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

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

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

[1] https://sites.google.com/tilburguniversity.edu/gitt2024
[2] https://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

Program

Thursday, June 27, 2024

09:00 - 09:15 *Opening Remarks*

09:15 - 10:15 *Keynote 1 (Kevin Robinson, Google DeepMind): Multilingual gender-inclusivity 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

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

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*

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

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

Thursday, June 27, 2024 (continued)

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

Lost in Translation? Approaches to Gender Representation in Multilingual Archives
Mrinalini Luthra and Brecht Nijman

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*

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

Joke Daems
Ghent University
joke.daems@ugent.be
International Quadball Association
joke.daems@iqasport.org

Abstract

Gender-inclusive translations are the default at the International Quadball Association, yet translators make different choices for the (timed) referee certification tests to improve readability. However, the actual impact of a strategy on readability and performance has not been tested. This pilot study explores the impact of translation strategy (masculine generic, gender-inclusive, and machine translation) on the speed, performance and perceptions of quadball referee test takers in German. It shows promise for inclusive over masculine strategies, and suggests limited usefulness of MT in this context.

1 Introduction

While the inherent importance of gender-inclusive language is clear (Sczesny et al., 2021), a commonly heard argument against the use of gender-inclusive language strategies is that they negatively impact readability and comprehensibility. With some notable exceptions (Friedrich et al., 2021), however, this impact has not been empirically tested.

At the International Quadball Association (IQA), gender-inclusive language is of critical importance, given the sports' commitment to gender inclusivity. While IQA translators aim to produce gender-inclusive translations, the desire for readability can outweigh the desire for inclusivity, particularly in the context of timed assessment for the referee certification tests (Daems, 2023)[1]. The German IQA translation team currently uses the colon as the non-binary marker in most translations, which also seems to be the strategy preferred by professional translators (Paolucci et al., 2023). This pilot study was conducted to answer the following research questions about referee certification test takers in German:

- Does inclusive language lead to slower answer times than generic masculine?
- Does inclusive language lead to lower test scores than generic masculine?
- Is machine translation (MT) a viable alternative when there are no translators available, considering answer time and test scores?
- What are test takers' perceptions about the understandability, readability, speed, and correctness of different conditions?

2 Methodology

A survey was created in Qualtrics (Qualtrics, Provo, UT) and distributed in April and May 2024. The main test block was randomised so as to evenly present the three conditions - gender-inclusive, generic masculine, and machine translation - to participants. It consisted of 14 multiple-choice questions taken from the official referee tests. Only questions with multiple references to people were selected to guarantee differences between the conditions. Questions were translated by IQA translators into the generic masculine and the gender-inclusive variant. For MT, each question and answer was translated using DeepL (translations generated in April 2024).

[1]Referees need to be certified to serve during official IQA games. Certification tests are created and hosted by the IQA and can be taken online at any time via https://hub.iqasport.org/

Proceedings of the 2nd International Workshop on Gender-Inclusive Translation Technologies, pages 56–57
June 27, 2024 ©2024 Association for Computational Linguistics

At the end of the survey, participants were asked how strongly they agreed with the following statements: "I could understand the questions and answers", "I found it easy to read the questions and answers", "I answered the questions as fast as I would in a real referee test", and "I answered most questions correctly" (Likert scale of 1 = "Not at all" to 5 = "Completely").

Twenty-four valid survey responses were collected (eight for each condition). All participants were native German-speaking quadball players, and all but one were currently or soon to be certified referees. Statistical analyses were conducted using Microsoft Excel (ANOVA) and RStudio (linear mixed effects models), but no statistical differences were found between conditions so only descriptive numbers are presented here.

3 Results

Speed: Surprisingly, test takers were fastest in the inclusive condition, despite the text being 8% longer than the masculine condition (Table 1).

condition	mean	median	stdev	min	max
inclusive	378	396	110	182	551
masculine	465	453	144	216	780
MT	460	495	144	231	706

Table 1: Descriptive statistics for total time needed to answer all 14 questions (in seconds) per condition.

Reading speed is especially high for correctly answered questions in that condition (Figure 1).

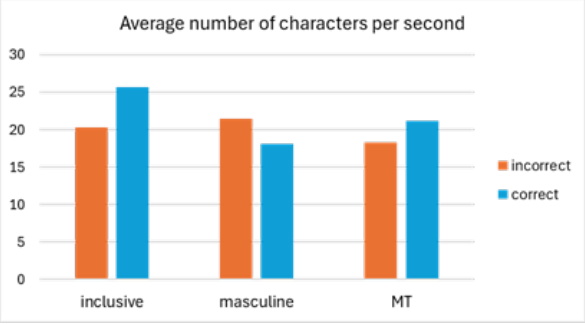

Figure 1: Average number of characters per second for the three conditions for questions answered incorrectly and correctly (excluding 'I don't know').

Performance: Participants in the inclusive condition scored highest on the test (Table 2), followed by those in the generic masculine condition.

Perceptions: MT scores worst overall (Figure 2). The inclusive condition scored highest on understandability and perceived correctness.

condition	mean	median	stdev	min	max
inclusive	10	10	2,1	6	13
masculine	9	8	3	5	13
MT	7,8	9	2,2	3	9

Table 2: Descriptive statistics for test scores per condition, max score = 14.

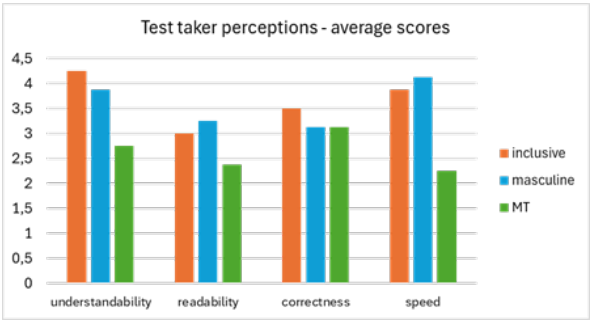

Figure 2: Average test taker perceptions per condition.

4 Conclusion & Future Work

Results suggest that (contrary to oft-heard criticism) speed and test scores are actually highest for the inclusive condition, showing its potential going forward. Based on the pilot study findings and participant feedback, MT is not currently seen as a viable strategy for referee test translation. Given the small sample size, statistically significant differences could not be identified, so for future work, we will expand this work by creating two variants (gender-inclusive and generic masculine) of the official IQA referee tests, in order to collect data from the entire population of referee test takers in a real-life setting.

References

Daems, Joke. 2023. Gender-inclusive translation for a gender-inclusive sport: Strategies and translator perceptions at the in-ternational quadball association. In *Proceedings of the First Workshop on Gender-Inclusive Translation Technologies*, page 37–47.

Friedrich, Marcus C. G., Veronika Drößler, Nicole Oberlehberg, and Elke Heise. 2021. The influence of the gender asterisk ("genderstern-chen") on comprehensibility and interest. *Frontiers in Psychology*.

Paolucci, Angela Balducci, Manuel Lardelli, and Dagmar Gromann. 2023. Gender-fair language in translation: A case study. In *Proceedings of the First Workshop on Gender-Inclusive Translation Technologies*, page 13–23.

Sczesny, Sabine, Franziska Moser, and Wendy Wood. 2021. Beyond sexist beliefs: How do people decide to use gender-inclusive language? *Personality and Social Psychology Bulletin*, 41(7):943–954.