech translation use cases including TED talks or European Parliament sessions, the meeting scenario poses several challenges for MT technology. For instance, when speaking spontaneously, people introduce hesitations and repetitions, and, when interacting with other participants, they generate truncated, overlapped, and malformed utterances. So, in addition to the speech recognition errors, the MT system needs to simultaneously deal with all these factors to generate the optimal translation in real time. In my presentation, I will initially focus on highlighting the main challenges of meeting translation, paying attention to those phenomena that have a critical impact on the final output. Then, I will present some solutions that can be used to mitigate these problems and enhance translation quality in meetings.



# EAMT 2023 Best Thesis Award — Anthony C Clarke Award

Nine PhD theses defended in 2022 were received as candidates for the 2022 edition of the EAMT Best Thesis Award, and all nine were eligible. 28 reviewers and 6 EAMT Executive Committee members were recruited to examine and score the theses, considering how challenging the problem tackled in each thesis was, how relevant the results were for machine translation as a field, and what the strength of its impact in terms of scientific publications was. Two EAMT Executive Committee members also analysed all theses. It became very clear that 2022 was another very good year for PhD theses in machine translation.

All theses had merit, all candidates had strong CVs and, therefore, it was very difficult to select a winner.

A panel of two EAMT Executive Committee members (Carolina Scarton and Helena Moniz) was assembled to process the reviews and select a winner that was later ratified by the EAMT executive committee.

We are pleased to announce that the awardee of the 2022 edition of the EAMT Best Thesis is **Biao Zhang's thesis "Towards Efficient Universal Neural Machine Translation"** (University of Edinburgh, UK), supervised by Dr Rico Sennrich and Dr Ivan Titov.

The awardee will receive a prize of €500, together with a suitably-inscribed certificate. In addition, Dr. Zhang will present a summary of their thesis at the 24rd Annual Conference of the European Association for Machine Translation (EAMT 2023: <https://events.tuni.fi/eamt23/>) which will take place from June 12th to 15th in Tampere, Finland. In order to facilitate this, the EAMT will waive the winner's registration costs and will make available a travel bursary of €200.

Helena Moniz, EAMT President  
Carolina Scarton, EAMT Secretary

