# 24th Annual Conference of the European Association for Machine Translation (EAMT 2023)

Tampere, Finland
12-15 June 2023

**Editors:**

**Mary Nurminen**
**Judith Brenner**
**Maarit Koponen**
**Sirkku Latomaa**
**Mikhail Mikhailov**
**Frederike Schierl**
**Tharindu Ranasinghe**
**Eva Vanmassenhove**
**Sergi Alvarez Vidal**
**Nora Aranberri**
**Mara Nunziatini**
**Carla Parra Escartín**
**Mikel Forcada**
**Maja Popovic**
**Carolina Scarton**
**Helena Moniz**

ISBN: 978-1-7138-8993-9

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

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

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

Copyright© (2023) by the authors.
All rights reserved.

These articles are licensed under a Creative Commons Attribution 4.0 International License
(CC BY 4.0).

Printed with permission by Curran Associates, Inc. (2024)

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

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

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

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

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

# Contents

**Foreword from the General Chair** . . . . . **v**

**Message from the Organising Committee Chair** . . . . . **ix**

**Preface by the Programme Chairs** . . . . . **xi**

**EAMT 2023 Organising Committees** . . . . . **xv**

**Keynote Addresses** . . . . . **1**

**EAMT 2023 Best Thesis Award — Anthony C Clarke Award** . . . . . **3**

Biao Zhang. *Towards Efficient Universal Neural Machine Translation* . . . . . 5

**Research: Technical** . . . . . **7**

Javad Pourmostafa Roshan Sharami, Dimitar Shterionov, Frédéric Blain, Eva. *Tailoring Domain Adaptation for Machine Translation Quality Estimation* . . . . . 9

Elise Bertin-Lemée, Annelies Braffort, Camille Challant, Claire Danet and Michael Filhol. *Example-Based Machine Translation from Textto a Hierarchical Representation of Sign Language* . . . . . 21

Mikko Aulamo, Ona de Gibert, Sami Virpioja and Jörg Tiedemann. *Unsupervised Feature Selection for Effective Parallel Corpus Filtering* . . . . . 31

Antoni Oliver González and Sergi Álvarez. *Filtering and rescoring the CCMatrix corpus for Neural Machine Translation training* . . . . . 39

Taisiya Glushkova, Chrysoula Zerva and André F. T. Martins. *BLEU Meets COMET: Combining Lexical and Neural Metrics Towards Robust Machine Translation Evaluation* . . . . . 47

Aarón Galiano-Jiménez, Felipe Sánchez-Martínez, Víctor M. Sánchez-Cartagena and Juan Antonio Pérez-Ortiz. *Exploiting large pre-trained models for low-resource neural machine translation* . . . . . 59

Nathaniel Berger, Miriam Exel, Matthias Huck and Stefan Riezler. *Enhancing Supervised Learning with Contrastive Markings in Neural Machine Translation Training* 69

Francesco Fernicola, Silvia Bernardini, Federico Garcea, Adriano Ferraresi and Alberto Barrón-Cedeño. *Return to the Source: Assessing Machine Translation Suitability* 79

Jianfei He, Shichao Sun, Xiaohua Jia and Wenjie Li. *Empirical Analysis of Beam Search Curse and Search Errors with Model Errors in Neural Machine Translation* 91

Varun Gumma, Raj Dabre and Pratyush Kumar. *An Empirical Study of Leveraging Knowledge Distillation for Compressing Multilingual Neural Machine Translation Models* . . . . . 103

Pedro Henrique Martins, João Alves, Tânia Vaz, Madalena Gonçalves, Beatriz Silva, Marianna Buchicchio, José G. C. de Souza and André F. T. Martins. *Empirical Assessment of kNN-MT for Real-World Translation Scenarios* . . . . . 115

Shenbin Qian, Constantin Orasan, Felix Do Carmo, Qiuliang Li and Diptesh Kanojia. *Evaluation of Chinese-English Machine Translation of Emotion-Loaded Microblog Texts: A Human Annotated Dataset for the Quality Assessment of Emotion Translation* . . . 125
Benoist Wolleb, Romain Silvestri, Georgios Vernikos, Ljiljana Dolamic and Andrei Popescu-Belis. *Assessing the Importance of Frequency versus Compositionality for Subword-based Tokenization in NMT* . . . 137
Harritxu Gete, Thierry Etchegoyhen and Gorka Labaka. *What Works When in Context-aware Neural Machine Translation?* . . . 147
Rachel Bawden and François Yvon. *Investigating the Translation Performance of a Large Multilingual Language Model: the Case of BLOOM* . . . 157
Flavia De Camillis, Egon W. Stemle, Elena Chiocchetti and Francesco Fernicola. *The MT@BZ corpus: machine translation & legal language* . . . 171
Sonal Sannigrahi and Rachel Bawden. *Investigating Lexical Sharing in Multilingual Machine Translation for Indian Languages* . . . 181
Tom Kocmi and Christian Federmann. *Large Language Models Are State-of-the-Art Evaluators of Translation Quality* . . . 193
Ali Vardasbi, Telmo Pessoa Pires, Robin Schmidt and Stephan Peitz. *State Spaces Aren't Enough: Machine Translation Needs Attention* . . . 205
Malina Chichirau, Rik van Noord and Antonio Toral. *Automatic Discrimination of Human and Neural Machine Translation in Multilingual Scenarios* . . . 217
Yasmin Moslem, Rejwanul Haque, John D. Kelleher and Andy Way. *Adaptive Machine Translation with Large Language Models* . . . 227
Angel Navarro, Miguel Domingo and Francisco Casacuberta. *Segment-based Interactive Machine Translation at a Character Level* . . . 239

**Research: Translators and users** . . . **249**
Manuel Lardelli and Dagmar Gromann. *Gender-Fair Post-Editing: A Case Study Beyond the Binary* . . . 251
Miguel A. Jimenez-Crespo. *"Translationese" (and "post-editese"?) no more: on importing fuzzy conceptual tools from Translation Studies in MT research* . . . 261
María Do Campo Bayón and Pilar Sánchez-Gijón. *A social media NMT engine for a low-resource language combination* . . . 269
Hadeel Saadany, Constantin Orasan, Rocio Caro Quintana, Felix Do Carmo and Leonardo Zilio. *Analysing Mistranslation of Emotions in \\Multilingual Tweets by Online MT Tools* . . . 275
Janiça Hackenbuchner and Ralph Krüger. *DataLitMT – Teaching Data Literacy in the Context of Machine Translation Literacy* . . . 285
Salmi Leena, Aletta G. Dorst, Maarit Koponen and Katinka Zeven. *Do Humans Translate like Machines? Students' Conceptualisations of Human and Machine Translation* . . . 295
Lieve Macken, Bram Vanroy and Arda Tezcan. *Adapting Machine Translation Education to the Neural Era: A Case Study of MT Quality Assessment* . . . 305
Sergi Alvarez and Antoni Oliver. *PE effort and neural-based automatic MT metrics: do they correlate?* . . . 315
Susana Valdez, Ana Guerberof Arenas and Kars Ligtenberg. *Migrant communities living in the Netherlands and their use of MT in healthcare settings* . . . 325

Vicent Briva-Iglesias and Sharon O'Brien. *Measuring Machine Translation User Experience (MTUX): A Comparison between AttrakDiff and User Experience Questionnaire* . . . 335
Fred Bane, Anna Zaretskaya, Tània Blanch Miró, Celia Soler Uguet and João Torres. *Coming to Terms with Glossary Enforcement: A Study of Three Approaches to Enforcing Terminology in NMT* . . . 345
Miguel Angel Rios Gaona, Raluca-Maria Chereji, Alina Secara and Dragos Ciobanu. *Quality Analysis of Multilingual Neural Machine Translation Systems and Reference Test Translations for the English-Romanian language pair in the Medical Domain* . . . 355
Maja Popovic, Ekaterina Lapshinova-Koltunski and Maarit Koponen. *Computational analysis of different translations: by professionals, students and machines* . . . 365
Bettina Hiebl and Dagmar Gromann. *Quality in Human and Machine Translation: An Interdisciplinary Survey* . . . 375
Fadi Al-Ghawanmeh, Alexander Jensenius and Kamel Smaili. *How can machine translation help generate Arab melodic improvisation?* . . . 385
Sheila Castilho, Clodagh Quinn Mallon, Rahel Meister and Shengya Yue. *Do online Machine Translation Systems Care for Context? What About a GPT Model?* . . . 393

**Implementations and case studies** . . . **418**
Zeynep Yirmibeşoğlu, Olgun Dursun, Harun Dallı, Mehmet Şahin, Ena Hodzik, Sabri Gürses and Tunga Güngör. *Incorporating Human Translator Style into English-Turkish Literary Machine Translation* . . . 419
Konstantinos Chatzitheodorou, M.[a] Ángeles García Escrivá and Carmen Grau Lacal. *Machine translation of anonymized documents with human-in-the-loop* . . . 429
Marjolene Paulo, Vera Cabarrão, Helena Moniz, Miguel Menezes, Rachel Grewcock and Eduardo Farah. *Context-aware and gender-neutral Translation Memories* . . . 437
Jose-Manuel De-la-Torre-Vilariño, Juan-Luis García-Mendoza and Alessia Petrucci. *Improving Machine Translation in the E-commerce Luxury Space. A case study* . . . 445
Mariana Cabeça, Marianna Buchicchio, Madalena Gonçalves, Christine Maroti, João Godinho, Pedro Coelho, Helena Moniz and Alon Lavie. *Quality Fit for Purpose: Building Business Critical Errors Test Suites* . . . 451
Maja Popovic, Vasudevan Nedumpozhimana, Meegan Gower, Sneha Rautmare, Nishtha Jain and John Kelleher. *Using MT for multilingual covid-19 case load prediction from social media texts* . . . 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. *Building Machine Translation Tools for Patent Language: A Data Generation Strategy at the European Patent Office* . . . 471
Rebecca Knowles, Samuel Larkin, Marc Tessier and Michel Simard. *Terminology in Neural Machine Translation: A Case Study of the Canadian Hansard* . . . 481

**Products and projects** . . . **489**
Paola Ruffo, Joke Daems and Lieve Macken. *Developing User-centred Approaches to Technological Innovation in Literary Translation (DUAL-T)* . . . 491
Marie-Aude Lefer, Romane Bodart, Adam Obrusnik and Justine Piette. *The Post-Edit Me! project* . . . 493
Antoni Oliver, Mercè Vàzquez, Marta Coll-Florit, Sergi Álvarez, Víctor Suárez, Claudi Aventín-Boya, Cristina Valdés, Mar Font and Alejandro Pardos. *TAN-IBE: Neural Machine Translation for the romance languages of the Iberian Peninsula* . . . 495

Cristina Toledo Báez. *GAMETRAPP: Training app for post-editing neural machine trans-lation using gamification in professional settings* . . . 497
Bram Vanroy, Arda Tezcan and Lieve Macken. *MATEO: MAchine Translation Evaluation Online* . . . 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. *SignON: Sign Language Translation. Progress and challenges.* . . . 501
Mirella De Sisto, Vincent Vandeghinste, Lien Soetemans, Caro Brosens and Dimitar Shterionov. *GoSt-ParC-Sign: Gold Standard Parallel Corpus of Sign and spoken language* . . . 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. *MaCoCu: Massive collection and curation of monolingual and bilingual data: focus on under-resourced languages* . . . 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. *First WMT Shared Task on Sign Language Translation (WMT-SLT22)* . . . 507
Maarit Koponen, Mary Nurminen, Nina Havumetsä and Juha Lång. *DECA: Democratic epistemic capacities in the age of algorithms* . . . 509
Yves Scherrer, Olli Kuparinen and Aleksandra Miletic. *CorCoDial - Machine translation techniques for corpus-based computational dialectology* . . . 511
Julian Hamm and Judith Klein. *How STAR Transit NXT can help translators measure and increase their MT post-editing efficiency* . . . 513
Lucía Ormaechea, Pierrette Bouillon, Maximin Coavoux, Emmanuelle Esperança-Rodier, Johanna Gerlach, Jerôme Goulian, Benjamin Lecouteux, Cécile Macaire, Jonathan Mutal, Magali Norré, Adrien Pupier and Didier Schwab. *PROPICTO: Developing Speech-to-Pictograph Translation Systems to Enhance Communication Accessibility* . . . 515
Mikko Aulamo, Nikolay Bogoychev, Shaoxiong Ji, Graeme Nail, Gema Ramírez-Sánchez, Jörg Tiedemann, Jelmer van der Linde and Jaume Zaragoza. *HPLT: High Performance Language Technologies* . . . 517

**Sponsors** . . . **521**