Sparks of Fairness: Preliminary Evidence of Commercial Machine Translation as English-to-German Gender-Fair Dictionaries

Manuel Lardelli^{*}, Timm Dill^{\(\lambda\)}, Giuseppe Attanasio^{*}, Anne Lauscher^{\(\lambda\)}

 University of Graz, Austria
 University of Hamburg, Germany
 Instituto de Telecomunicações, Lisbon, Portugal manuel.lardelli01@gmail.com

Abstract

Bilingual dictionaries are bedrock components for several language tasks, including translation. However, dictionaries are traditionally fixed in time, thus excluding those neologisms and neo-morphemes that challenge the language's nominal morphology. The need for a more dynamic, mutable alternative makes Machine Translation (MT) systems become an extremely valuable avenue. This paper investigates whether commercial MT can be used as bilingual dictionaries for gender-fair translation. We focus on the English-to-German pair, where notional gender in the source requires gender inflection in the target. We translated a dataset with person-referring terms using Google Translate, Microsoft Bing, and DeepL and discovered that while each system is heavily biased towards the masculine gender, DeepL often provides gender-fair alternatives to users, especially with plurals.

1 Introduction

"The past is print dictionaries; the present is print dictionaries with some electronic versions of the same text; the future must be print dictionaries and truly electronic dictionaries, compiled afresh for the new medium, enriched with new types of information the better to meet the needs of the multifarious users."

- Beryl Sinclair, 1996

The way we speak about humans influences our mental representation of them – psychological research shows that thus, using gender-fair language can reduce gender-related stereotyping and discrimination (Sczesny et al., 2016). Accordingly, many national and international organizations in Europe and beyond (e.g., universities and even the European Parliament¹) are increasingly adopting gender-fair language, and, for instance, publishing guidelines and recommendations on the topic. In translation, the topic of gender-fair language (GFL) is specifically interesting as we are often facing a gender-neutral person word (e.g., the workers in English), which needs to be translated to a language like German, in which using a gendered form would be the most traditional choice (e.g., die Arbeiter or die Arbeiterinnen in German). Often, using a gendered form only will, however, simply reflect existing stereotypes (e.g., occupational stereotypes) and also lead to the reinforced exclusion of individuals who do not identify with the specific grammatical gender chosen, like non-binary individuals (Dev et al., 2021).

In this work, we hypothesize that, given the widespread use of language technology, Machine Translation (MT) can be a key enabler in the adoption of gender-fair language for non-native speakers and in scenarios involving organizations that act internationally. Still, the existing research landscape on the behavior of commercial MT systems concerning gender-fair language is scarce: existing studies have looked at a few specific language pairs (and translation directions) scenarios, and domains only (e.g., (Savoldi et al., 2023), *inter alia*). For instance, there exists barely information on gender fairness in English-to-German MT. As MT systems are increasingly used as vocabularies

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

¹https://www.europarl.europa.eu/cmsdata/ 151780/GNL_Guidelines_EN.pdf

(Cotelli Kureth et al., 2023) (i.e., to translate single words without any further context), we pose the following research question: *Can commercial MT be used as English-to-German gender-fair dictionaries?* If we would find commercial MT to produce gender-fair translations, one could think of leveraging this potential for bootstrapping more research on gender-fair MT, further motivating our question.

Contributions. We present the first study on English-to-German gender-fair language in commercial MT focused on dictionary-like translations to date. To this end, we employ a communitycreated gender-fair dictionary for German, from which we sample seed nouns, which we translate into English. We then start from the English terms, and (1) conduct a pre-study in which we assess the general potential of three popular commercial MT systems (Google Translate, Microsoft Bing, and DeepL) for gender-fair MT. Based on these findings, we (2) conduct an in-depth study on DeepL, in which we test singular and plural forms and provide statistics on the exact type of gender-fair language we observe. Our findings show, for instance, that DeepL often provides gender-fair alternatives to the users, but that the system is heavily biased towards masculine translations (roughly 67% of the outputs). Interestingly, in the plural, gender-fair outputs are much more frequent than in the singular, with the participial form and BinnenI being the most common. We hope that our work fuels more research on gender-fair language in English-to-German MT.

Bias Statement. We collect English-to-German system outputs and analyze the overt gender of the translations. If a gender-fair output is present, we categorize it into its specific form. Our work therefore addresses overt gender bias in the output, and, accordingly, the issue of *representational harm* (stereotyping and exclusion) (Barocas et al., 2017).

2 Related Work

Bilingual dictionaries are bedrocks of various linguistic applications, including language learning (Thompson, 1987) and translation. Motivated by such an important role, several efforts have studied how to extract them at scale (Nagata et al., 2001) or integrate them in neural machine translation systems (Duan et al., 2020; Zhang et al., 2021). In this paper, we study modern commercial translation systems' as EN-DE bilingual dictionaries with a focused eye on gender-fair forms.

Our work is part of a broader discourse on fairness and inclusivity in machine translation. Neural and commercial systems are known to encode stereotypical views on genders (Savoldi et al., 2021), leading to brittle gender inflection capabilities in grammatical gender languages (Stanovsky et al., 2019; Attanasio et al., 2023), covert biases in genderless languages (Ciora et al., 2021), and inadequate handling of neo-morphemes (Lauscher et al., 2023) and named entities (Saunders and Olsen, 2023). Further, systems are nearly incapable of gender-neutral translation for human entity nouns in EN-IT (Piergentili et al., 2023) and DE-EN (Savoldi et al., 2023). To ground automatic translation with human practices, recent studies have reported on neutralization and genderinclusive strategies used by professional translators (Daems, 2023; Paolucci et al., 2023) and MT post-editors (Lardelli and Gromann, 2023a).

These findings and research efforts underscore the need for our research on commercial MT dictionary capabilities. As Sinclair poses it, static, bilingual dictionaries suffer from gaps in coverage, e.g., failing to include neologisms (Atkins, 1996). This work studies whether modern commercial systems have fixed on traditional gender forms or are indeed "meeting the needs of the multifarious user."

3 Background

As a basis for this work, we first introduce the relationship between gender and language ($\S3.1$), followed by the definition of gender-fair language (GFL) ($\S3.2$), and possible strategies in German ($\S3.3$).

3.1 Linguistic Gender

The term gender may refer to a linguistic feature and an extra-linguistic reality. Linguistic gender can be divided into grammatical, lexical, and referential (Cao and Daumé III, 2020; Corbett, 1991). Grammatical gender pertains to the classification of nouns into categories such as masculine, feminine, and neuter. For instance, "sun" is masculine in Italian ("il sole") but feminine in German ("die Sonne"). Lexical gender describes the semantic property of femaleness or maleness of a noun, such as "mother" and "father". Referential gender refers to the extra-linguistic reality, i.e., the gender of a noun reflects the gender identity of the referent, e.g., "Schauspieler" (EN: male actor) and "Schauspielerin (EN: female actor) in German.

Based on linguistic gender, languages can be classified into grammatical gender, notional gender, and genderless languages (Stahlberg et al., 2007; McConnell-Ginet, 2013). The first, e.g., German, have grammatical, lexical, and referential gender. Consequently, they are highly inflected and mark gender very often. The second, e.g., English, usually have lexical and referential gender. Therefore, they are sometimes marked for gender. Genderless languages, e.g., Turkish, have only lexical gender and rarely carry gender inflection.

3.2 Gender-Fair Language

A common linguistic phenomenon in grammatical and notional gender languages is the masculine generic, i.e., the use of masculine forms to refer to both men and people in general. This specific language practice has drawn the attention of feminists who, in the field of linguistics and translation studies amongst others, have analysed how patriarchal language is used to oppress women and consequently advocated for GFL (Simon, 1996; Kramer, 2016).

As in Sczesny et al. (2016), we use "genderfair" to subsume both gender-neutral and genderinclusive approaches. The former avoid gender marking by using passive constructions, indefinite pronouns, and gender-neutral nouns. The latter make all genders visible through typographical characters (e.g., gender star (*) in German), symbols (e.g., schwa (ə) in Italian), and neomorphemes (e.g., "e" in Spanish).

Furthermore, the relationship between linguistic gender and gender identity is not one-to-one (Cao and Daumé III, 2020). In many European languages, only the masculine and feminine gender are used in reference to people (Deutscher, 2010). Therefore, non-binary representation requires breaking traditional grammar rules and new GFL strategies have been proposed in the last few years (Lardelli and Gromann, 2023b; López, 2019).

3.3 Gender-Fair German

Lardelli and Gromann (2023b) provide an overview of GFL strategies in German, which we summarise here due to space constraints. The researchers identify four main approaches:

- 1. Gender-Neutral Rewording: strategies to avoid gender marking, e.g., the use of participial forms, passive constructions, and gender-neutral terms.
- Gender-Inclusive Characters: e.g., gender star (*) is used to separate masculine and feminine forms of words as in "*der*die Autor*in*" (EN: the author), usually to avoid masculine generics.
- 3. Gender-Neutral Characters: e.g., "x" is used to replace gender suffixes (e.g., "*dix Autorx*") in contexts where gender is unknown or irrelevant to the context of the conversation.
- 4. Gender-Fair Neosystems: for instance, "ens" is used as a morpheme to create new articles (e.g., "dens"), pronouns (e.g., "dens"), and nouns (e.g. "Authorens"). These strategies are usually devised by non-binary people as a means to be included in language.

4 Method

The proposed method is inspired by research on gender bias in MT (Savoldi et al., 2021), combining the creation of a dataset – containing common nouns referring to people –, its automatic translation with three commercial MT systems (DeepL², Google Translate³, and Microsoft Bing⁴), and their output analysis.

Since there is currently no standard for GFL and, as found in studies on translation and postediting (Lardelli and Gromann, 2023a; Lardelli, 2023), its implementation varies greatly. Therefore, we started from the "Genderwörterbuch"⁵. This is a community-created German vocabulary where users add gender-fair, usually neutral, alternatives to terms commonly marked for gender. The terms contained in the vocabulary are usually nouns referring to people, but the resource also contains expressions with pronouns (e.g. "der eine oder der andere", EN: "one or the other") and short phrases (e.g. "Das Angebot richtet sich an Anfänger und Fortgeschrittene", EN: "the offer is aimed at beginners and advanced students"). We focused on nouns referring to people and used the vocabulary to select suitable terms for our study.

²https://www.deepl.com/translator

³https://translate.google.com

⁴https://www.bing.com/translator

⁵https://geschicktgendern.de/

We first randomly selected 128 lemmas. We filtered out those that are already neutral, e.g., "*Star*", whose grammatical gender is masculine, but it is used for all genders and has no other inflected variants. We also removed polysemous terms, e.g., "*aid*", to facilitate translation into English. The final size of our datasets is 115 lemmas.

After translating each of the sampled lemmas into English, one of the authors also enriched the dataset with the English plural form. An extract from our dataset is shown in Table 1. To date, most research on gender bias in MT focuses on the translation of professions only (Prates et al., 2020). Conversely, our dataset includes common nouns too (e.g., "donor").

ish Singular English Plura
eader the readers

Table 1: Examples for entries in our dataset

MT systems are increasingly used as bilingual dictionaries (Cotelli Kureth et al., 2023). We were interested in widely used commercial MT systems, such as DeepL, Google Translate, and Microsoft Bing, as possible English-to-German gender-fair dictionaries. Although to a different extent, these tools offer dictionary functions and/or propose some alternatives for each translation. Therefore, between December 2023 and April 2024, we back-translated the English terms included in our dataset into German, both in the singular and plural, via the User Interface (UI) of each of the selected MT systems.

First, we conducted an exploratory study by translating the first 20 terms in our dataset. For each English term, we pasted the translations along with all the alternatives proposed by the MT systems into an Excel sheet. We initially translated terms along with the definite article. This is important because in German some nouns have only a gender form but require masculine or feminine articles, e.g. "*der/die Bedienstete*" (EN: "the employee"). However, we noted that Google Translate and Microsoft Bing provide only one translation when doing so. Therefore, for these two systems, we re-translated the terms by omitting the definite article.

We discarded Google Translate and Microsoft Bing based on the initial findings (§5.1): the alternative translations proposed by both systems are usually in the masculine form. Conversely, DeepL's outputs contain gender-fair alternatives considerably more often. Hence, we translated the whole dataset only with DeepL. One of the authors - an expert in GFL and translation - evaluated the translations. In the first step, a quantitative analysis was conducted: the author annotated the overt gender of the translation, i.e., masculine (M), feminine (F), gender-inclusive (GI), and gender-neutral (GN). Wrong translations (W) were also annotated. In Table 2, an example of the annotation for the translations of the English term "the colleagues" is reported. In this context, wrong refers to semantics (i.e., the German term has a different meaning than the English source), and grammar (e.g., wrong number or no agreement between article and noun). In the second step, the focus was on the type of gender-fair language strategy used by the system. Finally, another author whose first language is German replicated the analysis in order to validate the results. The percentage agreement between the two raters was calculated. Differences in the annotation were discussed to reach a consensus.

5 Results

First, we summarise the results of the exploratory study (§5.1), then we provide an overview of the results obtained with DeepL and focus on the overt gender of the machine-translated outputs in German (§5.2). We subsequently analyze which gender-fair, i.e., inclusive and neutral, strategies are found in the singular (§5.2.1) and plural (§5.2.2).

5.1 Findings from the Exploratory Study

Table 3 presents an overview of the results of the exploratory study with Google Translate (GT), Microsoft Bing (MB), and DeepL. The table focuses on the overt gender of the machine translations for the first twenty seed words in our dataset: M indicates masculine, G feminine, GI gender-inclusive, GN gender-neutral, W wrong translation, T the sum of all translations including the alternatives proposed by the MT systems.

First, when translating single nouns without an article, both Google Translate and Microsoft Bing usually, but not always, provide gender-specific translations for the masculine and the feminine (Kuczmarski, 2018; Translator, 2023). This feature, however, does not seem to be available for

Source Term	Translations	Overt Gender			GFL Strategy		
		М	F	GI	GN	W	
The colleagues	Die Kollegen Die Kolleginnen und Kollegen Die Mitarbeiter Die Kollegen und Kolleginnen	x x		x x			double form

Table 2: Annotation example for the German translations of "the colleagues"

		М	F	GI	GN	W	Т
GT	Singular	32	13	0	2	5	52
	Plural	19	6	0	1	0	26
MB	Singular	25	1	0	3	3	32
	Plural	29	0	2	6	0	37
DeepL	Singular	49	17	2	1	7	76
	Plural	43	2	14	9	4	72



the German language in Bing Translator and it is not available at all in DeepL.

Second, we noted that all systems are systematically biased towards the masculine forms, which represent more than half of all translations. While all systems also provide possible alternative translations, Google Translate and Microsoft Bing generally default to the masculine. For instance, the first system outputs "Siedler" (EN: settler, masculine) and "Siedlerin" (EN: settler, feminine) as a gender-specific translation for the English noun "settler" but it also suggests two synonyms in the masculine form, i.e., "der Ansiedler" and "der Kolonist". Both gender-inclusive (0-2%) and gender-neutral (1-6%) are rare in Google Translate's and Microsoft Bings's outputs and occur more often in DeepL, i.e. up to respectively 14% and 9% in the plural.

Finally, the number of alternative translations provided by Google Translate considerably decreases in the plural. DeepL is the system that provides the highest amount of alternatives, e.g. 76 translations against 52 (Google Translate) and 32 (Microsoft Bing) in the singular. Based on these preliminary findings, we decided to continue the study by translating our entire dataset with DeepL only.

5.2 General Findings with DeepL

Table 4 summarises the results for the translation of both singular and plural words contained in our dataset. The table focuses on the overt gender of

the DeepL outputs. Note that the total of transla-
tions does not amount to 115 because we analysed
all alternatives suggested by the system.

		М	F	GI	GN	W
Singular	$_{\%}^{N}$	285 66	79 18	6 1	17 4	44 10
Plural	N %	279 67	9 2	45 11	62 15	20 5

Table 4: Translation results with DeepL: overt gender of the singular and plural terms

The percentage agreement between the two raters was 96% in the overt gender annotation. The differences were discussed. In most cases, one of the two raters made a mistake in the gender annotation. For instance, the English term "mountaineer" was translated amongst others as "Bergbewohner", which indicates a person who lives in a mountain area. The second rater annotated this alternative as semantically wrong, which is not. An interesting source of disagreement in the annotation was the use of neutral forms in the plural. For example, the term "prosecutors" was translated as "Staatsanwaltschaft" and its plural form "Staatsanwaltschaften" was suggested as well (EN: " office(s) of the Public Prosecutor). One rater considered "Staatsanwaltschaft" as a wrong translation. However, the term is a collective noun and it could therefore be argued that it may be used for one or more referents, in this case one or more prosecutor(s). It is not always possible or desirable to decide if the translation of a single term is correct without analysing its use in a broader context, which represents a limitation of the present study (see §6) and, more generally, of the use of MT systems as dictionaries.

DeepL is strongly biased towards the masculine gender, which appears in about 67% of the translations both in the plural and in the singular. Feminine translations occur less frequently, i.e., in 18% of the outputs for the singular. This value drops to 2% in the plural. The number of genderinclusive and neutral forms is very low in the singular, 1% and 4% respectively. This value, however, considerably increases in the plural to 11% and 15%. Since words in isolation were translated and DeepL lacked contextual information for the selection of an appropriate term, semantically and/or grammatically wrong translations make up 10% and 5% of the outputs respectively, e.g., "*Depotbank*" (EN: "custodian bank") as a translation for the person term "custodian".

5.2.1 Gender-Fair Forms in the Singular

In the singular, two gender-inclusive strategies were found, as shown in Table 5. The first is the use of masculine and feminine forms separated by a slash, e.g., "*der Sportler/die Sportlerin*" (EN:"sportsperson"), which occurred four times in the translated dataset. The second is using a slash to combine the masculine and feminine definite article and a participial form for the noun, e.g., "*die/der Vorsitzende*" (EN: "the chairperson"), which occurred twice. Note that these approaches are not inclusive of non-binary people: the use of gender star, e.g. "*der*die Sportler*in*" and "*der*die Vorsitzende*", would be the most common gender-fair alternative nowadays to indicate that there are more than two genders.

Gender-Inclusive		
masculine/feminine	4	
article with / + participial form	2	

Table 5: Gender-inclusive forms in the singular

Gender-neutral forms were slightly more frequent and three main strategies were found, as shown in Table 6. The first was the use of abstract, usually collective, nouns, e.g., "die Projektleitung" (EN: "the project leadership" instead of "the project leader"), which occurred nine times. The second was the use of a noun that is already gender-neutral, e.g., "der Neuling" (EN: "the beginner"), which occurred four times. In these cases, the German term has the masculine grammatical gender, but it is commonly used for all genders. The third strategy found in the translation outputs was the use of the term "Person" (EN: "person") or "Mensch" (EN: "human") to build gender-neutral compounds, e.g., "die Geschäftsperson" (EN: "the businessperson"). This strategy too occurred four times.

Gender-Neutral	N
Abstract Nouns	9
Neutral Nouns	4
Expressions with Person	4

Table 6: Gender-neutral forms in the singular

5.2.2 Gender-Fair Forms in the Plural

Gender-fair outputs were more frequent in the plural. Three gender-inclusive strategies were The first was the BinnenI, e.g., "die found. MinisterInnen" (EN: the ministers), and occurred twenty-two times in the translated dataset. The BinnenI is similar to gender-inclusive characters, such as gender star (*), which are now more common in German (Körner et al., 2022). The second strategy was the use of double forms, i.e., the masculine and feminine gender are mentioned as in "Die Koordinatorinnen und Koordinatoren" (EN: the coordinators). It occurred twenty times in the translations. The last gender-inclusive strategy is the use of a slash (/) as an inclusive character, e.g., "die Blogger/innen" (EN: the bloggers).

Gender-Inclusive	Ν
BinnenI Double Forms	22 19
Slash	4

Table 7: Gender-inclusive forms in the plural

As concerns gender-neutral language, the same strategies as in (§5.2.1) were found with the addition of participial forms, which were the most frequent with twenty-two occurrences. German verbs can be nominalized by using participial forms, e.g., "*die Abgefraten*" (EN: "the respondents") as found in the analysed translations. While the articles and/or the noun declension is gender-specific in the singular, participial forms are gender-neutral in the plural – hence, they are a quite common strategy to avoid the generic masculine.

Gender-Neutral	N
Participial Forms	22
Abstract Nouns	20
Compounds with People	11
Neutral Nouns	7

Table 8: Gender-neutral forms in the plural

Abstract nouns occurred frequently too, i.e., twenty times. For instance, "the prosecutors" was translated into "*die Staatsanwaltschaft*" which,

back-translated into English, indicates the office of the Public Prosecutor. Expressions with the term "Leute" (EN: people) were found eleven times in the translated dataset, e.g., "die Bauleute" for "the builders". Finally, neutral nouns were the least common gender-neutral strategy with seven occurrences, e.g., "Die Grundschulkinder" for "the primary school pupils".

6 Discussion

In the present contribution, we were interested in using three commercial MT models as English-to-German gender-fair dictionaries. Unsurprisingly, the results suggest that commercial MT models are still systematically biased towards masculine forms when translating from a notional gender, English, into a grammatical gender language, German.

In our exploratory study, we find that Google Translate usually generates gender-specific translations, but only in the masculine and feminine genders. This feature is not available in Microsoft Bing and DeepL. Both Google Translate and Microsoft Bing provide alternative translations, usually synonyms, in their dictionary interface. These alternatives are, however, generally masculine. Conversely, DeepL offers numerous alternatives that are also gender-fair.

The main study confirms that DeepL is heavily biased towards the masculine with about 67% of outputs having this overt gender both in the singular and plural. A great difference emerges between singular and plural: the number of feminine translations significantly decreases from 18% to 2%. Conversely, the number of gender-inclusive and neutral translations increases from 1% to 11% and 4% to 15% respectively. There are at least two main reasons for this phenomenon.

First, some nouns are gender-specific in the singular form, but not in the plural. For instance, the term "traveller" has different declensions. Without articles, the masculine form is "*Reisender*" whilst the feminine is "*Reisende*". In the plural, there is one form only, i.e., "*Reisende*". Second, Germanspeaking countries have a relatively strong feminist tradition and gender-fair language policies to avoid masculine generics were introduced several years ago (Sczesny et al., 2016). GFL is now quite common in, e.g., administrative texts where different gender-fair strategies, such as participial forms and gender star (*), are increasingly used for the declension of plural terms (Körner et al., 2022).

In the translations generated by DeepL, several gender-neutral and inclusive forms were found. Gender-neutral forms included compounds with neutral terms such as person (e.g. "der spekulative Mensch", the speculating person), abstract nouns (e.g., "die Staatsanwaltschaft", the prosecutor's office), and participial forms (e.g. "die Vorsitzenden", the presidents). Gender-inclusive forms were also found, including BinnenI (e.g. "die KoordinatorInnen", the coordinators), slash (e.g. "die Mitbürger/innen", the fellow citizens), and double forms (e.g. "die Betreuerinnen und Betreuer", the counsellors).

Though DeepL seems to be receptive of genderfair forms that, probably quite rarely, occur in the training data, gender-inclusive strategies found in the outputs are usually outdated mostly because they are considered inclusive of binary genders only. For instance, BinnenI was once commonly used and studies about its effect on mental representations date back more than twenty years ago (Stahlberg and Sczesny, 2001). This strategy has nowadays been replaced by the use of genderinclusive characters such as gender star (*) (Körner et al., 2022).

The findings of this study show how current commercial MT systems cannot keep up with linguistic change. The field of gender-fair language is constantly evolving and there is yet no onesize-fits-all solution to issues of gender representation (Gromann et al., 2023). In fact, the selection of a gender-fair language strategy is highly context-dependent (Lardelli and Gromann, 2023a; Gromann et al., 2023; Lardelli, 2023). For this reason, future research endeavours on MT debiasing should be the result of interdisciplinary efforts, involving computational linguistics, sociolinguistics, and translation studies amongst others.

To conclude, we discuss three major limitations of the present study. The first concerns the nonreplicability of the results. Unfortunately, we don't have insights into the system used by Google, DeepL, and Microsoft Bing. Updates and/or retraining may lead to changes in the outputs over time. The second involves the analysis of genderneutral and inclusive strategies. As already explained, there is no standard for GFL and creativity is often required. The soundness of genderfair solutions might hence be judged differently among experts and, more importantly, depends on the broader context, i.e. at least the text, in which such solutions are used. Finally, we considered one language pair only because of the high amount of manual work involved in the translation of our dataset and its analysis.

7 Conclusion

In this study on gender bias in MT, we investigated the use of three commercial systems as Englishto-German gender-fair dictionaries. Drawing on a community-created gender-fair dictionary, we developed a dataset including 115 gender-specific terms for which gender-fair alternatives in German were proposed. We then provided the terms with a translation in English both in the singular and plural form. Subsequently, we conducted a brief exploratory study with DeepL, Google Translate, and Microsoft Bing by back-translating into German the first 20 seed nouns contained in our dataset.

The results from this exploratory study show that all systems default to male forms. Moreover, Google Translate usually provides gender-specific translations in the masculine and feminine, and Microsoft Bing offers synonymous translations in the masculine form only. For these reasons, we further conduct our study with DeepL which usually generates three to four translations per seed word.

In a nutshell, our findings seem to suggest that GFL is starting to appear in DeepL outputs, probably due to the relatively widespread GFL use in German-speaking countries. Nevertheless, DeepL still generates gender-fair forms inconsistently and far more often in the plural. Finally, the genderinclusive forms found in the machine translations are generally outdated and exclusive of genders beyond the binary – an issue still under-researched within translation studies and computational linguistics with few exceptions (Saunders et al., 2020; Lauscher et al., 2023; Lardelli and Gromann, 2023a).

Acknowledgements

This work is part of the GeFMT project (13223) sponsored by the European Association for Machine Translation (EAMT) under the EAMT Sponsorship of Activities 2023.

References

- [Atkins1996] Atkins, Beryl T Sue. 1996. Bilingual dictionaries: Past, present and future. EURALEX'96 Proceedings. Göteborg: Department of Swedish, Göteborg University, pages 515–546.
- [Attanasio et al.2023] Attanasio, Giuseppe, Flor Miriam Plaza del Arco, Debora Nozza, and Anne Lauscher. 2023. A tale of pronouns: Interpretability informs gender bias mitigation for fairer instruction-tuned machine translation. In Bouamor, Houda, Juan Pino, and Kalika Bali, editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 3996–4014, Singapore, December. Association for Computational Linguistics.
- [Barocas et al.2017] Barocas, Solon, Kate Crawford, Aaron Shapiro, and Hanna Wallach. 2017. The problem with bias: Allocative versus representational harms in machine learning. In 9th Annual conference of the special interest group for computing, information and society. Philadelphia, PA, USA.
- [Cao and Daumé III2020] Cao, Yang Trista and Hal Daumé III. 2020. Toward gender-inclusive coreference resolution. In Jurafsky, Dan, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 4568– 4595, Online, July. Association for Computational Linguistics.
- [Ciora et al.2021] Ciora, Chloe, Nur Iren, and Malihe Alikhani. 2021. Examining covert gender bias: A case study in Turkish and English machine translation models. In Belz, Anya, Angela Fan, Ehud Reiter, and Yaji Sripada, editors, *Proceedings of the 14th International Conference on Natural Language Generation*, pages 55–63, Aberdeen, Scotland, UK, August. Association for Computational Linguistics.
- [Corbett1991] Corbett, Greville G. 1991. *Gender*. Cambridge University Press.
- [Cotelli Kureth et al.2023] Cotelli Kureth, Sara, Alice Delorme Benites, Mara Haller, Hasti Noghrechi, and Elizabeth Steele. 2023. "i looked it up in deepl": Machine translation and digital tools in the language classroom. Studie e Ricerche: Human Translation and Natural Language Processing Towards a New Consensus?, 35:81–96.
- [Daems2023] Daems, Joke. 2023. Gender-inclusive translation for a gender-inclusive sport: strategies and translator perceptions at the international quadball association. In Vanmassenhove, Eva, Beatrice Savoldi, Luisa Bentivogli, Joke Daems, and Janiça Hackenbuchner, editors, *Proceedings of the First Workshop on Gender-Inclusive Translation Technologies*, pages 37–47, Tampere, Finland, June. European Association for Machine Translation.
- [Deutscher2010] Deutscher, Guy. 2010. Through the Language Glass: Why the World Looks Different in Other Languages. Metropolitan Books.

- [Dev et al.2021] Dev, Sunipa, Masoud Monajatipoor, Anaelia Ovalle, Arjun Subramonian, Jeff Phillips, and Kai-Wei Chang. 2021. Harms of gender exclusivity and challenges in non-binary representation in language technologies. In Moens, Marie-Francine, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1968–1994, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.
- [Duan et al.2020] Duan, Xiangyu, Baijun Ji, Hao Jia, Min Tan, Min Zhang, Boxing Chen, Weihua Luo, and Yue Zhang. 2020. Bilingual dictionary based neural machine translation without using parallel sentences. In Jurafsky, Dan, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings* of the 58th Annual Meeting of the Association for Computational Linguistics, pages 1570–1579, Online, July. Association for Computational Linguistics.
- [Gromann et al.2023] Gromann, Dagmar, Manuel Sabrina Burtscher, Lardelli. Katta Spiel, Lukas Daniel Klausner, Arthur Mettinger, Igor Miladinovic, Sigrid Schefer-Wenzl, Daniela Duh, and Katharina Bühn. 2023. Participatory research as a path to community-informed, gender-fair machine translation. In Vanmassenhove, Eva, Beatrice Savoldi, Luisa Bentivogli, Joke Daems, and Janica Hackenbuchner, editors, Proceedings of the First Workshop on Gender-Inclusive Translation Technologies, pages 49-59, Tampere, Finland, June. European Association for Machine Translation.
- [Körner et al.2022] Körner, Anita, Bleen Abraham, Ralf Rummer, and Fritz Strack. 2022. Gender representations elicited by the gender star form. *Journal of Language and Social Psychology*, 41(5):553–571.
- [Kramer2016] Kramer, Elise. 2016. Feminist linguistics and linguistic feminisms. In Lewin, Ellen and Leni M. Silverstein, editors, *Mapping Feminist An*thropology in the Twenty-first Century, pages 65–83. Rutgers University Press.
- [Kuczmarski2018] Kuczmarski, James. 2018. Reducing gender bias in google translate.
- [Lardelli and Gromann2023a] Lardelli, Manuel and Dagmar Gromann. 2023a. Gender-fair post-editing: A case study beyond the binary. In Nurminen, Mary, Judith Brenner, Maarit Koponen, Sirkku Latomaa, Mikhail Mikhailov, Frederike Schierl, Tharindu Ranasinghe, Eva Vanmassenhove, Sergi Alvarez Vidal, Nora Aranberri, Mara Nunziatini, Carla Parra Escartín, Mikel Forcada, Maja Popovic, Carolina Scarton, and Helena Moniz, editors, *Proceedings* of the 24th Annual Conference of the European Association for Machine Translation, pages 251–260, Tampere, Finland, June. European Association for Machine Translation.

- [Lardelli and Gromann2023b] Lardelli, Manuel and Dagmar Gromann. 2023b. Translating non-binary coming-out reports: Gender-fair language strategies and use in news articles. *The Journal of Specialised Translation*, 40:213–240.
- [Lardelli2023] Lardelli, Manuel. 2023. Gender-fair translation: a case study beyond the binary. *Perspectives*, pages 1–17.
- [Lauscher et al.2023] Lauscher, Anne, Debora Nozza, Ehm Miltersen, Archie Crowley, and Dirk Hovy. 2023. What about "em"? how commercial machine translation fails to handle (neo-)pronouns. In Rogers, Anna, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 377–392, Toronto, Canada, July. Association for Computational Linguistics.
- [López2019] López, Ártemis. 2019. Tú, yo, elle y el lenguaje no binario. *La Linterna del Traductor*, 19.
- [McConnell-Ginet2013] McConnell-Ginet, Sally. 2013. Gender and its relation to sex: The myth of 'natural'gender. In Corbett, Greville G, editor, *The Expression of Gender*, pages 3–38. De Gruyter Mouton.
- [Nagata et al.2001] Nagata, Masaaki, Teruka Saito, and Kenji Suzuki. 2001. Using the web as a bilingual dictionary. In Proceedings of the ACL 2001 Workshop on Data-Driven Methods in Machine Translation.
- [Paolucci et al.2023] Paolucci, Angela Balducci, Manuel Lardelli, and Dagmar Gromann. 2023. Gender-fair language in translation: A case study. In Vanmassenhove, Eva, Beatrice Savoldi, Luisa Bentivogli, Joke Daems, and Janiça Hackenbuchner, editors, *Proceedings of the First Workshop* on Gender-Inclusive Translation Technologies, pages 13–23, Tampere, Finland, June. European Association for Machine Translation.
- [Piergentili et al.2023] 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.
- [Prates et al.2020] Prates, Marcelo OR, Pedro H Avelar, and Luís C Lamb. 2020. Assessing gender bias in machine translation: a case study with google translate. *Neural Computing and Applications*, 32:6363– 6381.
- [Saunders and Olsen2023] Saunders, Danielle and Katrina Olsen. 2023. Gender, names and other mysteries: Towards the ambiguous for gender-inclusive

translation. In Vanmassenhove, Eva, Beatrice Savoldi, Luisa Bentivogli, Joke Daems, and Janiça Hackenbuchner, editors, *Proceedings of the First Workshop on Gender-Inclusive Translation Technologies*, pages 85–93, Tampere, Finland, June. European Association for Machine Translation.

- [Saunders et al.2020] Saunders, Danielle, Rosie Sallis, and Bill Byrne. 2020. Neural machine translation doesn't translate gender coreference right unless you make it. In Costa-jussà, Marta R., Christian Hardmeier, Will Radford, and Kellie Webster, editors, *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*, pages 35– 43, Barcelona, Spain (Online), December. Association for Computational Linguistics.
- [Savoldi et al.2021] 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.
- [Savoldi et al.2023] Savoldi, Beatrice, Marco Gaido, Matteo Negri, and Luisa Bentivogli. 2023. Test suites task: Evaluation of gender fairness in MT with MuST-SHE and INES. In Koehn, Philipp, Barry Haddow, Tom Kocmi, and Christof Monz, editors, *Proceedings of the Eighth Conference on Machine Translation*, pages 252–262, Singapore, December. Association for Computational Linguistics.
- [Sczesny et al.2016] Sczesny, Sabine, Magda Formanowicz, and Franziska Moser. 2016. Can gender-fair language reduce gender stereotyping and discrimination? *Frontiers in psychology*, 7.
- [Simon1996] Simon, Sherry. 1996. Gender in Translation: Cultural Identity and the Politics of Transmission. Routledge.
- [Stahlberg and Sczesny2001] Stahlberg, Dagmar and Sabine Sczesny. 2001. Effekte des generischen maskulinums und alternativer sprachformen auf den gedanklichen einbezug von frauen. *Psychologische Rundschau*, 52(3):131–140.
- [Stahlberg et al.2007] Stahlberg, Dagmar, Friederike Braun, Lisa Irmen, and Sabine Sczesny. 2007. Representation of the sexes in language. *Social Communication*, pages 163–187.
- [Stanovsky et al.2019] Stanovsky, Gabriel, Noah A. Smith, and Luke Zettlemoyer. 2019. Evaluating gender bias in machine translation. In Korhonen, Anna, David Traum, and Lluís Màrquez, editors, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1679– 1684, Florence, Italy, July. Association for Computational Linguistics.
- [Thompson1987] Thompson, Geoff. 1987. Using bilingual dictionaries. *ELT journal*, 41(4):282–286.
- [Translator2023] Translator, Microsoft. 2023. Bing's gendered translations tackle bias in translation.

[Zhang et al.2021] Zhang, Tong, Long Zhang, Wei Ye, Bo Li, Jinan Sun, Xiaoyu Zhu, Wen Zhao, and Shikun Zhang. 2021. Point, disambiguate and copy: Incorporating bilingual dictionaries for neural machine translation. In Zong, Chengqing, Fei Xia, Wenjie Li, and Roberto Navigli, editors, Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 3970–3979, Online, August. Association for Computational Linguistics.