Gender-Fair Language in Translation: A Case Study

Angela Balducci Paolucci
University of Vienna, Austria
angelabalducci4@gmail.com

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

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

Abstract

With an increasing visibility of non-binary individuals, a growing number of language-specific strategies to linguistically include all genders or neutralize any gender references can be observed. Due to this multiplicity of proposed strategies and gender-specific grammatical differences across languages, selecting the one option to translate gender-fair language is challenging for machines and humans alike. As a first step towards gender-fair translation, we conducted a survey with translators to compare four gender-fair translations from a notional gender language, English, to a grammatical gender language, German. Proposed translations were rated by means of best-worst scaling as well as regarding their readability and comprehensibility. Participants expressed a clear preference for strategies with gender-inclusive character, i.e., colon.

1 Introduction

Gender in language reflects on an extra-linguistic reality (Corbett, 1991) in the sense that it reflects gender associations and stereotypes of a society. To respect different gender identities, i.e., the sense of self and "who they are" (Barker and Iantaffi, 2019), it is vital to linguistically acknowledge their existence within and across languages. Machine translation (MT) is known to suffer from gender bias, which is problematic for many reasons. For instance, machine-translated online contents are consumed without people being aware that they are MT mediated (Martindale and Carpuat, 2018). In MT research, the idea to resort to gender-neutral language to avoid gender issues has been proposed (Piergentili et al., 2023). However, apart from information loss, this might not be the preferred gender-fair strategy by humans. To analyse human preferences, we propose a first survey[1] among language professionals of four distinct gender-fair translation strategies from English to German.

Translation studies has a long tradition of considering gender issues, such as in feminist (Von Flotow, 1997) and queer translation (Baer and Kaindl, 2017). However, gender beyond the binary has so far received little scholarly attention (e.g. Misiek (2020) and López (2022)). The same is true for the field of MT, where debiasing strategies focus on a binary conception of gender, with some important exceptions (Tomalin et al., 2021; Saunders and Byrne, 2020). Gender-fair language, which subsumes gender-inclusive and gender-neutral strategies, is particularly challenging in case of grammatical gender languages, i.e., several word classes require gender inflections.

In this case study, ten language professionals rated four gender-fair translations of online magazine articles in direct comparison and regarding their impact on readability and comprehensibility. The four German strategies consist of one gender-neutral neosystem, one gender-inclusive neosystem, a gender-inclusive colon with si:er, and the same colon with neopronoun xier. Since rating a translation is in general a highly subjective matter, the selected method is best-worst scaling, which allows participants to select and rate their subjectively most (best) and least (worst) preferred trans-

© 2023 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

[1]The survey is made available on Zenodo: https://zenodo.org/record/7951054

lation. In a previous gender-fair MT workshop we conducted with translators, non-binary people, and MT experts (Burtscher et al., 2022), readability and comprehensibility of gender-fair language strategies were repeatedly named as important factors in the selection process. Thus, we decided to include a rating of these two dimensions in the present survey. Furthermore, participants were requested to motivate their choice in the form of a free text answer. While the perspective of non-binary individuals and MT experts would be equally interesting, we believe that preferences and considerations of language professionals as producers of (gender-fair) translations are of vital importance to the field of translation studies as well as machine translation. The results of this survey contribute to the discussion on which gender-fair language strategy is preferred in (machine) translating to German and which considerations are particularly important for language professionals.

2 Related Work

Since the focus of this article is on analyzing gender-fair translation strategies as a first step, this section focuses on work on gender-fair translation. In spite of the recent development of queer translation studies (Baer and Kaindl, 2017), research in the field of translation studies rarely addresses non-binary genders (Lardelli and Gromann, 2023). Most research focuses on media translation, e.g. subtitled and dubbed series, and news articles (López, 2022; Attig, 2022; Misiek, 2020; Šincek, 2020).

López (2019; 2022) and Attig (2022) analysed the dubbed and subtitled versions of the Netflix series *One Day at a Time* in Spanish and French. They found that the gender-fair language strategies used varied between the dubbed and subtitled versions as well as from European to Latin American Spanish. The non-binary character was correctly addressed with non-binary neopronoun *elle* in the European Spanish dubbed version only. In the other cases, they were misgendered with female forms and/or literal translations of English singular *they*. Similarly, in the French dubbed version, non-binary neopronoun *ielle* was used whereas in the subtitles the character was referred to with indefinite pronoun *on* (one/we).

In their analysis of English TV series translated to Polish, Misiek (2020) found a systematic omission of the non-binary characters' gender identity. This phenomenon could also be observed in Croatian movie translations and articles on Sam Smith's coming out as non-binary where the third person masculine plural pronoun was generally used (Šincek, 2020). Šincek (2020) represents also one of the few works to include interviews with people, i.e., non-binary individuals, on the topic.

Recent developments in gender-fair language strategies have been studied in psycholinguistics with a focus on binary genders. For instance, Lindqvist et al. (2019) conducted experiments in Swedish and English and tested different strategies to reduce male bias in language, i.e., (i) binary paired forms, (ii) gender-neutral words as well as (iii) gender-fair pronoun *hen* and English singular *they*. Participants read a description of a candidate for a job position and were asked to select photos of men or women corresponding to the said description. The results suggest that (i) and (iii) actively reduce male bias.

In German, empirical research concentrated on the cognitive processing of textual information. Braun et al. (2007), for example, tested the effect of male generics and two binary gender-fair language forms on memory performance and text intelligibility. No differences in memory performance across strategies were found between men and women. However, as concerns intelligibility, women indicated no preferences, while men indicated a preference for male generics.

To the best of our knowledge, this is the first study to consult language professionals regarding their preferences regarding gender-fair language strategies. Since language professionals play the important role of producing gender-fair translations, needed to fine-tune MT models, we believe that their perspective is interesting for translation studies and the field of machine translation.

3 Preliminaries

In order to establish the theoretical foundation of the present survey, an introduction to the interaction of gender with language and translation is provided, followed by a brief overview of gender-fair language strategies in English and German.

3.1 Gender and Language

The relation between gender and language is complex because the term has multiple meanings. In the field of gender studies, it is defined as a biopsychosocial construct (Barker and Iantaffi, 2019). It

hence involves biological, e.g. hormonal, psychological, e.g. a person's sense of self, and social, e.g. normative and cultural expectations, factors. It is commonly used in reference to gender identity, i.e., a person's sense of their gender, and not the sex assigned at birth. In linguistics, the term is generally defined as "classes of nouns reflected in the behaviour of associated words" (Hockett, 1958, 231). In other words, associated word classes are inflected based on the grammatical gender of a specific noun.

Gender is realised differently in natural languages, which can be classified into (i) grammatical gender, (ii) notional gender, and (iii) genderless languages (Stahlberg et al., 2007; McConnell-Ginet, 2013). In (i), such as German and Italian, each noun has a gender (Corbett, 1991) and extensive gender marking is required. In (ii), such as English, third person singular pronouns, i.e., *he, she, it*, and specific nouns, e.g. *boy/girl*, are gender-specific. In (iii), such as Turkish, gender may be expressed, e.g. in kinship, but is not grammatically encoded in linguistic structures. Gender assignment in the case of human referents is based on the extra-linguistic reality of a society (Corbett, 1991) and reveals gender associations and stereotypes as well as connotations (Nissen, 2002; Jakobson, 1959).

3.2 Gender and Translation

Differences in linguistic structures and gender-specific connotations impact the translation process. In the first case, the translation from notional to grammatical gender languages can require choices that are not neutral (Nissen, 2002; Di Sabato and Perri, 2020). In several literary works, for instance *Written on the Body* (1993), a mysterious atmosphere is created by omitting gender markers. However, when translating to another language, this omission of gender might not be grammatically feasible, potentially forcing translators to assign a gender to characters (Di Sabato and Perri, 2020). This choice is often based on social gender, i.e., stereotypical associations to gender in a society (Nissen, 2002). In the second case, gender can be used to convey particular connotations through personifications and metaphors. This occurs, for example, in marketing texts and/or advertisement, where an animal, such as a male, fast tiger, is used to represent a car. Since the same animal can have different or no gender-specific connotations in other languages and cultures, translation choices that deliver the same source text message are required (Di Sabato and Perri, 2020).

3.3 Gender-Fair Language

Gender-fair language has a long tradition. Its development goes back to the 1960s, when differences in the linguistic treatment of men and women gained the attention of second-wave feminists (Kramer, 2016). With an increased visibility of non-binary people, new gender-fair language strategies have been accordingly proposed.

In English, singular *they* has become common to refer to people whose gender is unknown or irrelevant to the context of conversation as well as non-binary people (Apa Style, 2019). Furthermore, gender-neutral alternatives to gendered words, such as *chairperson* instead of *chairman*, are increasingly used (Weatherall, 2002). In German, a grammatical gender language that requires extensive gender marking, there are mainly four approaches:

- **gender-neutral rewording**: sentences are phrased in order to avoid gendered structures, e.g. person as gender-neutral word, indefinite pronouns, passive constructions and participial forms;

- **gender-inclusive characters**: typographic characters, such as gender star (*) or colon (:), are used to separate male forms from female endings and include all genders, e.g. *Leser*in* (reader). It is also possible to separate the stem from the noun ending as in *Lese*rin*, which should prevent binary thinking.

- **gender-neutral characters or endings**: for example *x* in *Lesx* (reader) are used to question the gender binary.

- **neosystems**:
  - gender-inclusive: a new gender is introduced in the language as in the case of the Sylvain system (De Sylvain and Balzer, 2008) with *Lesernin* (reader).
  - gender-neutral: the *ens* pronoun and suffix as in *Lesens* (reader) is introduced as gender-neutral form derived from M**ens**ch (human) (Hornscheidt and Sammla, 2021).

Furthermore, several neopronouns have been proposed. For instance, *xier* is the result of the combination of third person singular female *sie* and male *er* pronoun and has already been used in the translation of some English language TV series (Heger, 2020). Several more detailed overviews of gender-fair language in German are available (Hornscheidt, 2012; En et al., 2021; Hornscheidt and Sammla, 2021).

4 Method

In order to evaluate the perception, readability, and comprehensibility of gender-fair language strategies, two empirical methods targeted to measure subjective impressions were selected, i.e., Best-Worst Scaling (BWS) and the Likert scale. BWS (Louviere and Woodworth, 1990), a comparative annotation method, was used to select and evaluate the subjectively best and the worst translation strategy, whereas the Likert scale (Likert, 1932), a rating scale, was used to rate the readability and comprehensibility of the best and worst strategy chosen by the participants. Readability refers to whether a text written in a specific gender-fair language strategy is easy and enjoyable to read for the participants of this study subjectively. Comprehensibility refers to the ease to understand the message of a text written in a specific strategy for the participants of this study subjectively. The choice to combine these two methods is based on the desire to limit the granularity and inconsistencies that can occur when using solely a rating scale (Kiritchenko and Mohammad, 2017).

4.1 Data and Strategy Selection

Four English texts containing the use of singular *they* were selected from online articles to be translated using four different gender-fair language strategies. To be specific, the texts selected were interviews and reports on non-binary people in *Entertainment Weekly* (Text 1), *People* (Text 3) and on the website of the *Brown University* (Text 2) as well as a set of instructions on how to support a non-binary friend published on *Sociomix* (Text 4). Due to the fact that German is a grammatical gender language that associates gender with nouns in addition to pronouns, adjectives, and determiners, selected texts should allow to reflect this grammatical variety in the translation. For each original text, four gender-fair translations are provided in a set, which only differ in the utilized gender-fair language strategy. All translations were created manually and checked by three experts on gender-fair German. As strategies to be employed during the translation process, the choice fell on:

1. gender-neutral neosystem *ens*, because of its simple grammatical structure, where no declension and consequently easy use is expected;

2. gender-inclusive Sylvain neosystem, follows the grammatical rules of the German language, which is why it is expected to appear more natural;

3. colon after the word stem in combination with the pronoun *si:er*, because the colon is already widely known and used and with the two binary German pronouns combined should least impact readability and comprehensibility, and

4. colon after the word stem in combination with the *xier* pronoun, for the same reason of the colon and because the "x" explicitly emphasizes the inclusion of all genders, not only binary genders (Heger, 2013).

To exemplify the type of text and gender-fair translation strategies that were used in this study, we provide all four strategies for the sentence *Jim is a fierce pirate who journeys the seas seeking revenge on the people that killed their family.* of Text 1, an *Entertainment Weekly* interview with and article on Vico Ortiz who starred as non-binary pirate Jim in *Our Flag Means Death*:

1. Jim ist einens grimmig Piratens, dens durch die Meere reist, um sich an den Personen zu rächen, die ens Familie getötet haben.

2. Jim ist einin grimmigin Piratnin, din durch die Meere reist, um nimser an den Personen zu rächen, welche nimse Familie getötet haben.

3. Jim ist ei:ne Pira:tin, dier durch die Meere reist, um sich an den Personen zu rächen, welche siese Familie getötet haben.

4. Jim ist ei:ne Pira:tin, dier durch die Meere reist, um sich an den Personen zu rächen, welche xiese Familie getötet haben.

4.2 Participant Selection

For a principled selection of participating language professionals, a number of criteria had to be specified. First, their first language had to be German and they had to have a high command of English, i.e., C1 to C2 of the Common European Frame of Reference for Languages (CEFR), in order to be able to better identify which gender-fair strategy could be used as a translation for the English singular *they*. Second, participants were required to have completed or be about to complete a professional education in the field of translation. Finally, at least some practical translation experience beyond exercises during the education was required.

4.3 Survey Design

After introductory instructions and basic questions in a Google Forms survey, four translations corresponding to the four gender-fair strategies were presented side by side with the English original for each of the four source texts. For each pair of original and translations, participants were asked to select the best and the worst translation from the set and rate the former on a scale from +4 (very good) to 0 (neutral) and the latter from 0 (neutral) to -4 (very bad), a common scale and practice in BWS. Furthermore, participants were requested to rate the readability and comprehensibility of the best as well as worst translation selected on a Likert scale from 5 (very true) to 1 (not true). For the best strategy, the statements to be rated were that the best strategy does not impact the readability of the text and with the best gender-fair strategy the text is easy to understand. Thus, a rating of 5 means easy to read and highly comprehensible. For the worst strategy, the statements to be rated were that the worst strategy impacts the readability and makes the text hard to understand. Thus, a rating of 5 means hard to read and low comprehensibility. The general assumption was that the best strategy would have little impact on these two dimensions, while the worst is expected to achieve low ratings for both. Participants were also requested to optionally motivate their best/worst choices for each individual set of gender-fair translations as a free text answer. Furthermore, the demographic and general answers were analyzed to determine differences across participants and gather their prior experience with gender-fair language and translation as well as their opinion on the topic. The basic questions, thus, included participants' experience with and impressions on gender-fair language.

4.4 Analysis

The numeric BWS ratings are summed up by strategy across all four sets and all participants and divided by the number of times the strategy was rated to obtain the finally best and worst strategy on average in the survey. The same procedure was applied to the ratings on readability and comprehensibility. Finally, the free text answers and basic questions were analyzed and annotated for a topic-wise presentation of the results.

5 Results

After presenting participants' profiles, their preferences regarding the evaluated strategies, ratings for readability and comprehensibility, and overall comments on the topic are detailed.

5.1 Participant Profile

From the ten participants in the survey, nine identified as woman and one as man. In terms of age, 30% were between 18 and 25, 40% between 26 and 29, and 30% between 30 and 40. As required, all participants indicated to be professionally educated, have translation experience, and a high command of English (C1 or C2). All participants indicated to have prior knowledge of gender-fair language strategies, in particular neutral rewording and inclusive gender star and colon, and 90% indicated to be actively using gender-fair language in their daily lives. Another binary strategy that was indicated is to camel case plural endings with I to include men and women, e.g. *LeserInnen* instead of the female *Leserinnen* or the male *Leser*.

5.2 Ratings of Gender-Fair Translations

The detailed results of BWS ratings per participant, text, and gender-fair strategy are presented in Table 1. Each of the ten participants rated one translation per set as best and one as worst, resulting in a total of 40 positive/neutral and 40 negative/neutral ratings for four sets. Positive ratings are marked in green, negative ratings in red, and neutral ones in gray. The translation strategies in the columns correspond to the numbered list in Section 4.1, that is, S1 corresponds to the ens strategy, S2 the Sylvain system, S3 colon + si:er, and S4 colon + xier.

In Table 2, the counts of how often a strategy was selected as best or worst as well as the overall

	Text 1				Text 2				Text 3				Text 4			
Part.	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
P1		0		+3		0		+3		0		+3	+3	0		
P2		-4	+1			-2	+2			-3	+4			-1	+1	
P3		-3	+2			-3		+2	-2			+2		+3		-2
P4		-2	+3			-3	+3		-3		+4			-1	+4	
P5	-3		+2			-3		+3		-3	+3			-3		+4
P6		-4	0			-4	0		-4		0		-4		0	
P7	-4		0			0		+1	0			+2	0			+3
P8	+3	-2			+2	-1			+2	-2			+2	-2		
P9		-3	+3			-2	+3			-4	+3			-2	+3	
P10	-3			+2	-2			+2	-3			+2		-3		+2
Sum	-7	-18	+11	+5	0	-18	+8	+11	-10	-12	+14	+9	+1	-9	+8	+7

Table 1: Detailed BWS rating results per participant, strategy, and text

count and percentage it was selected are presented, alongside the sum, average and median BWS rating. The gender-fair translation strategy S4 obtained 18.75% of all ratings and achieved the best average rating of 2.13, followed by S3 with 25% of all ratings and on average 2.05 as can be seen from Table 2. While the Sylvain system obtained by far the most ratings, i.e., 36.25% of in total 80 ratings, from the numeric rating distribution and the color coding in Table 1 it becomes evident that it obtained mostly negative scores and the worst overall result with on average -1.97. Finally, S1 obtained 20% of all ratings and on average a final score of -1. Interestingly, S3 colon + si:er was never selected as worst strategy and did not obtain a single negative rating as can be seen from Table 1. Furthermore, it was most frequently selected as best strategy with 20 (50%) out of 40 positive rating counts. The overall best strategy S4 colon + xier was only selected once as worst strategy and obtained a negative rating by P3. Given that P3 breaks their previous pattern of rating S2 as worst, it might have been an accidental selection.

	S1	**S2**	**S3**	**S4**
Best C.	5	1	20	14
Worst C.	11	28	0	1
Total C.	16	29	20	15
Av. C. (%)	20.00	36.25	25.00	18.75
Sum R.	-16	-57	41	32
Av. R.	-1.00	-1.97	2.05	**2.13**
Median R.	-2.00	-3.00	3.00	2.00

Table 2: Summary of BWS rating results (C = Count; R = Rating; Av = Average)

While the decision that the S2 Sylvain system is the worst strategy was quite unanimous, some participants revealed individual preferences as can be seen in Table 1. Participant P8 showed a strong preference for the S1 ens strategy for all texts, while overall S1 obtained more negative than positive or neutral ratings. In terms of intra-annotator consistency, participants P2, P8, and P9 are completely consistent in their selection of strategies across texts. Other participants, especially P3, P5, and P7 changed their preferred strategies depending on the text, in particular with Text 3 and Text 4. This change could be attributed to the fact that the first two texts are equivalent in type since both are interviews, while Text 3 is a report on Demi Lovato and Text 4 represents a set of instructions of how to support a non-binary friend. Thus, there is considerable inter-annotator variation, however, overall the consensus is that colon with xier is the best and the Sylvain system is the worst gender-fair translation strategy for this group of participants.

5.3 Readability and Comprehensibility

Since in previous interactions with the target group of this study readability and comprehensibility were named as important factors for the choice of gender-fair language, participants were asked to rate both dimensions for the selected best and worst strategy. In Table 3 the average score for the best and worst strategy for both dimensions is provided, where for the best strategy 5 means high readability and comprehensibility and 1 means low readability and comprehensibility. For the worst strategy, participants were asked whether they agree that the strategy negatively influences readability and comprehensibility, which means

Best	**S1**	**S2**	**S3**	**S4**
Readability	3.40	3.00	2.95	3.07
Comprehensibility	3.60	3.00	3.00	3.21
Worst	**S1**	**S2**	**S3**	**S4**
Readability	4.55	3.86	0.00	4.00
Comprehensibility	4.55	3.50	0.00	4.00

Table 3: Average score on readability and comprehensibility

that full agreement (5) indicates low readability and comprehensibility, while 1 indicates a high rating for both dimensions.

As is to be expected, the strategies that were selected as worst were also rated as low in readability and comprehensibility, where S1 was on average indicated as the strategy with the highest impact on both dimensions. Table 3 confirms the fact that S3 was never selected as worst strategy by any participant. Ratings for the best strategy are more surprising, since even though participants considered a strategy the comparatively best from the set, they still indicated an impact on how easy to read and comprehend the gender-fair text is. On average, the ratings are rather neutral around 3, with only slightly worse ratings for the ens strategy (S1).

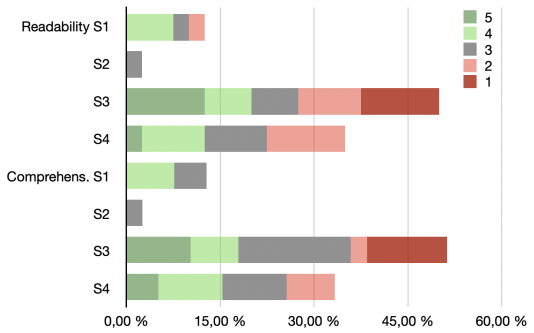

Figure 1: Detailed Scores of Best Strategy

To provide a closer look at the ratings of the best strategy, a detailed overview of scores is depicted in Fig. 1. S3 colon + si:er was selected most frequently as best strategy, which means it obtained most ratings for readability and comprehensibility. From the overall 20 ratings for S3 as best strategy, 9 (45%) ratings were very low with 1 or 2. Comprehensibility seems to be less of an issue, since only 6 (30%) ratings were below 3. In fact, as can be seen from Fig. 1, S3 is the only strategy to ever obtain a rating as low as 1 for both dimensions. However, it should be kept in mind here that with a small sample, single participants have an impact. Participant P6 consequently rated both dimensions as very low for a strategy across all texts, making up 4 of the 20 ratings for S3 as best strategy and of its readability and comprehensibility. This is in line with the overall evaluation of BWS, where P6 would never assign a higher score than 0 to any strategy (see Table 1). For S1 the results mostly rely on P8, who consistently selected it as the best strategy and considered both dimensions as high. P1 selected S1 ens only once as best strategy and provided a low rating for readability. An increase in scores from Text 1 to later texts could be explained by an increase in familiarity with the strategy, as commented by one participant. Overall readability seems to be a bigger issue than comprehensibility.

5.4 Participant Comments

Free text comments on the individual texts as well as on the survey in general reflected the overall negative attitude of participants towards the S2 Sylvain system. Participants remarked that texts written with this gender-fair language strategy would not be intelligible without the English original and especially the meaning of pronouns is hard to understand even in context, requiring an unnecessary cognitive effort. This comment reinforced our research design choice to provide the English original alongside the translations. One participant considered simply omitting possessive pronouns with this strategy as the best option. Other comments included that it generates texts that are perceived as grammatically incorrect, unnatural, and unnecessarily complicated, inhibiting the natural flow of the text.

In reference to S1, the ens system, participants mainly remarked on a detrimental effect on comprehensibility in their comments, which is in line with the fact that S1 obtained the worst overall ratings on comprehensibility in the survey. P5 remarked that they had to read the text several times in order to grasp its meaning and for P4 the text with this strategy seemed as if written in Dutch, distracting them from understanding it. P8, the only and most fervent advocate for S1, stated that to them it is the simplest strategy that is easy to use, both in written and spoken communication. Furthermore, to P8 ens imitates the English *they*, making it the ideal strategy for gender-fair translation from English. On the other hand, P8 remarks that the lack of noun declensions with this strategy might not be ideal.

In reference to the two best rated strategies with colon after the word stem, participants remarked that pronouns at times still seem unfamiliar, especially possessive pronouns, e.g. *sieser*, might at first be confused with demonstrative articles, e.g. *dieser*. Nevertheless, participants noted that it is comparatively easy to familiarize themselves with this strategy, which has little negative impact on the readability of the text.

One interesting change of comments from the first to the last set of English original and translations could be observed with participant P7, who on Text 1 commented that all of the proposed strategies have an entirely negative impact on readability. For Text 2, the remark changes to colon with xier might not be entirely unreadable, which progresses to relatively easy to read in Text 4. This change of heart is reflected in P7's ratings of the dimension readability, which progresses from 1 for the best strategy in Text 1 to 3 for the best strategy in Text 3 and 4. Comprehensibility never obtains a higher rating than 2. P7 finally concludes that the colon is less distracting also for pronouns than xier and its corresponding declensions.

As an overall evaluation of the entire survey, P7 provides an explicitly negative attitude towards the topic as such and an explicitly low opinion of gender-fair language in general, indicating to not use any such strategy privately and expressing the belief that a general public can hardly be expected to utilize such "creations". This overall belief is shared by P6, who provides low ratings for the best strategy as well as its readability and comprehensibility and at the end of the survey remarks that, while the topic is interesting, none of the proposed strategies find their liking and will hardly be used in everyday communication.

All participants but one considered the topic of gender-fair language strategies in translation interesting and important. Beyond the proposed strategies, the repetition of names instead of pronouns or rewording of nouns in the translation were indicated. One important aspect that was mentioned is the familiarity with and prior knowledge of the topic. Participants indicated that readability and comprehensibility improved from Text 1 to Text 4, highlighting how fast they were able to get more accustomed to these strategies. This factor of being accustomed and familiar with the individual strategies might in the end also change the overall evaluation, which for now leans towards S3 and S4 as the strategies closest to the current German language use.

6 Discussion

One initial assumption of this survey was that the gender-inclusive Sylvain system might be preferred on the basis that it follows the grammatical rules of German and thus, might seem more natural than other strategies. However, this strategy was overwhelmingly rated as the worst in the set, appearing unnatural, erroneous, confusing, and overly complex. Its ratings on readability and comprehensibility reflect these comments. The overall correspondence between BWS ratings and Likert scores indicates a tight link between personal preferences for gender-fair translation strategies and their subjective readability and comprehensibility, emphasizing the importance of the two dimensions chosen for this study. However, in future research these dimensions should take the specific needs of people with physical and/or cognitive disabilities into consideration, e.g. by conducting a survey with a more diverse group of participants.

The gender-neutral ens system is comparatively easy to use from a grammatical point of view, since it requires no declensions. However, this grammatical simplicity considerably alters the language with a detrimental effect on readability and comprehensibility, as shown by the overall ratings and participants' comments. Even its only advocate in the survey doubted the general applicability of a language system without declensions in German.

In the set of proposed strategies the colon after the word stem emerged as the clear winner, with a slight preference for its use with the neopronoun xier over introducing another colon in the pronouns as in si:er. Since these two strategies, S3 and S4, are the ones closest to the current language use, it can be assumed that familiarity with strategies plays a role in the selection of preferences, which was reflected in a participants' comment. Thus, it would be interesting to evaluate whether a thorough introduction including exercises to the other strategies would alter the final selection of preferred strategies. In fact, in a previously conducted workshop (Burtscher et al., 2022), participants obtained such a thorough introduction and then in exercises opted for the ens strategy (S1). One participant in this case study even remarked on the fact that familiarity, readability, and comprehensibility already increased from the first to

the last text, where each of which was very short. This indicates that familiarizing participants with different strategies might be feasible in a large-scale survey or experimental setting and would be an interesting alternation for future studies on gender-fair translation.

Due to the multiplicity of proposed gender-fair language strategies in German, we opted for a small selection in this survey in order not to overwhelm participants. This selection was driven by the intention to compare gender-neutral and gender-inclusive strategies. However, for the former we only included one neosystem, where other strategies, such as rewording, are available. For the latter category, three strategies were included, where even for the best strategy a change in typographical character might already impact the responses, i.e., underscore or star instead of colon as well as placement of the character. The colon after the word stem was explicitly chosen to reduce the emphasis on male/female endings.

Since we only included one gender-neutral option, a subselection of gender-inclusive strategies and a sample limited in size, no conclusions on the preference of either category can be drawn, for which future studies with a different setup are foreseen. However, the importance to be equipped with gender-fair translation strategy that finds acceptance by the general public was emphasized. Thereby, common translation problems, such as involuntary or accidental misgendering in the translation, could be mitigated or solved. For instance, if the source text is rather vague on the gender of a character/person in a notional gender language, some gender-fair translation strategies enable equal vagueness in a grammatical gender language. While the best strategy ultimately depends on the context not only in the textual sense but also in the sense of the translation assignment, target group, purpose, etc., some strategies might be easier to use, comprehend, and read than others and might impact the transfer from the source to the target text differently.

In terms of implications for translation technologies and in particular machine translation, we believe that this survey reveals how complex and language-specific the topic of gender-fair language and translation truly is. While overall preferred strategies could be identified, individual participants showed different preferences, e.g. one participant clearly preferred the ens strategy. Thus, machine translation might need to be able to accommodate different gender-fair language strategies depending on the language, context, purpose, and target audience of a translation. These preferences or requirements might also change with the domain of texts, where in this case study the degree of domain-specificity of media texts is rather low. In this case study, the task was also to select from a set of existing translations. It would be interesting to evaluate the performance and preferences of professional translators when asked to perform gender-fair post-editing of machine translated texts.

7 Conclusion

In order to socially and linguistically include different gender identities, a multiplicity of gender-fair language strategies, in particular for grammatical gender languages, has been proposed. The transfer of gender-fair strategies across structurally different languages is challenging for machines and humans. Thus, it is interesting to evaluate the position and preferences of language professionals on the topic of gender-fair translation from a notional to a grammatical gender language. In the presented survey results based on best-worst scaling, ten language professionals revealed a preference for the gender-inclusive strategy of colon after the word stem in combination with the neopronoun xier over the gender-inclusive Sylvian and gender-neutral ens neosytem. The alternative of colon with si:er was rated only slightly lower than with xier, where participants commented on a preference for pronouns without typographical character that they considered more natural. For both strategies with colon the overall rating on the dimensions of readability and comprehensibility was neutral to positive, whereas ens was considered to negatively impact both dimensions the most. A correspondence between the expression of preferences and the ratings of readability and comprehensibility as well as explicit references to these two dimensions in free text comments confirmed their importance within the context of gender-fair translation strategies.

In the present study, a preference for a gender-inclusive strategy could be observed, however, with a limited selection of strategies and a small number of respondents. To obtain a general preference regarding gender-neutral or gender-inclusive strategies, a large-scale study with a stronger va-

riety of gender-fair language strategies across languages and a larger, potentially more diverse target group would be required. The results indicate that preferences might also vary depending on the participants' degree of familiarity with individual strategies, which is a factor worth investigating in future endeavors. Finally, the impact of the level of domain specificity and text type would be interesting factors. Nevertheless, with this first study on gender-fair translation among language professionals we hope to have provided a methodological contribution as well as first results on gender-fair translation strategies from the perspective of language professionals, a method that can easily be transferred to future studies and even evaluating machine translation results.

References

Apa Style. 2019. Singular "they".

Attig, Remy. 2022. A call for community-informed translation: Respecting queer self-determination across linguistic lines. *Translation and Interpreting Studies*.

Baer, Brian James and Klaus Kaindl. 2017. Introduction: Queer (ing) translation. In Baer, Brian James and Klaus Kaindl, editors, *Queering Translation, Translating the Queer*, pages 1–10. Routledge.

Barker, Meg John and Alex Iantaffi. 2019. *Life isn't Binary*. London, UK: Jessica Kingsley Publishers.

Braun, Friederike, Susanne Oelkers, Karin Rogalski, Janine Bosak, and Sabine Sczesny. 2007. "aus gründen der verständlichkeit...": Der einfluss generisch maskuliner und alternativer personenbezeichnungen auf die kognitive verarbeitung von texten. *Psychologische Rundschau*, 58(3):183–189.

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.

Di Sabato, Bruna and Antonio Perri. 2020. Grammatical gender and translation: A cross-linguistic overview. In von Flotow, Luise and Kamal Hala, editors, *The Routledge Handbook of Translation, Feminism and Gender*, pages 363–373. Routledge.

En, Boka, Tobias Humer, Marija Petričević, Tinou Ponzer, Claudia Rauch, and Katta Spiel. 2021. Geschlechtersensible Sprache – Dialog auf Augenhöhe.

Heger, Illi Anna. 2013. Version 3.2 : Xier pronomen ohne geschlecht.

Heger, Illi Anna. 2020. Version 3.3 : Xier pronomen ohne geschlecht.

Hockett, Charles F. 1958. *A Course in Modern Linguistics.* New York: Macmillan.

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.

Hornscheidt, Lann. 2012. *feministische w_orte: ein lern-, denk- und handlungsbuch zu sprache und diskriminierung, gender studies und feministischer linguistik*. Frankfurt: Brandes & Apsel.

Jakobson, Roman. 1959. On linguistic aspects of translation. In Brower, Reuben Arthur, editor, *On Translation*, pages 232–239. Cambridge.

Kiritchenko, Svetlana and Saif M Mohammad. 2017. Best-worst scaling more reliable than rating scales: A case study on sentiment intensity annotation. *arXiv preprint arXiv:1712.01765*.

Kramer, Elise. 2016. Feminist linguistics and linguistic feminisms. In Lewin, Ellen and Leni M. Silverstein, editors, *Mapping Feminist Anthropology in the Twenty-first Century*, pages 65–83. Rutgers University Press.

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.

Likert, Rensis. 1932. A technique for the measurement of attitudes. *Archives of psychology*.

Lindqvist, Anna, Emma Aurora Renström, and Marie Gustafsson Sendén. 2019. Reducing a male bias in language? establishing the efficiency of three different gender-fair language strategies. *Sex Roles*, 81:109–117.

López, Ártemis. 2019. Tú, yo, elle y el lenguaje no binario. *La Linterna del Traductor*, 19.

López, Ártemis. 2022. Trans (de) letion: Audiovisual translations of gender identities for mainstream audiences. *Journal of Language and Sexuality*, 11(2):217–239.

Louviere, Jordan J and George G Woodworth. 1990. Best worst scaling: A model for largest difference judgments [working paper]. *Faculty of Business*.

Martindale, Marianna J. and Marine Carpuat. 2018. Fluency over adequacy: A pilot study in measuring user trust in imperfect MT. *CoRR*, abs/1802.06041.

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.

Nissen, Uwe Kjær. 2002. Aspects of translating gender. *Linguistik online*, 11(2):25–37.

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. *arXiv preprint arXiv:2004.04498*.

Stahlberg, Dagmar, Friederike Braun, Lisa Irmen, and Sabine Sczesny. 2007. Representation of the sexes in language. *Social Communication*, pages 163–187.

Tomalin, Marcus, Bill Byrne, Shauna Concannon, Danielle Saunders, and Stefanie Ullmann. 2021. The practical ethics of bias reduction in machine translation: Why domain adaptation is better than data debiasing. *Ethics and Information Technology*.

Von Flotow, Luise. 1997. *Translation and Gender: Translating in The 'Era of Feminism'*. Routledge.

Weatherall, Ann. 2002. *Gender, language and discourse*. Hove: Routledge.

Winterson, Jeanette. 1993. *Written on the Body*. Random House.

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.