GITT 2024

2nd International Workshop on Gender-Inclusive Translation Technologies

Proceedings of the Workshop

June 27, 2024

The GITT organizers gratefully acknowledge the support from the following sponsors.

Sponsors









The papers published in this proceedings are —unless indicated otherwise— covered by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC-BY-NCND 4.0). 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/4.0/deed.en ©2024 The authors

ISBN 978-1-0686907-2-3

Preface

This volume contains the proceedings of the 2nd International Workshop on Gender-Inclusive Translation Technologies (GITT 2024)¹, hosted by the 25th Annual Conference of The European Association for Machine Translation (EAMT 2024)². GITT is set out to focus on gender-inclusive language in translation and cross-lingual scenarios. The workshop brings together researchers from diverse areas, including industry partners, MT practitioners and language professionals. Also, GITT aims to encourage multidisciplinary research that develops and interrogates both solutions and challenges for addressing bias and promoting gender inclusivity in MT and translation tools, including LLM applications for translation.

The workshop welcomed three types of contributions: research papers, research communications, and extended abstracts. GITT-2024 received a total of 6 novel submissions (5 research papers, 1 extended abstract) and 2 research communication. Following the review process, all 6 submissions received positive reviews, highlighting an increase in the quality of the submissions received (i.e. GITT-2023 resulted in an acceptance rate of 75%). It is worth noting that the research communications did not undergo the review process as it had previously undergone peer-review at a top-tier conference. Of the accepted papers, 4 have been assigned to oral presentations, while the remaining 1, as well as the accepted abstract, have been assigned to the poster session. The research communications, which are not included in the proceedings, are also to be presented during the poster session in order to promote dissemination of research aligned with the scope of the workshop.

The accepted papers cover a diverse range of topics related to the analysis, measurement, and mitigation of gender bias in (Machine) Translation, as well as to the investigation of inclusive language. We are glad to attest to the interdisciplinary perspectives and methods represented in GITT submissions. The contributions range from technical papers proposing novel methods to position papers, user-centric experiments on the use of inclusive language, including reflection on the translation and localization of archival data with an inclusive and historically-grounded perspective.

In addition to the technical programme, we are honoured to have four invited speakers: Kevin Robinson (Google Research), with a keynote entitled "*Multilingual gender-inclusivity in translation and beyond*"; Begoña Martínez Pagán (University of Murcia) with the keynote "*Intersectionality and gender in translation* — *how ethical must one automatically be?*". Finally, the program includes a panel session on "Navigating Gender Inclusivity: From Research to Professional Practice", which – on top of the invited keynote speakers – includes two additional panelists: Paula Manzur (Booking.com) and Helena Moniz (University of Lisbon, INESC-ID/Unbabel).

We sincerely thank all the people and institutions that contributed to the success of the workshop: the authors of the submitted papers for their interest in the topic; the Programme Committee members for their valuable feedback and insightful comments; the EAMT organizers for their support. Finally, we thank our sponsors, Google, the Faculty of Arts and Philosophy at Ghent University, and Tilburg University for their generous contributions.

We hope you enjoy reading the papers and are looking forward to a fruitful and enriching workshop!

June 2024,

Beatrice Savoldi, Janiça Hackenbuchner, Luisa Bentivogli, Joke Daems, Eva Vanmassenhove, and Jasmijn Bastings

¹https://sites.google.com/tilburguniversity.edu/gitt2024
²https://eamt2024.sheffield.ac.uk/

Organizing Committee

Program Chairs

Beatrice Savoldi, Fondazione Bruno Kessler (FBK), Trento, Italy

Janiça Hackenbuchner, Department Translation, Interpreting and Communication, Ghent University, Ghent, Belgium

Luisa Bentivogli, Fondazione Bruno Kessler (FBK), Trento, Italy

Joke Daems, Department Translation, Interpreting and Communication, Ghent University, Ghent, Belgium

Eva Vanmassenhove, Department Cognitive Science and Artificial Intelligence, School of Humanities and Digital Sciences, Tilburg University (TiU), Tilburg, The Netherlands

Jasmijn Bastings, Google DeepMind

Program Committee

Program Committee

Dennis Fucci, Fondazione Bruno Kessler Andrea Piergentili, University of Trento Pushpdeep Singh, Tata Consultancy Services Limited, India Danielle Saunders, DeepL Jan Niehues, Karlsruher Institut für Technologie Guillaume Wisniewski, LLF / Université Paris Cité Anna Currey, Amazon Beatrice Spallaccia, University of Bologna Manuel Lardelli, Universität Graz Gabriel Stanovsky, Hebrew University of Jerusalem Nizar Habash, New York University Abu Dhabi María Isabel Rivas Ginel, Dublin City University Michal Měchura, Dublin City University Johanna Monti, University of Naples L'Orientale Matúš Pikuliak, Kempelen Institute of Intelligent Technologies Ranjita Naik, Microsoft Hannah Devinney, Umea University Giuseppe Attanasio, Instituto de Telecomunicações

Invited Speakers

Kevin Robinson, Google DeepMind Begoña Martínez Pagán, University of Murcia

Panelists

Paula Manzur, Booking.com Helena Moniz, School of Arts and Humanities, University of Lisbon Kevin Robinson, Google DeepMind Begoña Martínez Pagán, University of Murcia

Keynote Talk Multilingual gender-inclusivity in translation and beyond

Kevin Robinson

Google DeepMind

Abstract: Multilingual capabilities are increasingly available in general-purpose systems, rather than from dedicated MT systems alone. This shift impacts many practical concerns for improving gender inclusivity such as understanding downstream developer usage patterns, improving the validity of upstream evaluations, and scaling to global cultural contexts. It also raises sociotechnical research challenges in creating new kinds of transparently multilingual user experiences, improving controllability of gender-inclusive representations, and enabling new modalities like multilingual image understanding and audio generation. I discuss empirical work to measure potential misgendering harms in PaLM 2, and share experiences from more recent research at Google.

Bio: Kevin Robinson is a Senior Research Engineer at Google, working on developing new techniques for inclusive, controllable, and robust machine learning systems by effectively blending technical and sociocultural perspectives. Kevin has worked on research efforts like PaLM, PaLM-FLAN and PaLM 2, and contributed to products like Bard and Gemini. Kevin has separately co-authored publications on language models related to pre-training data, synthetic data generation, and measuring misgendering harms in translation systems. He is currently focused on measuring cultural and representational harms in ways that incorporate community-informed perspectives. Kevin has also worked as a special education teacher, and a computer science education researcher at MIT focused on bias within CS classrooms.

Keynote Talk Intersectionality and gender in translation — how ethical must one automatically be?

Begoña Martinez Pagan

University of Murcia

Abstract: To which extent should ethical considerations inform (automated) inclusive translation processes? This talk will present a reflection on criteria for the minimum requirements of translation ethics that could be applied systematically to any text, from the point of view of intersectional, queer and feminist principles. By critically examining the ethical dimensions of translation through these lenses, this talk will seek to illuminate the path toward more inclusive, equitable, and socially responsible translation practices.

Bio: Begoña Martínez Pagán is a translator, interpreter, and author based at the English Studies Department of the University of Murcia. Her activism, lecturing, and research include intersections of her profession with feminist and LGBTIQ+ literature, inclusive language, human rights, business organization, and open-source software.

Table of Contents

Gender Bias Evaluation in Machine Translation for Amharic, Tigrigna, and Afaan Oromoo
Walelign Tewabe Sewunetie, Atnafu Lambebo Tonja, Tadesse Destaw Belay, Hellina Hailu Nigatu, Gashaw Kidanu Gebremeskel, Zewdie Mossie, Hussien Seid and Seid Muhie
Yimam1
Sparks of Fairness: Preliminary Evidence of Commercial Machine Translation as English-to- German Gender-Fair Dictionaries
Manuel Lardelli, Timm Dill, Giuseppe Attanasio and Anne Lauscher
Gender and bias in Amazon review translations: by humans, MT systems and ChatGPT Maja Popovic and Ekaterina Lapshinova-Koltunski
You Shall Know a Word's Gender by the Company it Keeps: Comparing the Role of Context in Human Gender Assumptions with MT
Janiça Hackenbuchner, Joke Daems, Arda Tezcan and Aaron Maladry
Lost in Translation? Approaches to Gender Representation in Multilingual Archives Mrinalini Luthra and Brecht Nijman
Pilot testing gender-inclusive translations and machine translations for German quadball re- feree certification test takers
Joke Daems

Program

Thursday, June 27, 2024

- 09:00 09:15 Opening Remarks
- 09:15 10:15 *Keynote 1 (Kevin Robinson, Google DeepMind): Multilingual genderinclusivity in translation and beyond*
- 10:15 10:30 Poster Boaster
- 10:30 11:00 *Coffee Break*
- 11:00 12:30 Oral Presentations

Gender Bias Evaluation in Machine Translation for Amharic, Tigrigna, and Afaan Oromoo

Walelign Tewabe Sewunetie, Atnafu Lambebo Tonja, Tadesse Destaw Belay, Hellina Hailu Nigatu, Gashaw Kidanu Gebremeskel, Zewdie Mossie, Hussien Seid and Seid Muhie Yimam

Lost in Translation? Approaches to Gender Representation in Multilingual Archives

Mrinalini Luthra and Brecht Nijman

Gender and bias in Amazon review translations: by humans, MT systems and ChatGPT

Maja Popovic and Ekaterina Lapshinova-Koltunski

Sparks of Fairness: Preliminary Evidence of Commercial Machine Translation as English-to-German Gender-Fair Dictionaries Manuel Lardelli, Timm Dill, Giuseppe Attanasio and Anne Lauscher

- 12:30 13:30 Lunch
- 13:30 14:30 *Keynote 2 (Begoña Martínez Pagán, University of Murcia): Intersectionality and gender in translation — how ethical must one automatically be?*
- 14:30 15:30 Poster Session

You Shall Know a Word's Gender by the Company it Keeps: Comparing the Role of Context in Human Gender Assumptions with MT Janiça Hackenbuchner, Joke Daems, Arda Tezcan and Aaron Maladry

Building Bridges: A Dataset for Evaluating Gender-Fair Machine Translation into German Manuel Lardelli, Giuseppe Attanasio and Anne Lauscher

Thursday, June 27, 2024 (continued)

Hi Guys or Hi Folks? Towards Gender-Neutral Machine Translation Andrea Piergentili, Beatrice Savoldi, Dennis Fucci, Matteo Negri and Luisa Bentivogli

Pilot testing gender-inclusive translations and machine translations for German quadball referee certification test takers Joke Daems

- 15:00 15:30 *Coffee Break*
- 15:30 17:00 Panel: Navigating Gender Inclusivity: From Research to Professional Practice
- 17:00 17:00 Closing Remarks