How adaptive is adaptive machine translation, really?
A gender-neutral language use case

Aida Kostikova
Ghent University
aida.kostikova@ugent.be

Joke Daems
Ghent University
joke.daems@ugent.be

Todor Lazarov
New Bulgarian University
tdlazarov@gmail.com

Abstract

This study examines the effectiveness of adaptive machine translation (AMT) for gender-neutral language (GNL) use in English-German translation using the ModernMT engine. It investigates gender bias in initial output and adaptability to two distinct GNL strategies, as well as the influence of translation memory (TM) use on adaptivity. Findings indicate that despite inherent gender bias, machine translation (MT) systems show potential for adapting to GNL with appropriate exposure and training, highlighting the importance of customisation, exposure to diverse examples, and better representation of different forms for enhancing gender-fair translation strategies.

1 Introduction

With increasing adoption of GNL, its reflection in natural language processing (NLP) applications like MT becomes vital. Ignoring GNL trends can perpetuate biased representations and inequalities (Savoldi et al., 2021), and also much of prior work in the field of gender bias is built on techniques which assume gender is binary. AMT, which learns from users and adjusts to personal linguistic preferences (Bentivogli et al., 2015), may offer promise for ensuring GNL use in MT, especially given the fact that it reduces post-editing efforts and has shown certain potential for empowering the human in the loop (Martikainen, 2022).

This study evaluates the effectiveness of AMT for GNL use, focusing on non-binary language in English-German translation using the ModernMT[1] engine. Our research, while somewhat aligned with the works of Saunders et al. (2020), Sun et al. (2021) and Vanmassenhove et al. (2021) in their intent to address gender neutrality in MT, offers a distinct approach. While Sun et al. and Vanmassenhove et al. advocate for post-processing steps to rewrite gendered sentences into gender-neutral ones using either rule-based or neural approaches, and Saunders et al. fine-tune a base model with synthetic datasets, our research investigates the real-time adaptability of AMT systems to naturally integrate GNL into its translations. Their approaches, although effective, often require access to and control over the internal mechanisms of the MT system or extensive task-specific datasets, which is not always feasible with black-box systems like ModernMT. Our study, on the other hand, explores how these black-box systems can dynamically adapt to GNL in their translation process without the need for post-processing steps or specialised fine-tuning. In this way, we assess the system's intrinsic ability to adjust to GNL, offering a more dynamic and contextual view of how these systems may cope with evolving language trends over the long run.

We examine two distinct GNL strategies, such as gender asterisk and De-E-System. The gender asterisk form, which is increasingly utilised in German to inclusively represent all genders, is constructed by appending an asterisk before the gender-specific suffix (e.g., *Ärzt*innen*). Conversely, the De-E-System, which was developed by the Association for Gender Neutral German (Verein für geschlechtsneutrales Deutsch[2]) employs the concept of "Inklusivum", a proposed fourth grammatical gender. This system introduces new declination rules and modifications for forming gender-neutral articles, nouns, and pronouns: for example, in this system, nouns typically end with either *-e* or *-re*, accompanied by the gender-neutral article *de*: *de Schülere (student), de Autore (author)*.

Two versions of ModernMT were deployed in this study: a system without the use of TM, as well as its customised variant, which was exposed to a TM containing gender-neutral forms. The research investigates whether AMT exhibits gender bias in initial output, analyses its adaptability when working with two conceptually different GNL strategies, the role of TM in improving results, as well as factors influencing engine adaptation success.

---

[1] https://www.modernmt.com/

[2] https://geschlechtsneutral.net/gesamtsystem/

2 Data and Methodology

Four scenarios were studied: using gender asterisk and De-E-System, each with and without TM, using texts based on the 2006 Court Procedures Rules for the Australian Capital Territory, which have been selected due to the fact that they inherently follow GNL principles in English. However, they also contain numerous elements — specifically, terms denoting professional roles or titles (as *administrator, real estate agent, director of public prosecutions*) — which do not hold gendered connotations in English but could potentially introduce gender bias when translated into German. The study includes 15 sentences for each of the 15 selected role-related nouns that could require gender-neutral adaptation during translation. For instance, consider the sentence: *The real estate agent might be appointed to market the land and conduct the sale*. In German, this could be translated using the gender asterisk as *Die*der Immobilienmakler*in könnte beauftragt werden, das Grundstück zu vermarkten und den Verkauf durchzuführen*, or with the De-E-System as *De Immobilienmaklere könnte beauftragt werden, das Grundstück zu vermarkten und den Verkauf durchzuführen*.

The evaluation of the engine's adaptive performance was conducted through a blend of quantitative and qualitative methods. For a quantitative perspective, we utilized two automatic metrics, character-based translation edit rate (CharacTER) and keystrokes ratio (KSR), and we computed the total number of adapted segments. While these metrics do not directly measure gender bias or adaptivity rate, they provide essential insights into the editing effort required to correct the engine's output, thus serving as proxies for its adaptive performance, as lower values for CharacTER and KSR would suggest that the engine is adapting well to the GNL. These metrics also enable a comparison of results across different experiments, such as those employing various GNL strategies or comparing performance with and without TM utilisation.

For a more nuanced and targeted understanding of gender bias in the translation process, we complemented these metrics with a qualitative content analysis of specific instances of adaptation and a close examination of gender bias in the initial output.

3 Results

3.1 Gender bias in the adaptive MT output

Our findings align with previous research in the field of gender bias in MT (Monti, 2020; Savoldi et al., 2021), indicating that ModernMT also exhibits gender bias in their output when using both gender-neutral strategies. Analysis showed significant masculine bias in both experiments, with De-E-System having higher bias with TM (87% and 88% masculine translations in untrained and trained engines, respectively; presence of gender-neutral forms increased by 2.7%). Conversely, gender asterisk strategy exhibited less bias: without TM, 78% segments were masculine; with TM, masculine translations dropped by 17.2% and gender-neutral translations increased by 22.7%. It should be noted that these percentages represent an absolute increase in gender-neutral translations. By focusing on absolute changes, we could directly observe the shifts in translation behaviour due to the integration of the TM with gender-neutral forms, thus offering a clearer understanding of the potential of AMT for adopting GNL strategies.

Interestingly, some nouns were consistently translated with feminine gender, and all such instances were related to specific roles, such as *appellant, defendant* and *plaintiff*, who are less likely to be involved in decision-making, managing, and investigating functions (as opposed to, for instance, *employer, examiner, expert, liquidator*, which were predominantly translated with masculine gender). This discrepancy potentially indicates two sources of bias: pre-existing and technical bias (Friedman and Nissenbaum, 1996).

Pre-existing bias refers precisely to any asymmetries which are rooted in society at large or which reflect personal biases of individuals responsible for the system development. Technical bias, in contrast, manifests in the stages of data collection, system design, training, and testing procedures for MT models. In the context of our research, the fact that some roles were predominantly assigned masculine translations, while others consistently appeared with feminine connotations, might indicate such deep-rooted pre-existing biases within the training data. Moreover, according to the study of European Commission for the Efficiency of Justice (CEPEJ, 2016), although women frequently succeed in entering the legal field, their progression into senior positions tends to be slower. Thus, the achieved result might be explained by asymmetries present in the data used by the MT system.

3.2 Overall adaptivity to GNL

The percentage of gender-neutral translations for the experiments demonstrate diverse adaptability of ModernMT when handling GNL. And although the overall adaptation rate remains relatively low, along with inconsistent adaptation across all four experiments in this study, some clear trends were observed.

Firstly, the engine demonstrated better adaptation over time when working with the gender asterisk system, with an increased number of adapted segments

towards the end of the project (by the last two or three words) for both untrained and trained engines. Secondly, it was determined that, to achieve such progress, it is essential for the system to have sufficient exposure to gender-neutral words in various forms. This includes considering grammatical number, case, and different types of articles, as for instance, the system struggled with adapting to plurals due to an insufficient number of examples in the corpus. This highlights the importance of ensuring the variability and exposure to different grammatical and syntactical structures, possibly with the help of TM.

This discrepancy in the adaptation success of the gender asterisk method over the De-E-System could be attributed to a combination of factors. First, the wider prevalence of gender asterisk forms in the German language (Burtscher et al., 2022) may inherently favour this method, as its similarity to conventional language forms likely aids model recognition. Second, the instance-based learning approach of ModernMT (Piergentili et al., 2023) can help the system to learn from similar, even non-identical instances, and generalise to unseen examples. This ability, when coupled with the potential presence of gender-neutral structures resembling the gender asterisk method in ModernMT's training data, could have facilitated the engine's capacity to adapt more effectively to this method. These contributing factors provide a plausible explanation for the observed discrepancy, wherein the model displayed a superior adaptation to the gender asterisk approach as compared to the De-E-System.

4 Conclusion and Future Research

These findings suggest that AMT systems, despite being prone to gender bias, have the potential to adapt to GNL forms (especially if they are more widespread) with appropriate exposure, although further refinement and optimisation are necessary to improve their adaptability to GNL. TMs facilitated adaptation by exposing the system to diverse GNL examples, thus enhancing its recognition and adaptation of gender-neutral variants. Solutions for better customisation (Lardelli and Gromann, 2023) ensuring better representation of different forms for various strategies will be crucial in advancing gender-fair translation strategies.

It is important to acknowledge that the limited number of sentences may not fully capture all the nuances of translating GNL in broader contexts. This study, therefore, should be considered a preliminary exploration of this complex linguistic area. Future research is necessary to validate these results in more extensive and diverse corpora, which would provide a more comprehensive understanding of AMT systems' performance when dealing with GNL.

References

Burtscher, S., Spiel, K., Klausner, L.D., Lardelli, M. and Gromann, D. 2022. "Es geht um Respekt, nicht um Technologie": Erkenntnisse aus einem Interessensgruppen-übergreifenden Workshop zu genderfairer Sprache und Sprachtechnologie. *Proceedings of Mensch und Computer 2022 (pp. 106-118)*.

Bentivogli, L., Bertoldi, N., Cettolo, M., Federico, M., Negri, M. and Turchi, M. 2015. On the evaluation of adaptive machine translation for human post-editing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(2), pp.388-399.

CEPEJ Report, European Commission for the Efficiency of Justice. (2016). *Evaluation of European judicial systems - CEPEJ report*.

Friedman, B., and Nissenbaum, H. (1996). Bias in computer systems. In *ACM Transactions on Information Systems (TOIS), 14(3)*, pp. 330–347.

Lardelli, M. and Gromann, D. 2022. Gender-Fair (Machine) Translation. *New Trends in Translation and Technology 2022*, p.166.

Martikainen, H. 2022. *Ghosts in the machine: Can adaptive MT help reclaim a place for the human in the loop?*

Monti, J. 2020. Gender issues in machine translation: An unsolved problem? *The Routledge Handbook of Translation, Feminism and Gender*, pp. 457-468.

Savoldi, B., Gaido, M., Bentivogli, L., Negri, M., Turchi, M. 2021. Gender Bias in Machine Translation. Transactions of the Association for Computational Linguistics; 9, pp. 845–874.

Saunders, D., Sallis, R., and Byrne, B. (2020). Neural machine translation doesn't translate gender coreference right unless you make it. *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*, pp. 35–43.

Sun, T., Webster, K., Shah, A., Wang, W.Y. and Johnson, M. (2021) They, them, theirs: Rewriting with gender-neutral English. *arXiv preprint arXiv:2102.06788*.

Piergentili, A., Fucci, D., Savoldi, B., Bentivogli, L. and Negri, M. 2023. From Inclusive Language to Gender-Neutral Machine Translation. *arXiv preprint arXiv:2301.10075*.

Vanmassenhove, E., Emmery, C. and Shterionov, D. (2021). NeuTral Rewriter: A Rule-Based and Neural Approach to Automatic Rewriting into Gender-Neutral Alternatives. *arXiv preprint arXiv:2109.06105*.