GeFMT: Gender-Fair Language in German Machine Translation

Manuel Lardelli
University of Graz / Austria
manuel.lardelli@uni-graz.at

Anne Lauscher
University of Hamburg / Germany
anne.lauscher@uni-hamburg.de

Giuseppe Attanasio
Instituto de Telecomunicações / Lisbon, Portugal
giuseppeattanasio6@gmail.com

Abstract

Research on gender bias in Machine Translation (MT) predominantly focuses on binary gender or few languages. In this project, we investigate the ability of commercial MT systems and neural models to translate using gender-fair language (GFL) from English into German. We enrich a community-created GFL dictionary, and sample multi-sentence test instances from encyclopedic text and parliamentary speeches. We translate our resources with different MT systems and open-weights models. We also plan to post-edit biased outputs with professionals and share them publicly. The outcome will constitute a new resource for automatic evaluation and modeling gender-fair EN-DE MT.

1 Background

A wealth of research in the field of Machine Translation (MT) focuses on gender bias (Savoldi et al., 2021). However, most recent efforts predominantly focus on binary gender only (Lardelli and Gromann, 2023) or few languages. For instance, first contributions benchmark gender-neutral language in English-to-Italian MT (Piergentili et al., 2023). Gender-neutrality refers to avoiding gender-specific elements, often by rewording sentences with indefinite pronouns and passive constructions, amongst others. In contrast to gender-neutrality, new proposals advocate for gender-inclusive translation, e.g., by accounting for neopronouns (e.g., *xe/xem*), neomorphemes and characters such as the gender star (*) in German (e.g., *der*die Berater*in*). Such approaches increase gender visibility and might therefore be preferred for reference to non-binary individuals.

2 Project Overview

The present project is a one-year joint effort of the University of Graz and Hamburg, and the Instituto de Telecomunicações in Lisbon. It encompasses (*i*) the creation of a dataset we will share with the community for gender-fair MT,[1] (*ii*) human and automatic assessment of the use (or lack thereof) of gender-fair language in different commercial and non-commercial MT systems, and (*iii*) test new methods to generate translations based on gender-fair post-edited datasets. These steps are detailed in the following paragraphs.

(*i*) We use a community-generated gender-fair dictionary in German as a basis for our experiments.[2] This dictionary lists numerous common nouns and gender-neutral alternatives proposed by people who engage in language inclusivity. Drawing on this resource, we create a dataset by randomly selecting 128 entries and expand this list by adding all possible forms in German in the singular and plural, i.e., masculine, feminine, as well as gender-neutral and gender-inclusive alternatives and the English translation. We filter out those that were already neutral, e.g., *"Star"*, which is an Anglicism and does not have variants for other genders in German. Additionally, we remove polysemous terms, e.g., *"aid"*, to facilitate back-translation into English. Our final dictionary counts 115 in their singular and plural forms, containing both professions and common nouns. Ta-

---

[1]We use the term "gender-fair" to subsume both gender-neutral and inclusive approaches as in Sczesny (2016).
[2]https://geschicktgendern.de/

English	MS	MP	FS	FP	GN	GI
The counsellor	Berater	Berater	Beraterin	Beraterinnen	Beratende	der*die Berater*in
The pacifist	Pazifist	Pazifisten	Pazifistin	Pazifistinnen	Friedensbewegte	der*die Pazifist*in

Table 1: Example of the proposed dataset. For each English noun, we collect the German masculine/feminine singular (MS/FS), plural (MP/FP), gender-neutral (GN), and gender-inclusive (GI) forms.

ble 1 reports an example.

(*ii*) In order to test the use of gender-fair language among different systems, we back translate the singular and plural terms in our dataset from English into German with Google Translate, DeepL, GPT-3.5 and GPT-4, and open-weights models, including Opus-MT, FLAN, and Llama models. As a first quantitative inspection, we match the translation outputs with the alternatives in our data set. Next, we conduct a qualitative analysis to gain more insights into the translation of individual terms whose gender is ambiguous with no surrounding context. Subsequently, we investigate the influence of context by collecting and translating sentences that contain our words.

We collect an additional set of English passages that mention our dictionary entries. We sample sentences from Europarl (Koehn, 2005) and Wikipedia.[3] Europarl is a widely recognized benchmark dataset for MT displaying institutional language from parliamentary speeches–perhaps amongst the first contexts GFL was designed for (Piergentili et al., 2023). Wikipedia presents encyclopedic text, opening to new contexts where our seed nouns appear.

(*iii*) Finally, we plan to hire an expert in translation and gender-fair language to create different gender-neutral and inclusive versions of the outputs via post-editing. It will entail gender-neutral rewording, using a gender-inclusive character, e.g. gender star (*), and one or two different neo-systems. We will use these outputs as examples to test the few-shot learning capabilities of different Large Language Models (LLMs) for performing gender-fair translation and to develop a gender-fair MT benchmark.

3 Expected Outcome

This project will produce the following openly-accessible resources to the community: **i)** A new human-curated dictionary of English nouns with their German gender-fair inflections. **ii)** A parallel EN-DE corpus with source sentences from Wikipedia and European Parliament speeches that mention the dictionary nouns, and hypothesis sentences automatically translated with several state-of-the-art systems. **iii)** A new human-curated corpus of gender-fair German translations of the sentences above, obtained via post-editing machine-generated translations. Taken together, these resources will constitute the largest collection for studying automatic gender-fair translation from English into German.

References

Koehn, Philipp. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of Machine Translation Summit X: Papers*, pages 79–86, Phuket, Thailand, September 13-15.

Lardelli, Manuel and Dagmar Gromann. 2023. Gender-fair (machine) translation. In *Proceedings of the New Trends in Translation and Technology Conference - NeTTT 2022*, pages 166–177.

Piergentili, Andrea, Beatrice Savoldi, Dennis Fucci, Matteo Negri, and Luisa Bentivogli. 2023. Hi guys or hi folks? benchmarking gender-neutral machine translation with the GeNTE corpus. In Bouamor, Houda, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 14124–14140, Singapore, December. Association for Computational Linguistics.

Savoldi, Beatrice, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2021. Gender bias in machine translation. *Transactions of the Association for Computational Linguistics*, 9:845–874, 08.

Sczesny, Sabine, Magda Formanowicz, and Franziska Moser. 2016. Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in Psychology*, 7.

---

[3]We use the snapshot at 01–03–2022 at `https://huggingface.co/datasets/wikipedia`.