Gender-Fair (Machine) Translation

Conference Paper · March 2023

2 authors, including:

Manuel Lardelli
Karl-Franzens-Universität Graz

6 PUBLICATIONS · 8 CITATIONS

All content following this page was uploaded by Manuel Lardelli on 12 April 2023.

The user has requested enhancement of the downloaded file.
Gender-Fair (Machine) Translation

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

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

Abstract

Recent years have seen an increased visibility of non-binary people in public discourse. Accordingly, gender-fair language strategies that go beyond a binary conception of gender have been proposed. Such strategies pose a challenge for both translators and machine translation (MT), and gender-fair (machine) translation represents a relatively novel research field. With this survey and discussion, we hope to provide a starting point for this field and contribute to a detailed overview of (machine) translation strategies to counteract the misrepresentation of an individual’s gender. The results show that gender-fair translation studies (TS) approaches largely focus on media translation, such as subtitles or news articles, and the MT results show that the need to include non-binary debiasing methods is increasingly acknowledged, however, hardly ever implemented. Ideas on a closer mutually beneficial interaction between MT and translation studies are presented to advance multilingual gender-fair language use.

1 Introduction

Gender issues have always been a topic of interest in translation studies (TS) (Venuti, 1995; Tymoczko, 2010; Spurling, 2014). Accordingly, new paradigms emerged within the field of TS, such as the feminist and queer translation (Venuti, 1997; Baer and Kaindl, 2017). This led to new issues to be investigated, including (i) sexist tropes in translation theory (Chamberlain, 1988), (ii) the work of feminist and queer translators throughout history, (iii) the translation of misogynist, homophobic and/or anti LGBTQIA+ texts, the translation of feminist and queer texts (Venuti, 1991, 1997; Burton, 2010; Démont, 2017), and (iv) the translation of gender between languages with different structures (Nissen, 2002; Di Sabato and Perri, 2020).

Already with early approaches of machine translation (MT), gender issues surfaced, such as with reference to (i) anaphora resolution (Mitkov, 1999), (ii) named entity recognition (NER) (Babych and Hartley, 2003), and (iii) agreement between word classes (Frank et al., 2004). With substantial advances in quality of MT outputs, the focus has shifted to other aspects, such as gender bias. Such studies range from the analysis of commercially available systems (e.g. Prates et al., 2020) to the creation of (binary) benchmark datasets for MT (e.g. Cho et al., 2019). Several approaches for automatically debiasing MT have been proposed, from different types of gender tagging (e.g. Vanmassenhove et al., 2018) to domain adaptation (e.g. Saunders and Byrne, 2020). A recent interesting comparison establishes the effectiveness of domain adaptation after training in comparison to down-sampling, upsampling and counterfactual augmentation before training (Tomalin et al., 2021), however, only testing on a binary dataset. Savoldi et al. (2021) provide a comprehensive overview of gender bias in MT in general.

Lately, non-binary\(^1\) individuals have been gaining greater visibility in public discourse. For instance, TV series, such as Sex Education and Transparent, feature non-binary characters, and famous celebrities, such as Sam Smith and Demi Lovato, have come out as non-binary. Accordingly, linguistic strategies to represent them in discourse have been introduced. These include the use of the gender-neutral pronoun hen in Swedish (Bäck et al., 2015), gender star (*) in German (Hornscheidt, 2012; Hornscheidt and Sammla, 2021), and schwa (\(\ddot{a}\)) in Italian\(^2\) (Gheno, 2019). Nevertheless, there

\(^1\)Non-binary is here used as an umbrella term for all gender identities beyond male and female.

\(^2\)https://italianoinclusivo.it

---

New Trends in Translation and Technology, pages 166–177
4–6 July, 2022. ©NeTTT 2022
still seems to be a general lack of awareness of gender identities beyond the binary and misgendering, which denotes the assignment of a wrong gender to a person, can frequently be observed in public discourse and translation (Misiek, 2020).

In TS and MT, relatively few approaches explicitly address gender-fair language beyond the binary. However, the existing approaches have to date neither been systematically collected for TS or MT nor have they been compared. Building on systematic literature analysis methods (Kitchenham, 2004; Keyes, 2018) combined with a graph-based bibliometric search method (Ammar et al., 2018), we seek to answer the following research questions:

1. Which approaches have been proposed for gender-fair translation in the field of translation studies?

2. Which approaches have been proposed for gender-fair translation in the field of machine translation?

3. How do important projects and key activists in the field view the developments?

In both fields, we consider approaches that rely on a theoretical discussion as well as approaches that have been empirically tested. With this systematic literature review on gender-fair (machine) translation from the perspective of TS and MT, we seek to provide an excellent starting point for future research endeavors towards multilingual gender-fair communication.

2 Preliminary

As a theoretical foundation for this survey, a brief introduction to the interaction of gender with language and translation will be provided, followed by an introduction to gender bias and misgendering.

2.1 Gender and Language

Gender is a biopsychosocial construct (Barker and Iantaffi, 2019) that impacts people’s lives in numerous ways, e.g. it is used to regulate access to sport competitions or public spaces such as restrooms (Fae, 2016-12-14). The relation between gender and language is complex, since linguistic structures are used to refer to the extra-linguistic reality, i.e., to the gender identity of speakers and or referents (Corbett, 1991).

These structures vary across languages, which can be classified into (i) grammatical gender, (ii) notional gender, and (iii) genderless languages (Stahlberg et al., 2007). In grammatical gender languages, such as German and Italian, each noun has a gender, other word classes are inflected correspondingly, and the assignment occurs based on formal criteria (Corbett, 1991; Comrie, 1999; Eckert and McConnell-Ginet, 2013; Stahlberg et al., 2007). In notional gender languages, such as English, pronouns and some nouns (e.g. boy/girl) are gender-specific. In both language categories, gender assignment of nouns for human referents is based on the extra-linguistic reality (Corbett, 1991; Comrie, 1999; Eckert and McConnell-Ginet, 2013), i.e., the gender identity of speakers and/or referents. This poses challenges for the linguistic representation of non-binary individuals. Genderless languages, such as Turkish, are generally neutral. Nevertheless, since gender is not merely a grammatical phenomenon but also ideological, linguistic structures and different connotations can reveal gender associations and stereotypes (Nissen, 2002).

2.2 Gender and Translation

In the translation process, the omissions of gender is frequently not an option. First, since linguistic structures vary across languages, the translation from e.g. genderless to grammatical gender languages requires choices that are not neutral (Nissen, 2002; Di Sabato and Perri, 2020). For example, literary works, such as Edgar Allan Poe’s The Tell-Tale Heart (1843), avoid gender markers to create a mysterious atmosphere and translators to languages such as Italian are faced with an obligatory choice to assign a gender to the referent (Di Sabato and Perri, 2020). This choice is often based on social gender, i.e., stereotypical associations evoked by a word, e.g. a profession (Nissen, 2002). Second, gender conveys connotations, especially in the case of personifications and metaphors (Jakobson, 1959). For instance, animals are used in advertisement to connote the sponsored item in specific ways, e.g. the male, fast, strong jaguar to represent a car. Since the same animal might have different or no gender-specific connotations in another language, translation choices are required to deliver the source text message (Di Sabato and Perri, 2020). Finally, for humans and machines alike identifying gender references in text is not an easy task, as shown in several works (e.g. López, 2021a; Šincek, 2020; Cao and Daumé III, 2020) where, for instance, a lack of recognizing the use of singular
2.3 Gender Bias and Misgendering

This paper follows a sociological perspective of viewing bias as judgements based on preconceived notions or prejudices rather than an impartial decision (Barocas et al., 2017), which equally applies to models trained on biased data. Such bias can lead to different types of harms, which on a most coarse-grained level are allocative and representational harms (Barocas et al., 2017). The former refers to withholding opportunities or resources from a specific group, while the latter refers to reinforcing the subordination of groups along the lines of identities, i.e., race, class, gender, etc. For instance, a misrepresentation or omission of non-binary genders in the language representation of modern language and MT models perpetuates representational harm (Dev et al., 2021). Such misrepresentation can also be observed in human translations, e.g. of news reports (Lardelli and Gromann, 2022). This omission of an entire group from language has a direct representational harm for the social group as well as an allocative harm, e.g. disadvantages in hiring contexts (Horvath and Szcseny, 2016). Furthermore, misgendering can lead to emotional pain and feelings of identity invalidation (Zimman, 2019).

Inherent bias in MT systems is perpetuated in an unreflecting manner. This is particularly problematic since numerous online contents are automatically translated without people being necessarily aware of consuming machine translations and their inherent misgendering (Savoldi et al., 2021; Martindale and Carpuat, 2018). The fact that MT suffers from gender bias has been shown extensively in form of generating wrong gendered forms (Sun et al., 2019; Monti, 2020), implicitly relying on gender stereotypes (e.g. Prates et al., 2020; Cho et al., 2019), and having a strong male default (Monti, 2020). Automated debiasing techniques to reduce or at best eliminate gender bias in MT include gender tagging (Vanmassenhove et al., 2018; Stafanovićs et al., 2020), retraining with debiased word-embeddings (Escudé Font and Costa-jussà, 2019) and domain adaptation (Saunders and Byrne, 2020). With the exception of Saunders et al. (2020), these studies implicitly or explicitly focus on a binary view of gender.

3 Survey Methodology

To conduct a systematic survey on gender-fair (machine) translation, we combine the systematic review methodology proposed by Kitchenham (2004) and the literature analysis method of vision-based automatic gender recognition approaches put forward by Keyes (2018) with a literature-based bibliographic analysis tool Connected Papers\(^3\). Since this is such a novel research area, we decided to additionally include online representations of well-known individuals and activists that might not (yet) be quality-validated by means of peer-reviewing but have been validated by two experts in the field.

3.1 Search Strategy

The major objective of this survey is to systematically identify, analyse and compare approaches to multilingual gender-fair communication in the fields of translation studies and machine translation. To investigate the research questions introduced in Section 1, we formulated keywords to be submitted to a total of seven electronic scientific databases as well as a conventional search engine, i.e., Google. The electronic databases serve the purpose to identify scientific literature on the topic and are general (Google Scholar, Web of Science and Scopus), translation studies (Bibliography of Interpreting and Translation (BITRA) and Translation Studies Bibliography), computational linguistics (ACL Anthology), and computer science (DBPL) platforms. A conventional search engine was included in the search strategy in order to identify key research projects and online platforms of key players in the field of gender-fair communication and to investigate our third research question. In total, 23 keywords, e.g. [“gender-inclusive translation”], and keyword combinations, e.g. [“misgendering” AND “machine translation”], were submitted to the above platforms listed in Appendix A. Decisions to include or exclude individual publications were based on the identified eligibility criteria as well as on the consensus of both authors. The time period of the search was set to publications between 2009 and July 2022. Finally, the search strategy included snowballing, that is, consulting important references of publications in our result set, in order to ensure that important papers outside the time period would be considered.

\(^3\)https://www.connectedpapers.com
Papers deemed eligible from the first step were submitted to the tool Connected Papers in order to construct and analyse literature graphs as a complementary search strategy to ensure no key publications were overlooked. The graph is constructed by utilising a similarity metric based on co-citation and bibliographic coupling (Ammar et al., 2018), that is, analysing overlaps in citations and references of papers in the Semantic Scholar Paper Corpus. The underlying algorithm builds a force directed graph that visually clusters more similar papers closer together.

### 3.2 Eligibility Criteria

Given the linguistic capabilities of the two researchers performing this literature review, publications in the five languages English, German, Italian, Spanish, and French could be considered. Studies published in one of these languages were included in the result set of this survey based on the following eligibility criteria: (i) explicitly addressing gender beyond the binary in translation or (ii) explicitly addressing gender beyond the binary in machine translation, and (iii) proposing, discussing, or analysing translation strategies and/or debiasing methods for gender-fair language. Gender concepts in translation and machine translation only recently transitioned to consider gender beyond the binary. Thus, the eligibility criteria were intended to be as little restrictive as possible.

### 4 Results of Survey

From 48 search results, 31 items were included in the result set, three resulting from Connected Papers. Apart from compliance to the eligibility criteria, reasons to exclude items were the format (YouTube videos, forums, etc.), only briefly mentioning gender-fair language and/or translation, or a much broader focus, such as non-binary representation in language technology. This resulted in 21 English and 10 Spanish publications that could be considered. Since we tried to be as inclusive as possible, the result set included scientific publications (12), preprints (4), blog posts (6), websites (3), and bachelor’s or master’s theses (6). The vast majority of approaches focused on analysing translations from English to other languages, including Spanish, German, Croatian, Czech, and Polish. A detailed overview of the works in our result set is presented in Sections 5, 6, and 7.

### 5 Gender-Fair Translation

In TS, the majority of the studies focus on the analysis of audio-visual translations, e.g. subtitling or dubbing, from English into other languages. Misiek (2020) analysed the Polish dubbed version of three different English TV series featuring non-binary characters and found that translators often misgendered the characters or even completely erased their gender identity. Furthermore, gender-specific differences between English and Polish are detailed and translation strategies to avoid misgendering non-binary individuals are proposed, which include (i) asking the person interpreting the character for their preferred gender representation, (ii) consulting gender queer and non-binary communities, (iii) using Polish neuter or $u$, and (iv) using neologisms or gender-neutral words, such as *person*. Similarly, Šincek (2020) analysed the Croatian dubbed version of a movie as well as articles on Sam Smith’s coming out as non-binary. In this context, misgendering could frequently be observed, e.g. mistranslating singular *they to oni* (third person plural masculine), and Smith was either referred to with the third person singular pronoun *ono* (English *it*). A very extreme case is shown by the analysis of the English translation of the Japanese series Sailor Moon (Hoskin, 2018), in which queer characters, relationships and any concept of gender fluidity were censored, basically rewriting the original to fit a Western dichotomous conception of gender at the time of the translation.

Work by López (2019c; 2020; 2019a; 2021b; 2019b; 2021a) as well as Attig and López (2020) focuses on the translation of media products, mostly the dubbed as well as the subtitled version of English language TV series, such as *One Day at a Time*, from English into Spanish that feature non-binary characters. According to them, translators need to be able to use non-binary gender-fair language to both respect non-binary people as well as not to alter the message of the source text. Nevertheless, offensive or discriminatory words should be avoided even if present in the source text, e.g. “transsexual” which is a medical term used to pathologise trans and non-binary people. Finally, López distinguishes between Indirect Non-Binary Language (INL) and Direct Non-Binary Language (DNL). In the former, sentences are structured in a way that gender can be avoided, e.g. using passive constructions and gender-neutral terms, such
as *person*. This strategy should be used when texts address a broad audience that include different genders. In the latter, new gendered endings, e.g. *e* in Spanish, or typographical characters, e.g. @, may be used in order to specifically address non-binary individuals. Spoljar (2018) exemplified and analysed the use of Spanish *e* in the subbing of *One Day at a Time* as well as one web series and podcast. The focus is clearly on gender-inclusive Spanish language strategies and the necessity to truly understand the situation before translating is emphasized. This plea is reiterated by Brooks (2017), who emphasises the need for translators to not only understand the source text with all its implications but also pay attention to offensive terms ad conceptual differences between languages.

Moreover, *language variety* and inherent differences in gender-fair language use should be taken into account. Considerable differences between the Spanish and Latin American subtitles and dubbed versions in their reference to non-binary characters, e.g. *elle* (ES) vs. *ze* (LatAm), and frequency and forms of misgendering, e.g. *ellos* (plural male them) and *stuyos* (plural male belongings), could be observed (e.g. López, 2019b, 2021a).

Šincek (2020) also investigated how *non-binary people express their gender in language* by conducting three interviews. The main problem is that there are no specific strategies to gender-fairness in Croatian. With Šincek (2020) as starting point, Špoljar (2021) conducted similar interviews with eleven non-binary participants, finding that they (i) use binary forms according to their sex assigned at birth, (ii) switch gendered forms, (iii) use the third personal plural masculine pronoun, or (iv) avoid tenses and grammatical constructions that require gendering. Wehle (2020) provides an overview to gender-fair language in Czech, a fusional language in which it is difficult to refer to people without assigning them a binary gender. Nine students selected three different strategies to translate English novel extracts about non-binary individuals: (i) the use of the gender assigned at birth, (ii) the neuter grammatical gender and (iii) using orthographic symbols (e.g. asterisk) and adjustments (e.g. introducing gender-fair diacritics). In a subsequent questionnaire on the perception of these strategies, 31 non-binary community members showed a strong preference for strategy (iii).

The importance to *rely on the non-binary community* to produce gender-fair translation is equally highlighted by Attig and López (2020). Caudle (2019) translated a futuristic novel on a gender-nonconforming protagonist from Korean, a genderless language, into English. Translation strategies comprise the creation of a new term for the protagonist’s gender and correspondingly introducing new pronouns. The translator also highlights the importance of speaking with the author of the original work as well as having a knowledge of the queer source and target culture.

Besides public translations, studies have been conducted to investigate the language use and *attitude of translation and interpreting students* towards gender-fair language, which conventionally consist of a translation experiment and a questionnaire. Bailach Adsuara (2021) conducted an experiment with two students who were asked to translate the synopsis of an English language novel on a gender-fluid character to Spanish. In the translation, both displayed a preference for INL, which inevitably leads to a loss of information from the source to the target text. The participant who indicated more prior knowledge on the topic stated that in other circumstances they would have utilised DNL. Nevertheless, the experiment and questionnaire clearly showed a lack of awareness of the impact translators might have on the use of gender-fair language. In Pulgar et al. (2021), eleven students showed a more varied use of INL and DNL to represent non-binary identities. However, misgendering could be observed and especially in reference to mixed-gender groups the use of male generics in the translation was predominant. In Lópe León (2020), most of the twelve students utilised gender-fair language, only one used male forms and one person failed to recognise the English neo-pronouns “xe/xyr” (alternatives to singular “they/them”), interpreting them as proper nouns. The choice of translation strategy seemed to be correlated with the students’ prior knowledge of the community and language strategies. Only a third deemed the use of gender-fair language necessary when not officially accepted by language institutions and less than a fifth believes that gender-fair texts read naturally.

The impact of *prior knowledge and intent* in the use of gender-fair language in translation has been shown to affect attitudes towards the topic. Llanos Guerrero (2019) investigated whether delivering prior information on the utilised translation strategies as well as on the translator’s ideological
positioning increases the acceptance of gender-fair translations in readers. To this end, two translations of a French language comic, one using binary gender-fair language and the other using non-binary gender-fair language, were presented with such prior or with no information during four group interviews. Results show that a more positive reception of the target text is in fact achieved when prior information is delivered. One attempt to provide such explanations has been undertaken by Sobrien (2020), who translated an English website on the use of gender pronouns to Spanish, highlighting the importance of sharing information across communities and languages. Strategies from INL to DNL where used, including the suffix e and the pronoun elle. Di Biase et al. (2021) observe that in video games Spanish users are addressed with male generics. To counteract this, they propose a glossary for gender-fair localisation of English video games to Spanish.

Martínez Pagán (2020) discusses gender-fair language strategies for English and Spanish for INL and DNL and debates ethical aspects of the profession of translator, with the ability to harm or empower, and its political dimension. For instance, the case of handling discriminatory source texts is discussed, where the temporal, political, etc. context should clearly be taken into account and the client should be informed. Martínez Pagán (2020) warns that the role of the translator should not be to perform pinkwashing, whitewashing, genderwashing nor to conceal oppressive realities by “cleaning” discriminatory elements in the source text.

Gomez (2020) investigates in an online survey in different countries whether personal ideologies in regard to gender issues affect translation strategies. The 112 participants, including translation students and professional translators, were asked for their political/ideological orientation, religious beliefs, their country’s attitude towards homosexuality as well as if and how they would translate a Swedish book featuring a non-binary child. It emerged that professional ethics has a stronger impact than personal ideology when selecting the translation strategies. Only about five percent of the respondents would refuse to translate books on gender issues.

Within the context of sign language, Löffler et al. (2020) analyse the differences between German Sign Language (DGS) and German in order to produce a corpus-based dictionary for their translation. While DGS is gender-neutral, German clearly requires gender marking. Nevertheless, gender markings can be introduced to DGS lexically, by using the sign for man or woman, or morphologically, by adding the concept of female to the sign. The question is, however, whether the signer knows about the gender of one or more referents or their interpretation of gender is correct. In conclusion, the authors highlight the importance of transparency on decisions taken during signing when it comes to gender, since it represents a highly complex topic.

6 Gender-Fair Machine Translation

Challenges faced by human translators in terms of translating from genderless to grammatical or notional gender languages are shared by machine translation approaches. In general, the mapping of one-to-many forms is challenging for machine translation, and so are one-to-many gender predictions (Savoldi et al., 2021). A first proof of concept to address gender-fair Neural Machine Translation (NMT) beyond the binary has been proposed by Saunders et al. (2020), where a gender-balanced corpus (Saunders and Byrne, 2020) was extended by gender-neutral sentences with placeholders for gender inflections in German and French to train an NMT model. To test the effectiveness of their methods, they finally produced a gender neutral version of WinoMT (Stanovsky et al., 2019). While these experiments are very interesting, a low overall accuracy and a tendency to over-generalise to use exclusively gender-neutral language, even if the original text was clearly gendered, could be observed. Cho et al. (2019) collect gender-specific and gender-neutral Korean sentiment- and occupation-related sentences and their English translations to test the ability of MT systems to preserve gender neutrality.

Another way of achieving gender-fair language, even with machine translation, would be to perform rewriting (Vanmassenhove et al., 2021; Sun et al., 2021), that is, intralingual translations from gender-exclusive to gender-fair language. This could be considered a form of gender-specific automated post-editing of machine translation, however, might be difficult if the target text is very far from the source text, e.g. misinterpretation of singular they to refer to a group of individuals.

In the field of video-game localization, Theroine et al. (2022) compile an English to French corpus with games that also feature transgender and non-binary individuals and annotated it together with another corpus that was artificially augmented

171
by rewriting binary pronouns into non-binary pronouns. These corpora were used to train four NMT systems using the pre-trained text-to-text transformer model T5 (Raffel et al., 2020). The first preliminary results are modest but further research is currently being conducted.

7 Gender-Fair (Machine) Translation

In German, many strategies to gender-fair language have been proposed (e.g. Hornscheidt, 2012; Hornscheidt and Sammla, 2021). Nevertheless, there are no studies on their use in human translation and in machine translation, with the exception of Saunders et al. (2020). Therefore, the GenderFairMT project organised a three-day participatory workshop to consult translators, MT experts and members of the non-binary community in order to find an approach for English into German translation of gender-fair language that could be successively implemented in machine translation.

The results of this workshop are summarised in Burtscher et al. (2022) and Gromann4, and highlight the clear message that there is not going to be the one language strategy for German gender-fair translations or one-size-fits-all. Depending on the context and individuals involved, different strategies will need to be applied. Furthermore, Lardelli et al. (2021) note how the selection of a strategy in translation is extremely context-dependent. The author’s intention, target audience, text type, referenced genders (e.g. non-binary genders or mixed-gender groups) must be taken into account when translating or when post-editing. As a consequence, there is a need for MT approaches that enable a customization of language use in translation. To truly find a viable solution, the topic should be investigated with an inter-/trans-disciplinary multi-professional team, such as proposed and lived in the GenderFairMT project.

To combine TS and MT, Lardelli et al. (2021) organised a challenge on post-editing of machine translated texts that had been written both in German and in English with gender-fair language during a hackathon. Participants’ use of strategies differed amongst teams and their documented choices confirm that there is no one-size-fits-all approach to the translation or post-editing of gender-fair language. Common criteria to strategy selection include their readability, acceptability, and feasibility.

With the objective to develop a tool to reduce biases in texts, the Artificially Correct project6 has been gathering different stakeholders, such as translators, activists and language technology experts. Specifically, events such as webinars and hackathons are organised to create awareness on different types of biases and promote discussion among experts and laypeople, including gender bias beyond the binary.

8 Challenges and Cross-Community Strategies

Approaches from the field of TS show that the topic of gender-fair translation spans from studies of applied strategies, including studies on sign language, to investigations of people’s perceptions and ideologies. Major challenges from this perspective are the detection of non-binary gender references, such as the English singular they, or English neo-pronouns, such as xe, in order to even consciously take the decision to apply gender-fair language strategies and a lack of awareness of the role of translators as mediators between cultures and communities. One common strategy to overcome the issue of misgendering is to utilise gender-neutral language, however, in not all natural languages gender-neutral strategies are always feasible and its application inevitably leads to a loss of gender-specific information from source to target text. The call to actively involve the non-binary and queer community in the development of translation studies has been reiterated in several publications (e.g. Attig and López, 2020; Šincek, 2020).

In the field of MT, many experts acknowledge that non-binary genders are still neglected (e.g. Savoldi et al., 2021; Costa-jussà and de Jorge, 2020; Stefanović et al., 2020; Choubey et al., 2021). The proposed practical approaches in our result set range from gender-neutral tagging to intralingual rewriting of MT output. One key challenge is the lack of available datasets for training and testing methods for gender-fair machine translation. While Tomalin et al. (2021) focus on binary debiasing, they still argue that strongly gender-inflected languages lack conventions of non-binary inflections and that this is the reason why options for debiasing NMT beyond the binary is “severely limited”. This

---

4https://genderfair.univie.ac.at/index_en.html
5https://genderfair.univie.ac.at/index_en.html
6https://www.goethe.de/prj/one/en/aco.html

172
lack of conventions could also be observed in this survey, however, it could also be shown that gender-fair translation requires a more refined translation strategy depending on many factors, laid out below, than currently considered in machine translation approaches. Thus, it might be time to question the most fundamental MT paradigm of one input and one output based on linguistic context. Google Translate has recently started to provide a male and female output for a word or sentence where the gender is unclear⁷. This development supports the idea that the traditional one-to-one MT paradigm might have run its course.

As it emerges from the literature in the field of TS (e.g. López, 2019c; Lardelli et al., 2021) the selection of a gender-fair language strategy is extremely context-dependent. Translators must, for example, mediate among source text intention, target culture and audience and the genders addressed in the source text. In some languages, such as Spanish, the non-binary community seems to largely adopt one solution (elle), whereas in others, such as German, many approaches to gender-fairness have been proposed (Hornscheidt and Sammla, 2021). Furthermore, non-binary individuals have different preferences for which language to be used (Zimmerman, 2019) and people frequently have ideological presuppositions regarding gender. In regards to the latter, more information on translation strategies has been shown to achieve improvements.

We thus argue for both fields, TS and MT, that the quintessential implication of this comprehensive literature review is that the one (machine) translation strategy beyond the binary does not exist and due to many factors could never exist. Instead it is vital to perform a deep analysis of the source text regarding references to people and potential offensive and discriminatory language. For references to people, additional knowledge not available from the linguistic context of the source text might be required. For instance, knowledge of the fact that a specific person referenced in the text publicly identifies with the pronoun nin in German is indicative of the gender-fair language strategy to be used for this person in any translation to German. Simply resorting to a gender-neutral solution in all contexts is neither viable nor desirable and easily leads to a crucial loss of information. Instead, both TS and MT should consider adapting to the multiplicity of gender representations reflected in modern language use, as shown in numerous TV series and newspaper articles. A more flexible gender-fair (machine) translation strategy might implicate (i) the active inclusion of the non-binary communities across languages, (ii) explicit explanations of (machine) translation decisions on gender-specific translations, (iii) a highly informed method for identifying gender in natural language by humans and/or machines, and (iv) full awareness of the role of a translator, human or machine, as mediator between cultures and communities. This implies the inclusion of world knowledge, e.g. how a person refers to themselves, in the translation process and awareness that there might be more than one target text to be negotiated with clients and the non-binary community.

9 Conclusion

Gender-fair (machine) translation is a very complex yet important topic. First, misgendering can lead to emotional pain and feelings of identity invalidation for concerned individuals. Second, language has a very real impact on society and its perception of social reality. Third, (machine) translators have the important role of mediating between cultures and communities, thus, obtaining the chance to reduce or propagate gender bias. At the same time, the role of the translator, human or machine, should not be to conceal discrimination in source texts. This systematic literature review aimed at systematically and critically evaluating existing approaches to gender-fair (machine) translation and the foremost conclusion is that the one generally acceptable and widely applicable solution does not and could not exist.

In terms of practical implications for the field of TS and MT, a deep analysis of references to people and potentially discriminatory language use in the source text has evolved, with a strong recommendation to involve the non-binary community and world knowledge, e.g. identified pronouns of people, in the translation process. More gender-fair datasets and even only text samples would be needed across languages to understand the use of these strategies. Since prior knowledge on the topic and personal ideologies play a role in the acceptability of these strategies, these texts could serve as a basis to conduct studies on these criteria as well as readability and comprehensibility. The question of how machine translation can be debiased
beyond the binary and beyond strategies to automatically rewrite target texts, a form of gender-fair post-editing, is yet an open research question for which this review seeks to provide a basis.

References


Jane Fae. 2016-12-14. Non-gendered pronouns are progress for trans and non-trans people alike. The Guardian.


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


Ártemis López. 2021b. “they” is not“they”, it is “elle”: the challenges of non-binary language for translators.


Edgar Allan Poe. 1843. The tell-tale heart. Penguin UK.


Dominican Republic. Association for Computational Linguistics.


A Appendix: Keywords for Bibliographic Search

In this appendix, we provide a complete list of keywords and keyword combinations utilised for the bibliographic search of this survey. This list was compiled by two experts in the field of gender-fair (machine) translation and extended by keywords uncovered in important publications on the topic.

- “gender-fair” AND “translation”
- “gender-fair translation”
- “misgendering” AND “translation”
- “gender-inclusive translation”
- “gender-neutral translation”
- “gender-inclusive” AND “translation”
- “gender-neutral” AND “translation”
- “non-binary language”
- “non-binary translation”
- “non-binary language” AND “translation”
- “gender-fair” AND “machine translation”
- “misgendering” AND “machine translation”
- “gender-inclusive machine translation”
- “gender-neutral machine translation”
- “gender-inclusive” AND “machine translation”
- “gender-neutral” AND “machine translation”
- “non-binary” AND “gender bias” AND “machine translation”
- “non-binary gender” AND “translation”
- “non-binary gender bias” AND “translation”
- “non-binary inflection” AND “translation”
- “non-binary inflection” AND “machine translation”
- “non-binary coreference” AND “translation”
- “non-binary coreference” AND “machine translation”