

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).

¹Referees 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/>

© 2024 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

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).

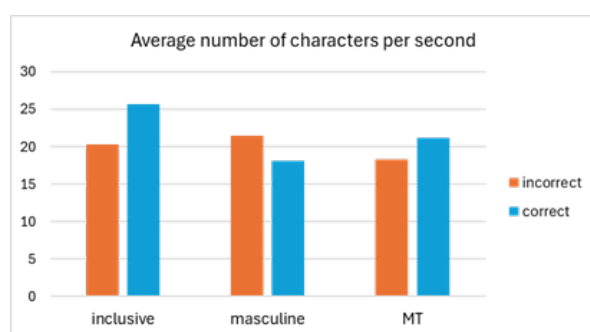


Figure 1: Average number of characters per second for the three conditions for questions answered incorrectly and correctly (excluding ‘I don’t know’).

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.

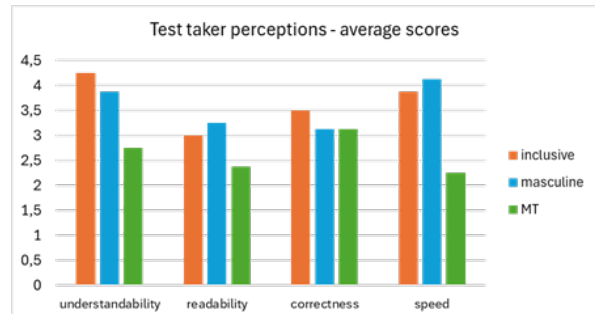


Figure 2: Average test taker perceptions per condition.

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 Dröbler, 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.