

Gender-Fair Post-Editing: A Case Study Beyond the Binary

Manuel Lardelli

University of Graz, Austria
manuel.lardelli@uni-graz.at

Dagmar Gromann

University of Vienna, Austria
dagmar.gromann@gmail.com

Abstract

Machine Translation (MT) models are well-known to suffer from gender bias, especially for gender beyond a binary conception. Due to the multiplicity of language-specific strategies for gender representation beyond the binary, debiasing MT is extremely challenging. As an alternative, we propose a case study on gender-fair post-editing. In this study, six professional translators each post-edited three English to German machine translations. For each translation, participants were instructed to use a different gender-fair language approach, that is, gender-neutral rewording, gender-inclusive characters, and a neosystem. The focus of this study is not on translation quality but rather on the ease of integrating gender-fair language into the post-editing process. Findings from non-participant observation and interviews show clear differences in temporal and cognitive effort between participants and GFL approach as well as in the success of using gender-fair language.

1 Introduction

Gender bias in Machine Translation (MT) has been studied from different angles and a lot of work has been published for debiasing MT, however, only recently from the perspective of bias beyond a binary conception of gender. Most approaches (Piergentili et al., 2023; Savoldi et al., 2021) discuss it from a theoretical perspective. One exception is Saunders et al. (2020), who proposed

a gender-tagging approach to translate from inflected to gender-neutral language with moderate success. Gender-fair debiasing MT models is challenging, since there is a lack of datasets and even human translators find it difficult to select and correctly apply gender-fair language strategies. As a first step, we propose a case study to investigate gender-fair language within the context of post-editing.

Post-editing approaches generally focus on speed, productivity, cognitive load, and quality (Jia et al., 2019; Toral et al., 2018). To the best of our knowledge, this is the first post-editing study to focus on gender-fair language. Non-binary individuals have become increasingly visible, such as in TV series like *One Day at a Time* or *Sex Education*, and the number and type of strategies to linguistically represent them differs considerably across languages. Gender-fair language (GFL) subsumes gender-neutral, that is, omitting any gender references, and gender-inclusive language, that is, linguistically including all genders. With grammatical differences and a multitude of strategies across languages, gender-fair translation is challenging for machines and humans. Nevertheless, human and machine translators can act as ambassadors for gender equality beyond the binary by using gender-fair language.

In this case study, we chose post-editing over translation from scratch to focus on the temporal and cognitive effort required to revise an existing translation in terms of gender references. Six professional translators post-edited three machine translations from English to German, which contain references to non-binary individuals. While for English singular they have become predominant, in German there are many different strategies. Participants were instructed to apply one

specific approach per text, i.e., gender-neutral rewording, gender-inclusive characters, and neosystems. Screen recordings allow to measure the temporal effort. Post-experiment interviews provide insights into the cognitive load felt by participants depending on the strategy. Finally, analyses of post-edited translations reveal the level of difficulty of the task.

The less familiar a person is with gender-fair language, the more difficult it is to correctly detect and revise gender references. Piergentili et al. (2023) argue to only use gender-neutral strategies in MT and only utilise gender-inclusive forms where necessary not to lose important information. In our experiment, participants equally expressed a clear preference for a combination of these two strategies over neosystems. With results from this case study, we contribute to guidelines for integrating gender-fair language into the translation workflow. We show which gender-specific constituents are particularly challenging in the source and target text, which potentially provides inspirations to MT debiasing.

2 Related Work

Research on gender-fair language in human and machine translation is still in its infancy and very few publications address the topic from either a translation studies or MT perspective (Lardelli and Gromann, 2023). One exception is Burtscher et al. (2022), who conducted a participatory workshop on both gender-fair human and machine translation, bringing together different stakeholders and working in a multidisciplinary team of experts from translation studies, MT, gender studies, and human-computer interactions. The results of this workshop highlight that strategy selection is highly dependent on different criteria, e.g., context, target audience, scope of the text, thus there being no “one-size-fits-all” solution (Burtscher et al., 2022). Since, to the best of our knowledge, this is the first gender-fair post-editing study, we will introduce gender-fair translation from the perspective of translation studies and MT.

López (2019; 2022) and Attig (2022) analysed, among others, the dubbed and the subtitled versions of the Netflix series *One Day at a Time* in Spanish and French. Both found that translation strategies varied based not only on the version, i.e., subbed or subtitled, but also on the language variety, i.e., European and Latin American Spanish.

In three of the four Spanish versions, a non-binary character is addressed with female forms, and/or a literal translation of English singular they. Only in the European Spanish dubbing, the non-binary neopronoun “elle” is utilised. In the French subtitles for the series, the French indefinite pronoun “on” (one/we) is used.

Misieć (2020) analysed the Polish translation of three different English language TV series that feature non-binary characters and found a complete omission of their gender identity. In Croatian articles on Sam Smith’s coming out as non-binary and movie translations (Šincek, 2020), the masculine plural pronoun was found as a frequent strategy, which is an instance of misgendering.

In a first attempt to debias NMT models beyond the binary, Saunders et al. (2020) extended a gender-balanced corpus (Saunders and Byrne, 2020) by gender-neutral sentences with placeholders for gender inflections in German and French for training. For testing, they produced a gender-neutral version of the WinoMT dataset (Stanovsky et al., 2019) and found a low overall accuracy and a tendency to over-generalise the use of exclusively gender-neutral language, even if the source text was clearly gendered.

From a theoretical perspective, Piergentili et al. (2023) propagate gender-neutral strategies for machine translation and propose gender-neutral constraint-based algorithms at training time, wider contexts than sentence-level, and injecting external knowledge as possible approaches. Furthermore, they highlight the difficulty of identifying gender references to be changed, e.g. the *mother* in *motherboard* might not be a candidate.

3 Preliminaries

In order to provide a basis for gender-fair post-editing, this section briefly introduces gender-fair language strategies for English and German. As a notional gender language, English generally requires gender-specification in third-person singular pronouns (e.g. he/she/it) and in some specific nouns, usually in reference to kinship (e.g. mother/father) or professions (e.g. chairman/chairwoman) (Corbett, 1991; Stahlberg et al., 2007; McConnell-Ginet, 2013). To achieve gender-fair English, singular they and gender-neutral nouns, e.g. *chairperson*, are often used to address non-binary people (APA Style, 2019). Other languages, such as German and Italian, are

grammatical gender languages and require extensive gender marking in pronouns, nouns and also in adjectives or participles (Corbett, 1991; Stahlberg et al., 2007).

In German, for example, four different approaches can be identified, i.e., (i) gender-neutral rewording; (ii) gender-inclusive characters; (iii) gender-neutral characters and forms; and (iv) neosystems. In (i), sentences are structured in order to avoid gender-specification using, e.g. gender-neutral words such as person, indefinite pronouns, passive constructions and participial forms. In (ii), characters such as gender star (*) are used to separate male forms from female endings, e.g. *Leser*in* (reader) to include all genders. In (iii), similar characters or new endings like “x” in *Lesx* (reader) are used to question the gender binary. In (iv), a fourth gender in addition to masculine, feminine and neuter is introduced in the German language as in the case of *Lesernin* (reader). Several comprehensive overviews of gender-fair language in German are available (Hornscheidt and Sammla, 2021; De Sylvain and Balzer, 2008; En et al., 2021).

The universal acceptance of gender-fair language can and has been debated. For instance, Vergoossen et al. (2020) provide four dimensions of resistance against the introduction of the gender-fair pronoun *hen* in Swedish, including distraction in communication, defending the status quo, and cisgenderism. However, linguistic change that does not come about naturally has always met initial resistance, but in the end facilitates social change towards gender equality (Sczesny et al., 2016). This, in turn, reduces linguistic and systematic identity invalidation and permits access to public spaces, e.g. restrooms, and services, e.g. personal identity cards. Translators and machine translation can act as ambassadors for such change.

4 Method

The proposed method, inspired by Translation Process Research (TPR) (Jakobsen, 2017) and Albl-Mikasa et al. (2017), combines non-participant observation, screen-recordings, retrospective interviews, and target text annotation. Six professional translators with at least two years of practical experience were recruited. Prior to the study, participants received instructions on the tasks, post-editing guidelines by the Translation Automation

User Society (TAUS)¹, and a handout on various strategies to gender-fair language in order to prepare for their participation.

As shown in Table 1, participants received three texts of approx. 150 words on three different English language TV series, namely *Sex Education*, *Grey’s Anatomy*, and *Sort Of*. The texts discussed non-binary actors joining such series and playing non-binary characters². They were retrieved from TV news websites and translated into German with DeepL in July 2022. Translators received a text file with a table containing the English source text as well as the German machine translation. Each text was to be manually post-edited adopting a different approach to gender-fair language, that is, (i) gender-neutral rewording, (ii) gender-inclusive characters, and (iii) neosystems. For each approach, participants could freely select specific strategies from the provided handout, e.g. gender star (*) or underscore (–) amongst others for (ii).

To ensure comparability of estimated PE times per text, readability scores were computed using the Flesch-Kincaid readability test (Kincaid et al., 1975). It takes into account the number of words and their length, but ignores semantics. The selected texts contained references both to non-binary individuals as well as mixed-gender groups in English. German was selected as a target language being a grammatical gender language which needs extensive gender marking when compared to English. For the translation analysis, respectively 9, 12, and 10 gendered phrases were identified. These phrases were composed of different word classes, such as nouns, adjectives, articles and different types of pronouns, mostly singular they.

The study was conducted online since it aimed for a most authentic and unintrusive experimental setting. Translators could work in their familiar environment and were instructed to work under usual conditions. Nevertheless, they were required to use one screen only in order for the whole post-editing process to be recorded. During the process, a video conference was open in the background, on which the shared screen was recorded. Subsequently, they were interviewed about their impressions, strategies, and which aspects of the study were particularly challenging. The interviews were conducted in German, transcribed ac-

¹<https://info.taus.net/mt-post-editing-guidelines>

²The instructions and texts of this case study are available at <https://doi.org/10.5281/zenodo.7898328>

Text No.	TV Series	Instructed Gender-Fair Approach	Word Count	Gendered Phrases
1	Sex Education	Gender-Neutral Rewording	152	9
2	Grey's Anatomy	Gender-Inclusive Characters	151	12
3	Sort Of	Neosystems	163	10

Table 1: Details on post-editing materials

ording to Dresing and Pehl's (2018) semantic transcription rules and then analysed by means of qualitative content analysis (Kuckartz, 2014) using the qualitative analysis software MAXQDA³.

In order to analyse the gender-fair post-editing process, Krings' (2001) division into temporal, technical, and cognitive effort was applied. The focus of this paper is on temporal and cognitive effort as well as an analysis of the final gender-fair translation and strategy. Screen recordings were used to measure post-editing times and, thus, temporal effort. To test whether the different approaches to GFL had an impact on translation speed, a linear mixed effects model was run with packages Imer4 and ImerTest in statistical analysis software R. GFL approach, participants' work experience and rates for GFL difficulty were used as independent variables while participants were used as a random factor. Observation protocols were produced by means of non-participant observation and aimed at reconstructing the post-editing process. Finally, interviews were used to gather data on the perceived cognitive effort of participants. In addition, gendered phrases in the post-edited texts were annotated based on the selected gender-fair language strategies and the success of their use.

5 Results

After presenting the participant's profile, the temporal and subjective cognitive effort for each text and gender-fair language approach used are presented. Furthermore, participants' impressions on the ease of using each strategy in post-editing are summarised.

5.1 Participants

Prior to the study, participants compiled a questionnaire to collect data regarding their profiles as well as their experiences with and use of gender-fair language. From the six participating translators, four identified as women and two as men. Unfortunately, no non-binary translator could be recruited for this post-editing task. Their work ex-

perience spanned from 3-5 to more than 20 years and all had extensive (4) or, at least, some (2) experience with PE. All participants indicated to already use gender-fair language to some extent in their daily work, with the exception of a patent translator who indicated that this is not desired in the field. An overview of the participants' profile is depicted in Table 2. All use gender-inclusive characters, such as gender star, two participants indicated alternating its use with gender-neutral rewording. Reasons for the use of GFL are to be more inclusive (3) and because it is becoming more common in written texts (2).

Participants were also asked to rate GFL difficulty on a Likert scale from one to five where one stands for very difficult and five for very easy. The vast majority was on the neutral to easy side as shown in Fig. 1.

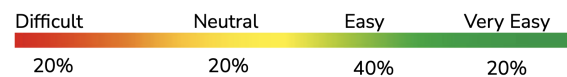


Figure 1: Rating of GFL difficulty

5.2 Temporal Effort

Differences in temporal effort were found among strategies and participants as shown in Fig. 2. Post-editing times for the first two GFL approaches, namely (i) gender-neutral rewording and (ii) gender-inclusive characters, were similar. Participants needed 00:19:59 minutes (SD = 00:06:03) on average to complete the first assignment and 00:17:49 (SD = 00:04:54) for the second. In the case of (iii) neosystems, the amount of time required was higher, i.e., 00:24:04 minutes (SD = 00:09:59).

In order to compare translation times across assignments, measurements were also normalised. A standard approach in research on PE is to divide the translation times for each task by the number of words in the machine translated source texts as in Table 3. Data showed a tendency for greater

³<https://www.maxqda.com>

Participant	Age	Gender	Work Experience	PE Experience	GFL Experience	GFL Use
P1	32-38	Woman	6-10	Extensive	Yes	Depends on client/assignment
P2	39-45	Man	11-15	Extensive	Yes	Gender-neutral characters
P3	53-59	Woman	20+	Some	Some	Gender-neutral characters
P4	25-31	Woman	3-5	Some	Some	Gender-neutral characters
P5	32-38	Man	6-10	Extensive	Yes	Depends on client/assignment
P6	39-45	Woman	16-20	Extensive	Some	No

Table 2: Participants' profiles



Figure 2: Post-Editing times for each assignment in minutes

temporal effort when post-editing in the third assignment, however, such difference was found to be not statistically significant (p -value>0.05).

Standard deviation for each task was high, indicating that there were considerable differences among participants in post-editing speed.

Assignment	Time (s/word)	Relative SD
1	7.9 ± 2.4	30%
2	7.0 ± 1.9	27%
3	9.0 ± 3.5	38%

Table 3: Average post-editing times, standard deviation (seconds per word), and relative standard deviation by text

The second task, requiring the use of gender-inclusive characters, was the fastest for four participants, P2, P4, P5, and P6. Only P1 and P3 took less time for the gender-neutral rewording assignment. In each case, PE times were shorter than for the third text with the use of neosystems. In the first assignment, four participants, i.e., P1, P2, P3, and P6, were faster than using neosystems. Interestingly, the temporal effort for P4 and P5 was lower in the third text than in the first.

In general, P2 was always the fastest participant and needed approximately ten minutes for each assignment. P2 and P5 showed no great variations in post-editing times with respect to the GFL approach used but needed both about 20 minutes for

each task. P6 generally took the longest to complete each task, i.e., 26 minutes for the first and 22 for the second. While using neosystems, P6 needed nearly 35 minutes to complete the post-editing but, in this case, P1 needed slightly more time. This last participant showed the greatest variation between the first two texts (completed in less than 20 minutes) and the third. Finally, P4 and P6 were the participants who showed the greatest variations among assignments, taking respectively about 26, 14 and 21 minutes, and about 26, 22, 35 minutes for each text.

5.3 Subjective Cognitive Load

During the post-study interviews, participants discussed each text and strategy commenting on their solutions, difficulties, and personal preferences. They also elaborated on their general experience as post-editors in the context of the study describing advantages and disadvantages concerning the use of machine translation for texts with references to non-binary individuals.

Generally, gender-neutral rewording was regarded as a feasible approach to gender-fair language, even though the majority of the participants (4) indicated gender-inclusive characters as the easiest GFL approach. There was also concordance that neosystems are the most difficult (4).

Even though most of the participants found the first approach easy to utilise (4), they also agreed that it is sometimes effortful to find neutral alternatives to gendered terms (3) that, for instance, “read well and do not repeat” (translated from German quote). The greatest difficulty in the post-editing process of this first text concerned the translation of the term “student” (5) because “for student (in a secondary school) there is no gender-neutral equivalent in German”. Half of the participants stated that they faced difficulties in finding a solution for “actor and musician” and thus “had to reflect a long time on how to phrase it”. The use of pronouns was mentioned as challenging by two

translators only, mostly because the main solution, i.e., the repetition of text referent's proper name, can negatively affect readability if the source text passage contains several third person singular pronouns.

Participants largely agreed (5) that gender-inclusive characters are easy to utilise because usual male and female forms of words are concatenated with a character. One participant, for instance, commented that “one does not need to reflect on what to do with a term [...] it is relatively clear how to handle it”. Nevertheless, a major concern (5) was that the sole use of this strategy could negatively affect readability. As a matter of fact, some text passages needed extensive gendering which was “found confusing while reading and perhaps also a bit challenging”. Accordingly, five participants admitted that they would prefer to use a mix of gender-neutral rewording and gender-inclusive characters for similar assignments. According to five participants, the greatest difficulty faced in the second text concerned the term “doctor”. The male and the female form of its German counterpart, namely “Arzt” and “Ärztin”, differ not only because of the ending, but also because of the umlaut on the first letter. Hence, a gender-inclusive form cannot be achieved by simply adding a star and the female ending. One participant even proposed to change the term “Arzt” to “Doktor” to avoid the issue. Finally, this approach to GFL challenges only partially the gender binary as most of the participants (4) specifically stated they needed to think of both female and male forms of words in order to utilise gender-inclusive characters.

The third assignment was the most difficult, notably because “it is something completely new” and participants “never used it in their work”. Some translators felt cognitively overwhelmed, as for example one who admits that they “have so strongly focused (on the use of the neosystem) that I had little attention left for the rest of the translation”. Since participants were not familiar with neosystems, they all had to use the handout they were provided with prior to the study as well as other resources for the whole duration of the post-editing process. The majority (5) indicated insecurity about the correct use of neosystems, which were perceived as a new, foreign or artificial language (4) which consequently negatively affects readability (5) and requires further training to be applied (5). Additionally, when consulting sources

on their use, a higher knowledge on the meta-level of language seems required. A good understanding and recognition of word classes is required, e.g. indefinite pronouns, relative pronouns, possessives, and grammatical structures, to be able to find gender-fair alternatives. One participant remarked “the German grammar should probably be revised to know which case to use”. Finally, for most participants (5) a specific challenge in the third text concerned the translation of the term “nanny”. The German equivalent, i.e., “Kinder mädchen” is gender-specific and even loaning the English word would grammatically be female. The term “Kinderbetreuer” (caregiver, male) could be used with gender-fair endings, but it differs in connotations from the English source word.

As regards the use of machine translation, most of participants (5) agreed that these were good and comprehensible even though they contained male generics and misgendering. Translators also felt that PE increased their speed and productivity. Two participants even stated that the MT draft allowed for more concentration on gendered elements. Only one participant felt that, due to mistakes in reference to gender-fair language, “(PE) was as effortful as translation from scratch”. The major difficulty when post-editing did not concern GFL but rather the decision on the extent to which MT outputs should be adapted, where three participants also highlighted that the style was not appropriate for the text type used in the study.

Even though all participants had previous experience with PE, half of them do not integrate MT translation in their usual workflow, thus using it only for PE assignments. Nevertheless, four participants mentioned using it sometimes as a source of inspiration. This is regarded as one of the main advantages of MT (3) alongside the fact that PE is generally faster and cheaper than translation from scratch (5). Only one participant, however, mentioned that they would use MT for further assignments requiring the use of non-binary GFL. When asked to comment on the disadvantages of the use of MT, participants did not mention the use of GFL but elaborated on the post-editing process in general. The majority feel that extensive PE is generally required for MT outputs (5) and that it is detrimental to creativity because they are constrained by the machine translated draft (4).

5.4 Strategy Selection

MT outputs of this study suffered from substantial gender bias. As shown in Table 4, nearly each gendered phrase was erroneously machine translated. For each non-binary noun, there were instances of misgendering. Singular they was translated with plural forms in German and plural nouns describing mixed-gender groups were translated with male generics. Consequently, participants post-edited all of these gender references in each text. The annotations of the final translations show great success of integrating GFL in the PE process, although with substantial differences in the use of strategies. From the three assignments, the first gender-neutral rewording required the highest rewriting effort of entire passages of text.

Gender-neutral rewording is a creative approach that can be realised differently, spanning from the use of neutral nouns to passive constructions. As a consequence, many different strategies were found. In the case of gender-inclusive characters, there was a clear preference (5) for gender star (*). However, this was applied quite differently by each participant. In the post-edited versions of the first two texts, misgendering, male generics or in general gender-specific mistakes were very rare as detailed below. In the third text, there was a strong preference (4) for a neosystem in particular, i.e., the Sylvain system. In this case, misgendering and mistakes occurred more frequently.

The first source text contained nine phrases with gendered elements that were of interest for this analysis. This amounted to 54 analysed phrases and a total of 58 annotations since some phrases were translated by combining different strategies. The most common strategy was the use of gender-neutral words/and or compounds (24%), e.g. “non-binary actor and musician” translated to “eine nicht-binäre schauspielerisch und musikalisch tätige Person” (a person active in music and acting). Many also opted for rewording whole phrases (22%). Some examples include “aus dem Schauspiel- und Musikbereich kommt” (who comes from music and acting) or “it (a loose uniform) makes them (Cal, the non-binary protagonist) feel more comfortable in who they are”, was translated by one participant as “weil sich diese angenehmer anfühlt” (because it feels better). 12% of the annotations also showed the omission of pronouns and 8% the repetition of the referent’s proper name. Other strategies included

the use of collective nouns, the omission of some information, and gender-inclusive characters, even though not permitted (each 3%). A participial form was used as well (2%). Finally, in 18% of cases no specific strategy was used as some source text segments contained the English pronoun they in reference to a mixed-gender group and the MT draft was appropriate. Only one instance of each misgendering and male generics was found in the 54 analysed phrases.

12 gendered phrases were analysed for the second text. 74 annotations were performed, meaning that in two phrases gender-inclusive characters were used along with rewording. In general, five participants opted for the use of gender star (*) which was, however, applied differently:

- male and female forms in the noun, e.g. “Schauspieler*in” (actor*actress) but female and male article or pronoun, e.g. “die*der” (the), switching the binary genders
- always male forms first, e.g. “der*die Schauspieler*in (the actor*actress);
- gender star to build nouns but slash (/) to build pronouns and articles, e.g. “der/die Schauspieler*in”
- female form first in pronouns and articles, but combined in a new form, e.g. “die*r Schauspieler*in”.

Switching the type of character within the same text is not recommended and in general female forms should be used first in articles and pronouns, e.g. “die*der” instead of “der*die” or the invented “die*r”. The remaining participant used colon instead of star in all instances, combined with a slash for articles and pronouns and male forms first. In the post-editing of the second text, no instances of misgendering were found and all of the participants’ solutions could be utilised, although some are less common than others, e.g. the combination of slash with another character. Male generics were used in two segments only by one participant, i.e., 3% of all analysed gendered phrases. In 17 segments (23%), strategies typical for the gender-neutral rewording approach were also used for passages that required extensive gendering and/or for the translation of the term “doctor” that, as mentioned before, was regarded as particularly challenging, e.g. “Ärzte*team” (team of doctors) and “Doktor*in”, which represents a change of terminology in the translation.

Text No.	MT Errors per Phrase	Types of MT Errors
1	7/9	Misgendering (2), Plural forms (3), Male generics (1)
2	10/12	Misgendering (6), Plural forms (3), Male generics (2)
3	9/10	Misgendering (7), Plural forms (3), Male generics (1), Co-reference (1)

Table 4: MT errors in reference to gender

Ten gendered phrases were analysed in the last assignment. 62 annotations were performed - in this case as well, two segments were post-edited with both a neosystem and rewording. Participants opted for different systems:

- Sylvain System (De Sylvain and Balzer, 2008), e.g. “einin muslimischin Schauspielernin” (a Muslim actor) (4);
- NoNA System (Geschlechtsneutrales Deutsch, nd), e.g. “eint muslimische Schauspieler*in” (1) ;
- Ens Forms (Hornscheidt and Sammla, 2021) e.g. “einens muslimisch Schauspielens” (1).

The choice for the Sylvain system was motivated by the impression that it was the most complete system, whereas participants selecting the NoNa System and the Ens forms perceived them as the easiest to use. Two participants admitted to arbitrarily deciding which neosystem to use. In this case, six instances of misgendering (13%) were found and all concerned the translation of “nanny”. Target text annotations also confirm participants’ doubts regarding the use of the neosystems. In 35% of the analysed segments, there was at least one mistake in the use of the selected system. One participant produced an error-free translation only utilising Ens. Interestingly, a tendency to overuse gender-fair forms was noted in one post-editing result. In a text passage, kids were mentioned and, even though the German equivalent “die Kinder” is also gender-neutral, one participant chose a gender-fair article (“dais Kinder”).

6 Discussion

From the results of this case study, quite a substantial variation in selecting gender-fair language for post-editing could be observed. When required to use gender-neutral rewording, participants omitted pronouns or repeated the character’s name to avoid gender marking. Nevertheless, the annotated post-edited segments also show a large degree of creativity with different rewording and terms used, e.g. “aus dem Schauspiel- und Musikbereich kommt” (who comes from music and acting). When

instructed to use gender-inclusive characters, the majority of the participants opted for gender star (*). However, its realisation was inconsistent in the case of pronouns and articles, at times erroneous. This included the use of other characters, such as slash (/), and a different order of male and female forms. When required to use neosystems, two trends could be observed: participants either opted for more sophisticated neosystems (the Sylvain System) or for easier ones (NoNa System and Ens Forms). While we ensured that the texts were equivalent in length and complexity, the fact that the neosystems came last after already two previous post-editing tasks could potentially have impacted the results. In the future, reordering the sequence between participants could account for this factor. The variation in the use of strategies will probably always occur, since even if used correctly there are many ways to reword a phrase. In terms of times, the use of different strategies did not impact PE speed. The great differences in time depended on the person more than the specific gender-fair strategy.

As regards perception, participants rated gender-inclusive characters as the easiest strategy, followed by rewording. Nevertheless, gender-neutral rewording requires considerable creativity which is sometimes perceived as challenging. Participants indicated a preference for a mix between rewording and gender-inclusive characters. There was general consent that neosystems are the most difficult approach to GFL as they are largely unknown and hence feel like a foreign language, which requires practice. This was also confirmed by the occurrences of mistakes found in the post-edited translations.

Participants were also interviewed on whether they consider MT in combination with PE as a viable option for producing gender-fair translations. As a general response, the MT draft was considered of good quality, requiring mostly stylistic adaptations, and PE was considered less time consuming than translation from scratch. Furthermore, the existing draft allowed for a focused revision of gender references. Nevertheless, half of the

participants stated they would not integrate MT in their workflow due to a negative view on the technology that, in their opinion, still requires extensive post-editing.

In a nutshell, the results of this case study suggest that even though unable to process gender-fair language, MT can still be a useful instrument for the translation of texts in which non-binary individuals are mentioned. Thus, we argue that post-editing might be a faster and viable option to generate test sets for gender-fair MT than producing translations from scratch. Moreover, even though differences in temporal efforts were not found among the strategies: (i) there is a tendency for longer PE times when neosystems are used which, in this study, is not statistically significant. This could be due to the small sample of participants, thus further experiments would be needed to shed light on this phenomenon; (ii) temporal effort does not necessarily correspond to the participants' perceived cognitive effort which was generally high, especially for neosystems.

In terms of methodology, interesting results could be obtained with the proposed mix of methods. However, to provide a less subjective evaluation of cognitive load, eye-tracking and key-logging experiments could be a potential alternative. It should also be noted that the group of participants had an overall positive attitude to gender-fair language, given that the vast majority already actively used it in their daily life and work. A repetition of the experiment with a larger, more varied population might lead to quite different results.

The results suggest substantial variation in the type of gender-fair language selected by translators, even if already restricted to a specific subtype. This has implications for MT in two regards. First, gender-fair translations or post-edited translations as future input texts might vary considerably in their gender references when describing non-binary individuals and MT should be able to handle these across languages. Second, gender-neutral MT as advocated by Piergentili et al. (2023) might not be the ideal option for all languages, since a clear preference for other strategies was stated by all participants in this study.

7 Conclusion

In this first gender-fair post-editing study, professional translators revised machine translated texts containing references to non-binary individuals

from the notional gender English to the grammatical gender German. Substantial variation in the implementation of the three gender-fair language strategies could be observed among participants, which implies for MT that a large variety of potential gender-neutral rewording and/or use of gender-inclusive characters, the two preferred strategies, need to be handled by the systems. The third strategy of utilising neosystems was perceived as requiring the highest temporal and cognitive effort.

Testing the cognitive and temporal load as well as success of using GFL on a larger scale and in different language pairs might be an interesting extension of the present study. For instance, eye-tracking would allow for a more detailed, objective analysis of the cognitive load of each strategy. Furthermore, a large-scale study across natural languages and their respective gender-fair language strategies would be interesting, especially when comparing post-editing to translation from scratch. This comparison could provide further insights into the effectiveness of post-editing within the context of gender-fair language use in the translation process.

8 Acknowledgments

We would like to warmly thank all the study participants for their motivation and interesting insights into the topic as well as UNIVERSITAS Austria for the financial and logistic support. We would also like to thank Associate Professor Anouschka Foltz for the support in performing the statistical analysis presented in the study.

References

- Albl-Mikasa, Michaela, Giovanna Fontana, Laura Maria Fuchs, Lena Meret Stüdeli, and Aline Zaugg. 2017. Professional translations of non-native english: 'before and after' texts from the european parliament's editing unit. *The Translator*, 23(4):371–387.
- APA Style. 2019. Gender.
- Attig, Remy. 2022. A call for community-informed translation: Respecting queer self-determination across linguistic lines. *Translation and Interpreting Studies*.
- Burtscher, Sabrina, Katta Spiel, Daniel Lukas Klausner, Manuel Lardelli, and Dagmar Gromann. 2022. "es geht um respekt, nicht um technologie": Erkenntnisse aus einem interessensgruppen-übergreifenden workshop zu genderfairer sprache und sprachtechnologie. In *Mensch und Computer 2022*. ACM.

- Corbett, Greville G. 1991. *Gender*. Cambridge University Press.
- De Sylvain, Cabala and Carsten Balzer. 2008. Die sylvain-konventionen-versuch einer “geschlechtergerechten” grammatik-transformation der deutschen sprache. *Liminalis*, 2:40–53.
- Dresing, Thorsten and Thorsten Pehl. 2018. *Praxisbuch Interview, Transkription & Analyse. Anleitungen und Regelsysteme für qualitativ Forschende*. dr dresing & pehl GmbH.
- En, Boka, Tobias Humer, Marija Petričević, Tinou Ponzer, Claudia Rauch, and Katta Spiel. 2021. Geschlechtersensible Sprache – Dialog auf Augenhöhe.
- Geschlechtsneutrales Deutsch. n.d. Das nona-system.
- Hornscheidt, Lann and Ja'n Sammla. 2021. *Wie schreibe ich divers? Wie spreche ich gendergerecht?: Ein Praxis-Handbuch zu Gender und Sprache*. Insel Hiddensee: w_orten & meer.
- Jakobsen, Arnt Lykke. 2017. Translation process research. In Schwieter, John W. and Aline Ferreira, editors, *The Handbook of Translation and Cognition*, pages 19–49. Wiley Online Library.
- Jia, Yanfang, Michael Carl, and Xiangling Wang. 2019. How does the post-editing of neural machine translation compare with from-scratch translation? a product and process study. *The Journal of Specialised Translation*, 31:60–86.
- Kincaid, J Peter, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Technical report, Naval Technical Training Command Millington TN Research Branch.
- Krings, Hans P. 2001. *Repairing texts: Empirical investigations of machine translation post-editing processes*, volume 5. Kent State University Press.
- Kuckartz, Udo. 2014. *Qualitative Text Analysis: A Guide to Methods, Practice and Using Software*. Sage.
- 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.
- López, Ártemis. 2019. Tú, yo, elle y el lenguaje no binario. *La Linterna del Traductor*, 19.
- López, Ártemis. 2022. Trans (de) lection: Audio-visual translations of gender identities for mainstream audiences. *Journal of Language and Sexuality*, 11(2):217–239.
- 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.
- Misiek, Szymon. 2020. Misgendered in Translation?: Genderqueerness in Polish Translations of English-language Television Series. *Anglica. An International Journal of English Studies*, 29(2):165–185.
- Piergentili, Andrea, Dennis Fucci, Beatrice Savoldi, Luisa Bentivogli, and Matteo Negri. 2023. From inclusive language to gender-neutral machine translation. *CoRR*, abs/2301.10075.
- Saunders, Danielle and Bill Byrne. 2020. Reducing gender bias in neural machine translation as a domain adaptation problem. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7724–7736, Online, July. Association for Computational Linguistics.
- Saunders, Danielle, Rosie Sallis, and Bill Byrne. 2020. Neural machine translation doesn’t translate gender coreference right unless you make it. In *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*, pages 35–43. 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.
- Stahlberg, Dagmar, Friederike Braun, Lisa Irmen, and Sabine Sczesny. 2007. Representation of the sexes in language. *Social Communication*, pages 163–187.
- Stanovsky, Gabriel, Noah A. Smith, and Luke Zettlemoyer. 2019. Evaluating gender bias in machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1679–1684. Association for Computational Linguistics, Florence, Italy, July.
- Toral, Antonio, Martijn Wieling, and Andy Way. 2018. Post-editing effort of a novel with statistical and neural machine translation. *Frontiers in Digital Humanities*, 5:9.
- Vergoossen, Hellen Petronella, Emma Aurora Renström, Anna Lindqvist, and Marie Gustafsson Sendén. 2020. Four dimensions of criticism against gender-fair language. *Sex Roles*, 83(5):328–337.
- Šincek, Marijana. 2020. *On, ona, ono*: Translating Gender Neutral Pronouns into Croatian. *Journal of the International Symposium of Students of English, Croatian and Italian Studies*, pages 92–112.