03

Neural machine translation and a queer perspective on gender bias

A qualitative study of how different strategies of *écriture inclusive* are translated into German by DeepL and Google Translate

03

Neural machine translation and a queer perspective on gender bias

A qualitative study of how different strategies of *écriture inclusive* are translated into German by DeepL and Google Translate

Andrea Chagas López, Hannah Hilß & Sebastian Müller

The issue of gender in translation has garnered significant attention since the establishment of (human aided) machine translation as previously existing gender biases in natural language are now being reproduced by translation machines. Although machine translation has a long history, recent advancements, such as neural machine translation (NMT), have revolutionized the field. NMT systems rely on training algorithms and large corpora which are influenced by human choices that can be negatively biased regarding gender, race, etc. In the context of written French, *écriture inclusive* strategies seek to establish gender-inclusive alternatives to promote gender equality in language use. However, the debate on gender and inclusive language still predominantly focuses on binary gender representations. This study explores how Google Translate and DeepL handle *écriture inclusive* strategies when they are translated into German. Three main aspects will be directly addressed in this section. First, we will take a look at the common translation practices offered by the machines regarding sentences in *écriture inclusive*; second, we will examine the target term strategies that differ from the source language's; and third, we will examine the instances where translations incorporate genders beyond the binary. We therefore aim to investigate how machine translation systems, specifically Google Translate and DeepL, perform in these cases. In this article, we argue that the absence of ethical frameworks for AI and data training has resulted in the reinforcement of gender biases and representational harms within machine translation systems.

1 Introduction

The issue of gender in translation began attracting scholarly attention in the 1980s, particularly among feminist translation theorists, and has since evolved and diversified. Grammatical gender systems vary significantly across languages, requiring both machine translation systems and human translators to make deliberate and informed choices when selecting lexemes that encode gender. This process demands close attention to the grammatical structures of the target language in order to appropriately select lexemes that correspond to specific social gender categories and to match them with suitable translation equivalents in each linguistic context (cf. Di Sabato & Perri 2020: 367). In the past, translation decisions have predominantly favored the use of masculine forms, which nowadays continue to be employed by a large segment of the population in a presumed gender-neutral manner.

The key aim of Germany's second-wave feminist movement, which emerged in the 1970s, was to advance gender equality in society. A central strategy of this movement was to pursue gender equality by promoting linguistic equality, thereby challenging the generic use of masculine forms and increasing the visibility of women through the explicit use of feminine grammatical forms. Contemporary queer feminist movements have extended these efforts, advocating for inclusive language that transcends the binary gender framework. Unlike earlier feminist approaches, queer feminist perspectives seek to foster gender equality by incorporating non-binary and gender-diverse categories into linguistic practices. Issues regarding gender remain a subject of contentious debate in broader social contexts, in research and more specifically in translation studies. Sun et al. (2019) and Prates et al. (2020), for example, have shown the need for this kind of awareness targeting gender bias in automated systems such as Google Translate. Prates et al. (2020) highlight how sentences originating from gender-neutral languages often yield translations that perpetuate stereotypical gender roles, with a marked tendency to default to male classifications over female ones. For instance, when translating from a gender-neutral language, Google Translate commonly produces results such as “He is an engineer” rather than “She is an engineer,” mirroring societal biases that associate engineering and related professions predominantly with men. This tendency is particularly salient in fields such as STEM (Science, Technology, Engineering, and Mathematics) which have traditionally been male dominated. In other words, the gender biases are expressed through linguistic practices and therefore are likely to be present in AI tools and neural machine translation (NMT). Studies on automated translation have demonstrated that gendered language can perpetuate stereotypes, thereby highlighting the growing importance of gender-inclusive language use. These findings prompt critical inquiry into how machine translation systems perform in comparison to human translators, who already grapple with the complexities of gender representation in language. While human translators are often equipped to make context-sensitive decisions, machine translation systems operate on computational strategies that have evolved significantly over time. Although the concept of translation machines dates back to the early days of modern digital computing, recent technological advancements have led to the development of systems based on neural machine translation (NMT), e.g. those used by DeepL and Google Translate. These systems employ artificial neural networks—modeled on the structural and functional characteristics of the human brain—to implement machine learning algorithms. Through exposure to large and diverse corpora, NMT systems learn to apply a variety of translation strategies, yet they remain susceptible to reproducing the gender biases embedded in their training data. A clear example can be found in Caliskan et al. (2017: 183) who analyzed whether semantics derived automatically from language corpora contain human biases, finding that “human-like semantic biases result from the application of standard machine learning to ordinary language—the same sort of language humans are exposed to every day.” Therefore, machine translators that are trained to learn from corpora will likely reproduce “historic biases” (ibid.) that were displayed through language use in the first place. Regarding the biases studied in this paper, we follow the argumentation by Tomalin et al. (2021: 420) who claim that gender biases and representational harms, such as stereotyping, exist due to datasets that demonstrate an asymmetric and/or prejudiced societal representation in which a white, cis-gendered male is defined as the norm.

Inclusive language use takes on varied forms and strategies across languages, which presents a problem not just in machine translation (MT) but also concerning the stereotypes and biases in the aforementioned datasets. In written French, *écriture inclusive* strategies aim to establish gender-inclusive alternatives (cf. Viennot 2020: 149).

Based on our own corpus containing sentences in *écriture inclusive* in French and their German translations, we aim to explore the following interrelated research questions: How do Google Translate and DeepL translate the diverse strategies of *écriture inclusive* into German? Which strategies are most commonly found in the translation from French to German? Are there any tendencies that are applied to the translations in the target language (German) that differ from the strategies in the source language (French)? Can we identify instances in German where the translations incorporate genders beyond the binary?

To comprehensively examine the established subject matters, we begin by analyzing the various grammatical and morphological structures referring to social gender in section 2, as well as inclusive language practices in both French and German. Section 3 addresses the influence of grammatical gender on machine translation and the technologies utilized in this field and inspects the persisting gender biases more closely. Section 4 describes the methodology of building and evaluating a corpus of French sentences in *écriture inclusive* for translation into German. Section 5 provides an analysis of the translations produced by Google Translate and DeepL, focusing on the translation strategies used for *écriture inclusive* to German and the assumed biases behind them. Finally, section 6 presents the conclusion.

2 Gender-inclusive language – grammatical and morphological differences between German and French

2.1 Gender markers in German and French

In German, all nouns have a specific grammatical gender. The grammatical gender system consists of a tripartite gender division that includes a feminine, a masculine, and a neuter gender. The grammatical gender of words for animate beings, however, does not necessarily correlate with the social gender (and/or sex). For example, *das Mädchen* ('the girl') is neuter in grammatical gender even though the word's meaning addresses girls. This is the direct result of the diminutive suffix *'-chen'*, which consistently produces neuter nouns regardless of the social gender lexically assigned in lexemes for animate beings. Nevertheless, a strong correlation between grammatical and social gender does exist, with only a few exceptions, where the correlation can be explained by semantics. Nübling (2020: 10) illustrates this close relationship between grammatical gender and sex (and/or social gender) through the example of the so-called *Differentialgenus*. When a German noun referring to a person is derived from a nominalized adjective or participle, the only grammatical gender marker is the article (or other determinants, if present). These associated function words serve primarily (but not exclusively) a grammatical function by marking syntactic agreement with the noun, in line with Hockett's definition: "Genders are classes of nouns reflected in the behavior of associated words" (1958: 231).[1] In such cases, grammatical gender and social gender are congruent. The noun itself does not suffice to identify its grammatical gender, as the grammatical category is expressed externally through these agreement markers. In plural, the associated function words no longer exhibit gender-specific marking, since the co-occurring number inflection neutralizes morphological gender distinctions. This phenomenon can be attributed to phonological coincidence whereby the diverse plural inflections underwent phonological simplifications, resulting in the current syncretism. Rather than representing a deliberate grammatical feature, this phonological coincidence is

[1] For example, in *die/der Studierende* ('the student', f./m.), only the article reveals the reference to the social gender in these cases. In plural, on the other hand, the article does not reveal any reference to the social gender, e.g., *die Studierenden* ('the students').

responsible for the absence of gender differentiation in plural forms (cf. Schwarze 2009: 111).

Moreover, personal pronouns (e.g., *er* 'he', *sie* 'she', *es* 'it') play a significant role in the expression of gender. As (co-)referential devices, they reflect the word's grammatical gender (cf. Busch & Stenschke 2018: 117). When referring to a person, they are also used to index social gender (cf. Kotthoff & Nübling 2018: 148).[2]

In summary, grammatical gender isn't necessarily congruent with the social gender of the referred person, but there is, in many cases, consistency between the grammatical gender of an animate noun and the social gender of its referent. This becomes especially obvious when talking about personal nouns like *die Mutter* ('the mother', f.) or *der Vater* ('the father', m.), in which grammatical gender reflects social gender, and which even carry gender stereotypes of their own.

French, on the other hand, has a binary grammatical gender system. French nouns are either grammatically masculine or feminine. Still, there are homonyms that may refer to either a man or a woman depending on their grammatical gender as reflected in accompanying function words. For example, this can frequently be found in invariant occupational titles such as *le ministre* ('minister', m.) and *la ministre* ('minister', f.). This is an example for *genre commun* or *Differentialgenus* as explained above. In French, the grammatical gender of a noun is not only marked by accompanying function words such as articles and pronouns, but also through the declension of attributive and predicative adjectives.

French determinants and pronouns[3] are also inflected according to the grammatical gender of the noun. For example, the demonstrative pronouns *celui/ceux* ('the one(s)', m., sing./pl.) and the demonstrative determinants *ce/cet* ('this', m.) have masculine grammatical gender while the demonstrative pronouns *celle/celles* (f., sing./pl.) and the demonstrative determinant *cette* (f., sing.) are grammatically feminine. The gender opposition is neutralized in the demonstrative plural determinant form *ces* (m./f., pl.). Other neutralizations occur, among other examples, before feminine nouns that begin with a vowel, where the masculine form is used (*mon amie* 'my friend', f. but not **ma amie*). Ordinal numbers and predicative or attributive adjectives are also inflected in agreement with the gender of the noun. For instance, *l'homme est beau* 'the man is beautiful' and *la femme est belle* 'the woman is beautiful'.

---

[2] We want to point out that personal pronouns are also used by individuals whose gender identity falls outside the cis-binary classification. In particular, people who do not identify with either of the two established binary gender categories are confronted with the absence of a neutral pronoun that can refer to human beings in German (as the neutral pronoun *es* is restricted to non-human referents). This includes many non-binary individuals, though not necessarily all trans people, some of whom identify with either the male or female gender and therefore use the corresponding binary pronouns. To address this gap, neopronouns such as *sier, xier,* or *nin* have emerged, alongside borrowed English pronouns such as *they*, which can refer not only to plural referents but also to singular individuals of unknown or non-binary gender (cf. Di Sabato and Perri 2020: 367). As a result, neopronouns like *dey*, derived from the English *they*, have been established to refer to non-binary individuals in German (cf. Verein für Geschlechtsneutrales Deutsch e.V.). These neopronouns, in turn, require the same morphological verb forms as those used for binary pronouns such as *he* and *she*.

[3] As opposed to German, there is no neuter grammatical gender in French that could be applied as such regarding social gender. Both in the singular *(elle, il*) and in the plural *(elles, ils)*, the personal pronouns indicate social gender. Therefore, the neopronoun *iel* in the singular and *iels* in the plural were suggested to refer to groups of mixed genders (cf. *Haut Conseil à l'Egalité entre les femmes et les hommes* 2023: 21).

2.2 Gender-inclusive language strategies in German and French

In both German and French, different strategies can be applied to avoid gender marking or to explicitly mark social gender linguistically. It is important to highlight that the different strategies carry strong ideological associations according to the markers being used and the gender categories included in the reference of the selected variant (Sarrasin et al. 2012). For instance, a double mention of the female and male variants of a lexeme reproduces the established binary understanding of gender. We will expand upon this topic later in the section.

We previously introduced the grammatical features of German used to mark gender, as well as the parts of speech that can carry gender markers. Function words, such as articles, always agree in grammatical gender with the noun they accompany. In cases where a noun referring to a human being does not itself indicate grammatical gender, the feminine article form is used exclusively to refer to individuals of the female gender. This is always true for the *Differentialgenus*, which often occurs in designations for individuals derived from nominalized adjectives or participles: *die Schöne* ('the beautiful woman') vs. *der Schöne* ('the handsome man') (cf. Nübling 2020).

Regarding the strategies for German that explicitly mark gender – according to a binary understanding of gender – feminists set their goal to linguistically highlight the visibility of women in the 1970s. This led to the use of paired forms like *Arzt und Ärztin* ('doctor', m./f.) or *Lehrer und Lehrerin* ('teacher', m./f.). In many cases, the feminine form is placed first purposefully like in *Damen und Herren* ('ladies and gentlemen'), stemming from historical developments.[4] The explicit use of feminine forms serves communicative functions, such as signaling the speaker's intention to refer to both women and men. These paired forms were subsequently abbreviated using parentheses or slashes (e.g. *Lehrer(in)* or *Lehrer/in*) to indicate both genders simultaneously. Another strategy capitalizes the *'-I'* of the morpheme *{-in}* utilized to mark the feminine gender (e.g. *LehrerIn*). Pusch (1979) also suggested replacing the masculine generic with a feminine generic, arguing that the lexemes inflected to mark female gender morphologically include both the masculine and the feminine markers (e.g., *Leser* is part of *Leserin*, 'reader'). Strategies that mark gender beyond the binary can be found in contemporary queer feminist movements that introduced the so-called gender-gap. This gap is made visible by using special characters such as an underscore (_, e.g., *Lehrer_in*), an asterisk or gender star (*, e.g., *Lehrer*in*), or the colon (:, e.g., *Lehrer:in*) before the suffix marking female gender (cf. Kotthoff & Nübling 2018: 218). The use of special characters signals a break with binarity by creating a visible disruption within the word itself. This form includes the representation of male and female gender and through the disruption, it leaves room for all possible gender categories. These solutions therefore make visible and create a gap between the poles of the binary gender system in order to account for the continuum of social genders that exist beyond this division.

[4] For a closer look on this behalf and the (ir)reversibility of gender order in binomials in German see Rosar (2022).

In French, as in German, different strategies[5] are used to create more gender-inclusive language forms. In France, in contrast to Germany and other Francophone countries like Canada, the form of *écriture inclusive* is still more gender-equitable rather than gender-inclusive (Dupuy 2020: 2), meaning that the focus is mainly on adding female representation in language but not to encompass all genders. The use of feminine job titles has been an important battleground in the fight for binary gender equality in France. Their usage has been gradually increasing in the last 30 years (cf. Viennot 2020). However, there are reservations and even certain restrictions[6], primarily regarding the use of feminine versions of job titles for positions of power or for professions that have been historically held by men. Consequently, feminine forms are still subject to controversy, and their usage is still not normalized. The French Academy has only approved of some of the female word forms as of 2019 (cf. Lessinger 2020: 380).

One approach to gender-fair language, referring only to individuals of both traditional genders, is to mention both the masculine and feminine forms *(doublets): les étudiants et les étudiantes* ('students', m./f.). An additional practice to mark multiple genders in French involves the use of two symbols around the feminine ending, in this case the interpunct: *étudiant·e·s (point médian)*. This method was recommended in the 2016 guide by the *Haut Conseil à l'Égalité entre les femmes et les hommes* as the best way to be more gender-inclusive beyond the binary. In the updated 2023 version, this strategy is no longer suggested for the sake of an even less explicitly gendered and rather gender-neutral language as will be discussed at a later point in this section (cf. Haut Conseil à l'Égalité entre les femmes et les hommes 2023: 16). This means that the first steps towards an all-encompassing gender-inclusive and even gender-neutral language are being taken. Still, even the previously explained and often favored use of special signs is not always neutral nor practical, as each strategy is loaded with ideologically driven decisions or other (technical) restrictions.[7] The use of *doublets*, for example, acknowledges the existence of two genders but does not leave any room to include people that identify with genders beyond that conception. The use of parentheses that enclose the feminine ending, such as *-e*, for example, merely suggests a possibility of marking both genders without indicating equivalence (e.g., *les étudiant(e)s*). The hyphen (*trait d'union*) can be separated by line breaks and is already used to form compounds (e.g., *étudiant-e-s*). The slash (*barre oblique*) tends to suggest an either/or distinction, rather than the intended inclusive interpretation (e.g., *étudiant/e/s*). The period, by contrast, already functions as a punctuation mark indicating the end of a sentence. Its use can even lead to unintended effects – for example, when not enclosing the feminine suffix, as in *étudiant.es*, it may be misinterpreted as a hyperlink or as referencing the Spanish top-

[5] Additionally, these methods vary across different countries in the Francophone community (cf. Dupuy 2020). This work, however, focuses only on *écriture inclusive* used in France.

[6] In 2017, France's prime minister decreed the generic use of masculine forms while explicitly prohibiting the use of *écriture inclusive* in official government documents (Ministère de la Justice 2017), and in 2021, the use of the interpunct and other symbols was prohibited in schools (Ministère de l'Éducation nationale, de la Jeunesse et des Sports 2021).

[7] However, the use of the so-called masculine generic is also often perceived as indexing an ideological stance.

level domain *.es*. Starting the feminine ending with a capital letter is rejected by Viennot (2020) on the ground that the use of capital letters within a word is not customary and that it disproportionately emphasizes the feminine form. The use of the interpunct is to date the most progressive and inclusive way of re-writing grammatically feminine and masculine gendered words in French.

In contrast to the explicitly gendered forms introduced above, there are also neutralizing strategies akin to the sociological concept of "undoing gender" (Diewald & Nübling 2022). In this approach, forms that allow to avoid referring to social gender are preferred. In German, general terms like *Kundschaft* ('customer base') or *Publikum* ('audience') are used instead of person-referential designations like *Kunde/Kundin* ('costumer', m./f.) or *Zuschauer/Zuschauerin* ('viewer', m./f.). The employment of compound words using *-person* or *-kraft* (e.g., *Lehrperson, Lehrkraft* 'teacher') is another strategy. Additionally, designations for individuals are formed by nominalizing attributive adjectives and the present participle forms of verbs (e.g., *Deutsche* 'German persons', *Studierende* 'persons who are doing their studies'). The latter forms are mainly suitable as gender-neutral designations when they are in the plural form, as the marking of the *Differentialgenus* in the singular form reintroduces gender. However, this can also be avoided through a (gender-inclusive) dual designation of feminine and masculine articles, such as *die/der*. Furthermore, there are various other strategies to avoid gender-marked designations, including the use of epicene words like *Person* 'person' or *Mensch* 'human' in combination with attributive adjectives (e.g., *befreundete Person* 'befriended person') (cf. Becker et al. 2022: 27, Ewels 2020: 56). Nevertheless, Klein (2022) demonstrates that seemingly gender-neutral epicene terms in person-referential nouns often function as pseudo-epicenes with their interpretation being guided by the reference context and predominantly aligned with their grammatical gender. In French, we can also find strategies which aim at undoing gender. One example is the explicit use of gender-neutral words, especially when they are used in plural and when they are accompanied by gender-neutral articles, such as *personnes* ('persons' or 'people') instead of *hommes et femmes* 'men and women'.

The differences between the strategies applied in both languages under observation result from the differences in the morphological systems. One difference between both systems is the existence of a neutral gender in German and the lack thereof in French. Regardless of this, the gender marking in plural function words is neutralized in most cases in German. On the other hand, in French, while plural articles like *les* and *des* do not mark gender, other plural elements such as adjectives and participles often retain gender distinctions. Furthermore, in French, gender marking is required in a wider range of contexts. For example, adjectives agree in gender with the noun in predicative constructions (e.g., *elle est fatiguée* 'she is tired' vs. *il est fatigué* 'he is tired'), whereas in German, they remain unmarked (*er ist müde, sie ist müde*). This, in turn, establishes a relatively strong gender specificity for French and a weaker one for German (cf. Schwarze 2009: 95).

Upon examining the various forms of gender-inclusive language and *écriture inclusive*, it becomes evident that despite substantial grammatical and morphological differences, significant commonalities exist. Across both languages, four main types can be classified: the paired mention of the feminine and masculine forms, their abbreviation through symbol integration at morpheme boundaries (representing either a binary distinction or a continuum between genders), the use of gender-neutral expressions, and the use of gender ambiguous words of the *genre commun/ Differentialgenus.*

3 Translation technologies: machine translation

3.1 Introduction to different technologies of machine translation

The term 'translation technology' generally refers to various technological means that render one expression in a language into another (cf. O'Brien & Rodriguez Vazquez 2020: 264). Two main technologies are used in machine translation nowadays.[8] The first is Computer Aided Translation (CAT) and the second is Machine Translation (MT) (cf. Werthmann & Witt 2014: 82).[9] Examples of CAT tools include Translation Management Systems, corpus analysis software, or terminology databases (cf. O'Brien & Rodriguez Vazquez 2020: 264). MT comprises different strategies using language and text data to produce translations. For instance, these strategies can be rule-based, example-based, or statistical, which is corpus-based, among other approaches (cf. Zhang & Liu 2023: 118).

Although machine translation has made significant progress over the past decades, it still faces challenges arising from the complexity of natural language (cf. Prates et al. 2020: 6364). The most recent form of statistical MT's evolution relies on corpus analysis to train the machine. The MT uses deep learning strategies that allow the technology to build neural networks inspired by the structure and function of the brain to perform translation tasks. Neural Machine Translation (NMT) proves to be a promising approach that can help solve the shortcomings of traditional translation systems. Wu et al. (2016: 1), for example, point out that the advantage of NMT

> lies in its ability to learn directly, in an end-to-end fashion, the mapping from input text to associated output text. Its architecture typically consists of two recurrent neural networks (RNNs), one to consume the input text sequence and one to generate translated output text.

Even though NMT provides a solid foundation to MT, this system also shows some weaknesses such as "its slower training and inference speed, ineffectiveness in dealing with rare words, and sometimes failure to translate all words in the source sentence"

8 The first machine translators were Rule-Based-Machine-Translation (RBMT) that are based on "linguistic resources, namely (1) bilingual dictionaries providing the morpho-syntactic and semantic information, and (2) a set of morpho-syntactic and sometimes also semantic rules for both the source and target languages" (Monti 2020: 459). The first prototypes provided a word-for-word translation based on a set of rules that parted from the linguistic structure of the target language. Further developments of RBMT include a three-step approach based on (1) analysis of the source language, (2) transfer, and (3) translation into the target language (ibid.).

9 Machine translation finds applications in various forms, both independently and in conjunction with human translation, for example, in the form of human-aided machine translation (HAMT) or machine-aided human translation (MAHT). The latter can also be grouped under the term Computer Aided Translations (CAT) (see Hutchins 1995: 431).

(ibid.). NMT predicts how likely certain linguistic structures, such as word sequences, are to appear in a context departing from patterns observed in datasets used to train it. Each word in the dataset is transcribed into a vector that receives a singular, unique value in the process of encoding and decoding. The machine therefore analyzes the source text, encodes it into vectors to then decode them to the target language. The engine predicts how correct the translation is according to the data used to train the machine. Like other kinds of MT, NMT also uses probabilistic models. Yet, NMT does not separate the input introduced into sub-categories but rather treats it as a whole. This difference allows NMT to make decisions according to the contextual information and not only on the linearity of a sentence (Choi et al. 2017). As Saunders & Byrne (2020: 7724) point out,

> natural language training data inevitably reflects biases circulating in our societies. For example, gender bias manifests itself in training data (corpus) which features more examples of men than of women. Tools trained on such data will then exhibit or even amplify the biases.

The translation of grammatical gender is one of the challenges NMT deals with when translating sentences from and/or into languages with a grammatical gender system. Sun et al. (2019) as well as Prates et al. (2020) have shown that translations of sentences that contain information about men and/or represent stereotypical gender roles are translated more accurately than cases where women are described, especially if they have a stereotypically male occupation.[10]

DeepL and Google Translate both use NMT based on AI to translate from source to target language; however, they show some key differences. DeepL draws from Linguee's corpus that consists of manually translated sentences, idioms, etc., while Google Translate originally drew from the Europarl Corpus[11] and now draws its dataset from different digital resources (Choi et al. 2017). Nevertheless, both MTs display similar limitations since they both use similar training strategies even though the data they draw upon to train their neural systems differ. Both MTs therefore struggle with the same difficulties that are endemic to AI and machine translation. NMT-generated output can be used to identify current ideologies, since the AI relies on algorithms – programmed by humans – to explore and learn from the examples in the dataset.

In the following section, we will discuss how the current circulating ideologies reflected in language use – in our case regarding gender – are embedded in the dataset used to train neural models (NM) and can therefore influence the development and also the functioning of NMT according to different demographics.

[10] The studies approaching problems related with gender biases in translation have concentrated on studying grammatical and pronominal gender. The main research areas included anaphora resolution (cf. Hardmeier 2012, Luong and Popescu-Belis 2016, Voita et al. 2018) and named entities recognition such as the classification of names of singular people, locations, etc. (cf. Babych and Hartley 2003).

[11] Europarl Corpus is a multilingual corpus that contains most of the documents published on the European Parliament's official website.

3.2 NMT and gender bias

Bias can be defined in several ways; we apply the definition introduced by Savoldi et al. (2021: 846) who take a human-centered approach to define the "motivated framing of bias" which parts from a sociolinguistic point of view: "[W]e consider as biased an MT model that systematically and unfairly discriminates against certain individuals or groups in favor of others" (ibid.). In this paper, we part from a (socio)cognitive definition of bias and recognize that biases have a (socio)cognitive function that causes people to evaluate a situation and/or a person more quickly. Hence, these biases have the cognitive function to create mental shortcuts to help people react in an interaction (cf. Tversky & Kahnemann 1973: 1974). This concept can be sustained for AI, since 'bias' in natural language processing (NLP) is conceived as the development of a norm or ideally expected value resulting from the examples in the dataset.

For example, Caliskan et al. (2017: 184) refer to bias as "a necessary prerequisite for intelligence," therefore 'bias' stands in this context for the prior information needed in order to train neural machines. However, "prior information" or biases can be problematic when this "prior information is derived from precedents known to be harmful" (ibid.). Therefore, biases have ethically negative effects on the learning process of NMT only in certain cases. Negative gender bias (or prejudices) in NMT results from different co-occurring causes that are intrinsically related to each other. Pre-existing prejudices in society can be reproduced in the development of technical tools and also new biases can emerge. For instance, Sweeney (2013: 1) demonstrates that personal, proper names are racially tagged in Google Search.[12] A linguistic bias, on the other hand, emerges in NMT when the system establishes a grammatical gender norm based on the predominance of masculine generic forms in the training data. As a result, AI-generated translations – unless the underlying algorithms are explicitly adjusted to mitigate the bias – default to and reproduce masculine forms, which represent the most frequently used grammatical gender in many languages with a grammatical gender system.

Technical biases, by contrast, refer to

> aspects related to data creation, model design, and training and testing procedures. If present in training and testing samples, asymmetries in the semantics of language use and gender distribution are respectively learned by MT systems and rewarded in their evaluation. (Savoldi et al. 2021: 849)

Different strategies used to train MT can magnify or minimize the effect of certain biases.[13] Emergent bias, however, arises when a system is developed for a 'norm' and therefore a specific demographic group. The corpora used to train MT – as we discussed above – show

[12] Sweeney (2013) studied "racially associated names and finds statistically significant discrimination in ad delivery based on searches of 2184 racially associated personal names across two websites". Google searches involving names stereotypically related to the Black or Hispanic community were more likely to get an ad suggestive of a criminal record while searches containing names associated to white demographics got more neutral ads.

[13] We will not deepen into this topic; however, to those interested in the topic, we recommend consulting Vanmassenhove et al. (2019, 2021) or Roberts et al. (2020).

gender disparity and the generic use of the masculine grammatical gender. Consequently, when women and people who belong to other genders beyond the binary interact with MT systems, they will most likely have to adjust the translations to the grammatical gender that best represents their social gender in the target language. Google Translate, however, has stated the aim to "promote fairness and reduce gender bias" (Kuczmarski 2018: n.p.). They aim to achieve this goal by introducing a double translation that includes a feminine and masculine translation for a specific word or sentence structure. This approach, however, perpetuates the idea that gender is divided into a binary.

These endemic difficulties are well known in the industry and also approached by the UNESCO Recommendations on the Ethics of Artificial Intelligence (2021). Policy area six of the UNESCO Recommendation report approaches the topic of gender in AI explicitly. For our purpose here, two recommendations are central. First, UNESCO recommends that the "Ethical Impact Assessment should include a transversal gender perspective" (UNESCO 2021: 32). This assessment, on the other hand, should explicitly include the analysis of certain circulating and already known biases. UNESCO's recommendation therefore states that

> 89 Member States should ensure that the potential of AI systems to advance the achievement of gender equality is realized. They should ensure that these technologies do not exacerbate the already wide gender gaps existing in several fields in the analogue world and instead eliminate those gaps. These gaps include: the gender wage gap; the unequal representation in certain professions and activities; the lack of representation at top management positions, boards of directors, or research teams in the AI field; the education gap; the digital and AI access, adoption, usage and affordability gap; and the unequal distribution of unpaid work and of the caring responsibilities in our societies. (ibid.)

One of the aspects highlighted by the recommendation report concerns the "unequal representation in certain professions and activities" (ibid.). The studies of Sun et al. (2019) and Prates et al. (2020) discussed above highlight that NMT can translate sentences that reproduce gender stereotypes about, for example, professions more accurately in terms of the linguistic coding of gender. As we will discuss in this paper, both MTs being tested show a clear tendency to reproduce stereotypical gender roles, especially when it comes to the representation of certain professions.

Regarding technological biases, we can observe that NMT uses semantic representation of words also known as 'word embedding.' Caliskan et al. (2017: 184), for example, explain how negative biases can show up in word embedding:

> These [biased implicit associations] are derived by representing the textual context in which a word is found as a vector in a high-dimensional space. Roughly, for each word, its relationship to all other words around it is summed across its many occurrences in a text. We can then measure the distances (more precisely, cosine similarity scores) between these vectors. The thesis behind word embeddings is that words that are closer together in the vector space are semantically closer in some sense.

> Thus, for example, if we find that programmer is closer to man than to woman, it suggests (but is far from conclusive of) a gender stereotype.

We already stated previously that the corpora used to train NMT shows a disparity in gender representation, which leads to the creation of a supposed norm. In order to counteract against the biases resulting from the corpora, programmers need training on diversity and equity topics in order to raise awareness within the people in charge of developing algorithms used to train[14] NMT. We suggest there is a need to raise awareness regarding gender disparity in society and also the corpora used to train AI in order to assure that certain biases are not reproduced by our technologies.

When it comes to genderqueer representation, it is noteworthy that both institutions – UNESCO and the European Parliament – do not mention genderqueer perspectives. The recommendations under Policy area six of UNESCO only address gender equality from a binary perspective. Legislations overall still rely on gender binarity and still do not fully recognize the role of social genders. This positioning exhibits once again that even though there are institutions aiming to minimize gender biases, the efforts being made only encompass the binary conception of gender. NMT still does not fulfill the recommendations proposed by UNESCO (not even for the gender binary). One of the aspects cited in the quote above will be analyzed, namely the representation problems resulting from both the corpora and the algorithms used to train NMT.

Kate Crawford (2017), co-founder of the AI Now Institute and professor at New York University, has studied the consequences of biased systems. She discusses the existence of two main errors in estimation and over- or under-representing populations in sampling.

First, she speaks of a **representational harm** that detracts the representation of social groups which affects their identity, attitudes, and beliefs. This representational harm displays two bifurcations. First, underrepresentation refers to the reproduction of visibility reduction of certain groups in society. An example of this underrepresentation is the low representation of women and non-binary people in texts and their translation. In many cases, sentences which describe women and non-binary people are translated into the generic form of the masculine grammatical gender, misrepresenting these social groups as illustrated in the results of this study. This can lead to an erasure of certain social groups and can also "fail to reflect their identity and communicative repertoires" (Savoldi et al. 2021: 846). Mostly women are underrepresented while non-binary people are omitted and/or are not acknowledged. The second effect is stereotyping, which contributes to problematic generalizations regarding a certain social group. For instance, women are categorized as caretakers or chefs, but not as pilots, doctors, or scientists. As such, the output provided by MT indexes social roles and gender identities that are currently associated with less prestigious occupations. The second harm presented by Crawford (ibid.) is **allocational harm**; this arises when a system "allocates or withholds opportunities or

[14] Authors like Bolukbasi et al. (2016), Bordia and Bowman (2019), and Zhao et al. (2018) have approached gender bias and developed methods to debias the machine during training. To deepen the strategies – e.g., soft vs. hard debiasing methods – we suggest consulting the literature just cited.

resources to certain groups" (Savoldi et al. 2021: 846) and is related to aspects of "data creation, model design, and training and testing procedures." Allocation is affected by representation since the results of translations are more accurate for men than for other genders. There is a performance disparity on NMT and therefore the quality of service is more efficient for those individuals who conform to the 'norm' than those who are treated as deviances by the systems. In our analysis, we do not include the allocational harms, since we would need to draw on other kinds of data in order to analyze the consequences of representational harm.

Savoldi et al. (2021: 845) point out that

> MT is a multifaceted task, which requires resolving multiple gender-related subtasks at the same time (e.g., coreference resolution, named entity recognition). Hence, depending on the languages involved and the factors accounted for, gender bias has been conceptualized differently across studies.

These differences contribute to the complexity of NLP and the mitigating strategies applied to close the gaps provided by a biased data set.[15]

MT processing based on statistical or neural approaches does not differ from other AI applications since it learns from the dataset, and it relies on an algorithm to train the engines. Caliskan et al. (2017: 325) point out that this kind of technology is trained to try to maximize overall prediction accuracy and therefore "[i]f a specific group of individuals appears more frequently than others in the training data, the program will optimize for those individuals because this boosts overall accuracy." We can hence observe that the dataset used influences the output that NMT produces. On the other hand, computer scientists "evaluate algorithms on 'test' data sets, but usually these are random sub-samples of the original training set and so are likely to contain the same biases" (ibid.). The way scientists evaluate the data, as well as the lack of attention being paid to differences in society by the programmers therefore reproduce society's biases and norms.

4 Method

We used a small corpus of French sentences that contain examples of gender-inclusive language and their translations made by two different machine translation systems (Google Translate and DeepL) in order to analyze how NMT translates French examples of gender-inclusive language into German. The corpus consists of 48 French sentences containing at least one word in *écriture inclusive*. They were selected to include various forms of *écriture inclusive* in differing syntactic contexts. In total, the corpus contains 60 individual words (types) in *écriture inclusive*. Nouns and their modifiers (e.g., articles and adjectives) were treated as a unit. Our research focus lies on the nouns in *écriture inclusive*, although other word classes in *écriture inclusive*, such as indefinite pronouns, are also part of the corpus. The sentences were taken from texts published on various feminist French websites between September 2022 and May 2023, including the website of the feminist organization *Osez le*

15 A 'biased data set' refers to a data set that reflects the differences, discrimination, and exclusion of certain people (women, BIPOC, disabled, etc.) that do not represent the norm (white, cis, able-bodied, men) in real life.

Féminisme!, which is part of the French *Haut Conseil à l'Egalité entre les femmes et les hommes*, and the feminist online newsletter *Les Glorieuses*. Regarding the text type, most of the examples were taken from online articles as well as online newsletters. After the compilation of the corpus in the source language, the utterances in *écriture inclusive* were entered into the machine translation software (Google Translate and DeepL). Even though in our analysis we concentrated on the translation of specific nouns, full sentences were introduced to the NMT to provide enough grammatical cuing allowing the machines to assign a gender marker based on the sentence and not an isolated word.

Each of the 48 source language sentences is attributed to a number and the translations are additionally marked with either D (DeepL) or G (Google Translate) to indicate which NMT produced the target language element:

(1) 41. ***Chacun·e*** *devrait se sentir libre, dans* ***son*** *corps et dans* ***sa*** *tête.*[16]
Own translation: '**Everyone** should feel free, in **their** body and **their** mind.'
41G. ***Jeder*** *soll sich frei fühlen, in* ***seinem*** *Körper und in* ***seinem*** *Geist.*
41D. ***Jede/r*** *sollte sich frei fühlen, sowohl in* ***ihrem/seinem*** *Körper als auch in* ***ihrem/seinem*** *Kopf.*

The decision to work with this corpus was made consciously due to our aim to analyze the corpus qualitatively, whereas most of the studies conducted until today have focused on a quantitative approach (Cho et al. 2019, Prates et al. 2020, Rescigno et al. 2020). It was further a response to the studies that have been conducted in the last years regarding gender bias only addressing the topic from a binary gender perspective (cf. Braun 2000, Hacker 2018, Koehn 2005). We therefore propose to analyze the data – both the source and target language strategies – from a queer perspective. As such, our main goal in this paper is to analyze the translation strategies from a non-binary gender perspective including genderqueer options. There we propose an approach to study the harms which (i) (under)representation and (ii) stereotypes of queer genders can potentially reproduce in translations.

Regarding our research design, after collecting the examples in our corpus and translating the sentences to the target language with both NMT, we coded the data with MAXQDA. The first step consisted of coding the strategies in *écriture inclusive* in the source language according to the system displayed in Table 1.

Each strategy was given a specific value which can be observed in Table 1. We classified the different strategies of *écriture inclusive* by genderqueer and binary gender representation. In our definition, *point median*, *point*, neutral expressions and paraphrasing are considered genderqueer-friendly, while *doublet* and *trait d'union* are defined as only representing the binary system. Regarding the genderqueer representation, we must however highlight that the first two categories depicted in the table above (*point median* and *point*) mark genderqueerness explicitly, making the current debate on gender marking visible. The last two genderqueer representations (neutral and paraphrase) do not

[16] Mouronval, Amélie. 2023. "Les Petites Glo: Personne ne veut voir de femmes rondes dans la mode." https://lesglorieuses.fr/femmes-rondes-mode/ (accessed July 2023).

Strategies of *écriture inclusive* in source text	Kind of gender representation
Point median (·)	Genderqueer representation
Point (.)	Genderqueer representation
Trait d'union (-)	Binary gender representation
Doublet	Binary gender representation
Neutral (epicenes)	Gender indifferent, genderqueer-friendly representation
Paraphrase	Genderqueer, genderqueer-friendly representation

Table 1: Kinds of gender representation according to different strategies of *écriture inclusive* in the source language

display explicit gender marking which we here define as queer-friendly, although this does not mean that the gender-inclusivity by way of neutrality has been achieved intentionally.

The translations in our target language – German – were also coded using a similar scheme as in the source language examples (cf. Table 2). The difference in the coding lies in the morphologic differences between German and French, for example the neutralization of gender in plural.

The first category 'symbols' encompasses all strategies of inclusive language previously introduced in section 3.2. The symbols (/) and (-) are considered to perpetuate the visibility of the gender binary. The code 'masculine generic' is self-explanatory and was applied to all the cases in which the masculine grammatical gender seems to be used in a generic function. Regarding the category 'plural (or formal)', we used this label to code all the examples in which the plural declination in German did not allow for an explicit analysis of gender due to the morphological structure of the language. The category 'neologism/anglicism' on the other side was applied to all translations that contained loan words and/or those cases in which a new word was derived to refer to a female person in the source text. However, we would like to highlight that not all neologisms are based on loan words. Many follow the derivation rules of French/German itself. Our categorization treats neutral translations as those that encompass either a paraphrasing strategy or exhibit the use of grammatically gender-neutral nouns in German.

After localizing the different strategies used, both in source and target language, we moved on to analyze the data applying the two main categories suggested by Crawford (2017) (see section 3.2). Our main aim is to categorize how automatic translations reproduce negative linguistic biases or prejudices regarding gender. As we already stated before, our aim therefore is not to determine how the biases emerge but rather to analyze the (i) (under)representation and (ii) stereotyping of women and non-binary individuals in the translations. Regarding the category 'underrepresentation', we analyzed the corpus and coded the sentences in French to categorize the translations according to the strategies introduced in Table 2. The second category refers to stereotyping gender roles. This category was analyzed by paying attention to the meaning behind the examples in our corpus. We observed how the NMT translated nouns which are stereotypically perceived as female or male occupations. We analyzed harm according to the following categories: For the category of representation, we divided 'underrepresentation' into subcategories of (1) female and (2) queer underrepresentation. Translations which use a masculine generic form were coded as examples of underrepresentation of women and queer people. For cases in which a strategy of inclusive language was provided by the NMT but included only men and women were labeled as 'queer underrepresentation'. The code 'stereotype' was used for those cases in which the NMT uses the mas-

STRATEGIES IN TRANSLATIONS IN THE TARGET LANGUAGE	KIND OF GENDER REPRESENTATION
Symbols (*, and/or : and/or _) (/ and/or -)	 Genderqueer representation Binary gender representation
Doublet	Binary gender representation
Masculine generic	No representation of women and/or genderqueer
Feminine/Plural/Formal	Gender indifferent
Neologism and anglicism	Genderqueer representation Binary gender representation
Neutral	Genderqueer-friendly representation

Table 2: Kinds of gender representation according to different translations in the target language

culine form generically (e.g., for occupations and/or the declination of a nominal phrase even though there are other alternatives).

5 Analysis

In the previous section we provided an overview of the codes used to categorize the data in our corpus. Our corpus consists of 3959 tokens and 1374 types, the type-token ratio is =0.3471. In order to visualize the data, we present a quantitative overview of the preliminary results before analyzing the translation strategies and possible phenomena (representation and/or stereotyping) the produced output can create qualitatively. Figure 1 depicts the frequency in percentage of how often a specific strategy of *écriture inclusive* appears in our corpus.

clusive as the most frequent occurrences of strategies encompass those solutions we define as queer-friendly. On the other hand, *doublets* and examples with *trait d'union* make up 15,8 % of the strategies present in the corpus.

Regarding the strategies of inclusive language that both NMT applied to translate the input into the target language, we can observe a clear difference between DeepL and Google Translate.

As can be seen in Figure 2, DeepL's most frequent strategy (33,3 %) is to neutralize the input from the source language, that is to say that the gendered source term is translated in a way that the target term does not refer to gender at all, such as the use of neutral words as seen in DeepL's translation of the word *l'aîné-e* ('the oldest', m./f.) in sentence number 2. While Google Translate erroneously transformed the

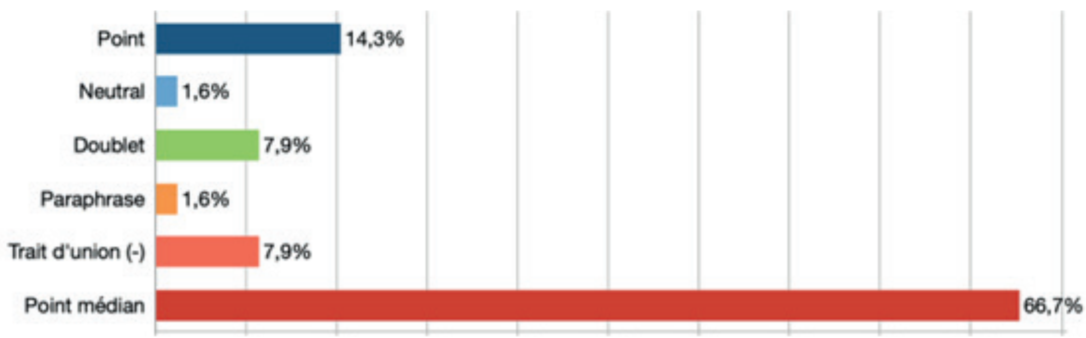

Figure 1: Frequency of different strategies of écriture inclusive

As we can see in the visualization of frequency, the strategy with the highest occurrence is the *point médian* (66,7 %). Hence, two thirds of the examples in our corpus are fully gender-inclusive. The second most frequent strategy is the *point* (14,3 %). In total, 81 % of the corpus consists of sentences that are gender-in-

French explicitly binarily gendered form into *der Älteste*, 'the oldest', which is the German masculine form, DeepL translated it as *das älteste Kind*, 'the oldest child', which strips the term of any notion of gender, seeing that the German *Kind* is semantically gender-neutral. For example:

(2) 5. *[L]es enfants d'une même fratrie doivent avoir le même nom que l'aîné-e.*[17]
Own translation: 'Siblings must have the same name as the oldest child.'
5G. *Die Kinder derselben Geschwister müssen denselben Namen haben wie der Älteste.*
5D. *Geschwister müssen denselben Namen wie* ***das älteste Kind*** *haben.*

Other examples frequently used in daily life include the usage of epicenes like *persons, people,* or *humans* instead of *women* and/or *men*, sometimes seen in suffixes as well such as *Fachmensch* 'professional (human)' instead of *Fachmann* 'professional (man)' or *Fachfrau* 'professional (woman)'[18] in either of the languages. The second most frequent strategy (28,3 %) is the generic use of the masculine grammatical gender. It is interesting to observe that oftentimes, sentences translated using the masculine generic (MG) were sentences which contained occupations in the source language that are considered stereotypically male. The 23 job titles in the source texts used either gender-inclusive variations, both the female and the male or solely the latter form, but no kind of ambiguous designation that might have been interpreted as a reason for NMT to struggle with an adequately gendered translation into the target language. Google Translate employed the MG for all of them except in one case, *Europaabgeordnete* 'members of the European Parliament', which is a form of *Differentialgenus* in plural form and thus does not indicate gender. However, it may very well be argued that this output resulted from the fact that this is the most commonly used form in German and there is no masculine generic alternative. This example shows that lexical choices in NMT often mirror the frequency and conventionality of certain forms in the

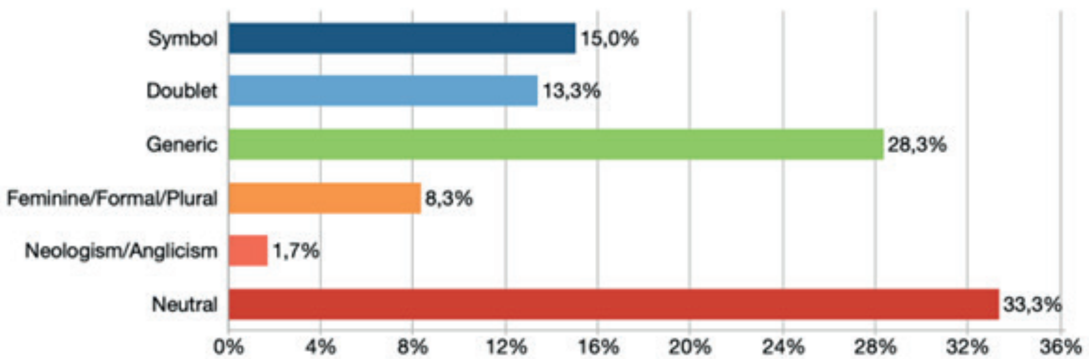

Figure 2: Frequency of occurrence of codes for DeepL

[17] Chloé. "Porte mon nom." https://feministoclic.olf.site/porte-mon-nom/ (accessed July 2023).

[18] Contrary to some opposing voices, many of these terms in both French and German have existed for decades and even centuries and thus are not simply an invention of modern queer-feminist efforts, cf. *Fachmensch*, in: Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm, Erstbearbeitung (1854–1960), digitalisierte Version im Digitalen Wörterbuch der deutschen Sprache.

training data even though (NMT) systems operate based on vector embeddings rather than relying solely on statistical frequency. Therefore, the distribution patterns within the training corpus continue to exert a significant influence on the outputs.

When it comes to *doublets*, Google Translate rendered the French phrase *les sénateurs et les sénatrices* ('senator', m./f.) into the German *die Senatoren* ('the senators', m., pl.), thereby collapsing the inclusive formulation into the masculine generic:

(3) 29. *Nous appelons toutes celles et tous ceux qui sont attaché·es*[19] *au droit à l'IVG à suivre la séance et à réagir aux débats au Sénat pour encourager* ***les sénateurs et les sénatrices*** *à voter en faveur de la loi.*[20]
Own translation: We call on all those who are attached to the right to abortion to follow the session and react to the debates in the Senate to encourage **senators** to vote in favor of the law.

29G. *Wir rufen alle, die sich für das Recht auf Abtreibung einsetzen, dazu auf, die Sitzung zu verfolgen und auf die Debatten im Senat zu reagieren, um* ***die Senatoren*** *zu ermutigen, für das Gesetz zu stimmen.*

On the other hand, sentences that contained female gendered occupation titles in the source language were translated using masculine generic counterparts in German: *une magistrate professionnelle,* ('a professional magistrate', f.), was translated as *ein professioneller Richter* (m.) instead of *eine professionelle Richterin* (f.) and *cheffes étoilées*, ('star chefs' f., pl.) was translated as *Sterneköche* (m., pl.) instead of *Sterneköchinnen* (f., pl.). Contrastingly, DeepL only used the MG 14 times. However, similar tendencies to Google Translates' strategies can be found in the sentences delivered by DeepL. The following example exhibits that also DeepL subsumed the gender-inclusive use of *point médian* into the MG counterpart in German.

(4) 8. *Les médicaments abortifs retirent la décision de mettre fin à une grossesse des mains* ***des docteur·e·s*** *et* ***des élu·e·s*** *et la remettent entièrement entre les mains des personnes qui ont besoin d'avorter.*[21]
Own translation: Abortion drugs take the decision to terminate a pregnancy out of the hands of **doctors** and **elected officials**, and put it entirely in the hands of those who need an abortion.

8D. *Abtreibungsmedikamente nehmen die Entscheidung, eine Schwangerschaft zu beenden, aus den Händen von* ***Ärzten*** *und* ***Abgeordneten*** *und legen sie vollständig in die Hände der Menschen, die das Bedürfnis haben abzutreiben.*

This example shows that DeepL also doesn't provide a gender-inclusive translation when dealing with occupation titles in this case.

When it comes to the examples *une magistrate professionnelle* ('professional judge', f.) and *cheffes étoilées*, ('star chef', f.) previously dis-

19 In this case the use of point médian can also be interpreted as a binary translation. We need to highlight that in French there is no uniform use of this strategy. Therefore, point médian can both be used to represent gender in binarity or beyond binarity.

20 Osez le Féminisme ! 2022. "Droit à l'IVG dans la Constitution: la bataille commence au Sénat!". https://osezlefeminisme.fr/droit-a-livg-dans-la-constitution-la-bataille-commence-au-senat/ (accessed July 5th 2023).

21 Clement, Megan. 2023. "IMPACT". https://lesglorieuses.fr/qui-a-peur-de-la-pilule-abortive/ (accessed May 9th 2023).

cussed for Google Translate, we can state that DeepL correctly translated both of the female job titles into *Berufsrichterin* ('professional judge', f.) and *Sterneköchinnen* ('star chefs', f.).

We can clearly observe that most of the job titles refer to prestigious occupations such as *Präsidenten* ('presidents', m., pl.), *Dekane* ('deans', m., pl.), or *Minister* ('ministers', m., pl.) which are considered as stereotypically male. However, jobs that could be considered 'modern' and are expressed with neologisms borrowed from English into French and/or German (e.g., *instagrammeur/instagrammeuse* 'instagrammer', m./f., and *Medienmanager* 'media manager', m.) are considered neither stereotypically male nor female. While the source terms are at times gender-inclusive, these occupations are translated using only the masculine grammatical gender.

Additionally, there are nouns in plural with *Differentialgenus* (e.g., *Minderjährige* 'underaged' or *Jugendliche* 'young people'), which do not index gender themselves but only through accompanying function words in singular form. Here are a few examples:

(5) 34. *Les Petites Glo, c'est la première newsletter destinée aux* ***adolescent·e·s***.[22]
Own translation: Les Petites Glo is the first newsletter aimed at **teenagers.**
34D. *Les Petites Glo ist der erste Newsletter, der sich an* ***Jugendliche*** *richtet.*

Therefore, we cannot generally speak of a translation that explicitly applies gender-inclusive strategies neutralizing gender in the translation, but they are rather the result of the coincidence that those lexemes do not index gender in German. Yet, there are examples where DeepL used neutral forms when a translation by a masculine generic form was also applicable and even more common at times: *rangées d'étudiant·e·s* ('rows of students', m./f.) became *Reihen der Studierenden*, whereas Google Translate used the more traditional masculine form *Studentenreihen*.

(6) 46. *Reste encore l'image des* ***rangées d'étudiant·e·s*** *devant la distribution d'aide alimentaire, des apéros-visios et des examens chamboulés…*[23]
Own translation: There's still the image of **rows of students** in front of the food aid distribution, of aperitif-visios and disrupted exams…
46D. *Es bleibt noch das Bild von den* ***Reihen der Studierenden*** *vor der Verteilung der Lebensmittelhilfe, von den Aperitifen und den Prüfungen, die durcheinander gebracht wurden...*

The results regarding occupations and the strategies used by DeepL coincide with the results of other studies that analyzed gender bias by measuring different translation equivalents of job titles (cf., e.g., Farkas & Németh 2022).

Google Translate, on the other hand, shows a clear tendency to provide translations in generic use of the masculine grammatical gender; 60,5 % of the results contain this strategy. The second most frequent category are neutral word forms with 31,6 %, as figure 3 depicts.

The tendency that we observed for DeepL, namely a preference for the generic use of the

[22] "Les Petites Glo". Accessed May 9th 2023. https://lesglorieuses.fr/les-newsletters/les-petites-glo/
[23] Mouronval, Amélie. 2023. "Les Petites Glo: Ceci n'est pas un problème isolé." https://lesglorieuses.fr/ceci-nest-pas-un-probleme-isole/?cn-reloaded=1(accessed July 3rd 2023).

masculine grammatical gender for stereotypically male occupations, is maintained in the translations provided by Google Translate. The main difference is that Google Translate utilizes masculine forms generically in a greater number of examples, such as *Praktikanten* 'intern' (m., pl.), while DeepL in certain cases applies other strategies for such input as discussed previously in this section.

Google Translate translates most of the occupations present in our corpus as masculine (e.g., *Parlamentarier* ['member of the parliament', m.], *Praktikanten* ['intern', m., pl.], *Minister* ['minister', m.], *Ärzte* ['doctors', m., pl.], *Richter* ['judge', m.], *Elektriker* ['electrician', m.], *Senatoren* ['senators', m., pl.], etc.). This is done even in the case of *Paläontologe* or *Minister,* which in French are words of the *Differentialgenus* that do not indicate grammatical or social gender without a determinant – this was not present in the source texts.

(7) 28. *Plusieurs* ***ministres*** *et près de 440* ***parlementaires*** *au Sénat et à l'Assemblée nationale, se sont déjà* ***engagé·es*** *dans une démarche d'inscription du droit à l'IVG dans la Constitution.*[24]

Own translation: Several **ministers** and nearly 440 **members** of the French Senate and National Assembly have already committed to enshrining the right to abortion in the Constitution.

28G. *Mehrere* ***Minister*** *und fast 440* ***Parlamentarier*** *im Senat und in der Nationalversam mlung haben sich bereits dafür eingesetzt, das Recht auf Schwangerschaftsabbruch in die Verfassung aufzunehmen.*

Google Translate further tends to employ masculine forms for other nouns which do not necessarily imply gender roles: for example, *leurs ami·e·s* is translated as *Freunde* 'friends' (m.).

When it comes to apparently neutral translations, on the other hand, we see a similar pattern in the translations provided by DeepL.

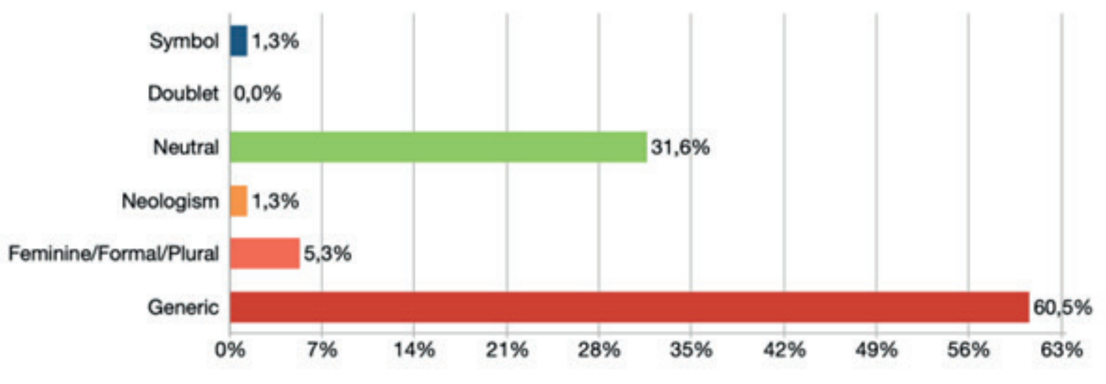

Figure 3: Frequency of occurrence of codes for Google Translate

24 Osez le Féminisme ! 2020. "Droit à l'IVG dans la Constitution: la bataille commence au Sénat !". https://osezlefeminisme.fr/droit-a-livg-dans-la-constitution-la-bataille-commence-au-senat/ (accessed July 5th, 2023).

Figure 4 quantifies the translations we coded to interpret the harms resulting from the translation procedures used by both NMT. The percentages were obtained by applying the categories presented in Figure 2. Translations could be coded with more than one variable. E.g., 'Harms_Female underrepresentation' could also at the same time be labeled as 'Harms_queer underrepresentation'.

As can be observed in Figure 4, 'queer underrepresentation' occurs most frequently in our corpus (31,0 %), followed by female underrepresentation (24,3 %). In 22,2 % of the cases, the translations provided by both NMT represent gender from the source language in concert with the target language. Concerning stereotypes, 22,6 % of the translations represent a reproduction of stereotypes according to gender marking either on a syntactic (e.g., a correction in declination may be needed) or lexeme level (e.g., the nouns regarding occupation might need adjusting). This was especially shown with the examples that contained profession titles (lexeme) and the syntactic structure surrounding the profession titles.

In order to better understand the negative effects, or 'harms', as Crawford (2017) defines them, we analyze the data qualitatively. In the translations made by Google Translate and DeepL, the use of masculine word forms predominates, thus leading to the underrepresentation of both women and genderqueer people. This becomes most apparent in terms of occupation. While DeepL frequently translates occupation-related terms using the masculine generic, it occasionally employs gender-inclusive strategies. Google Translate, on the other hand, uses masculine generics rather broadly. We should note that there is an ongoing debate, concerning French, regarding the usage of the feminine grammatical gender in occupational terminology (cf. Lessinger 2020, Viennot 2020). The use of feminine job titles has been increasingly accepted over the last 30 years; still, the *Académie Française* – as the official authority on changes in the linguistic policies regarding the French language – agreed only recently, in 2019, to the use of certain feminine terms for occupations, and these will be gradually represented in corpora used to train NMT (cf. Haut Conseil à l'Egalité entre les femmes et les hommes 2023: 4). However, the hesitance and lack of norms regarding gender-inclusive language in society fits with our findings, where terms denoting oc-

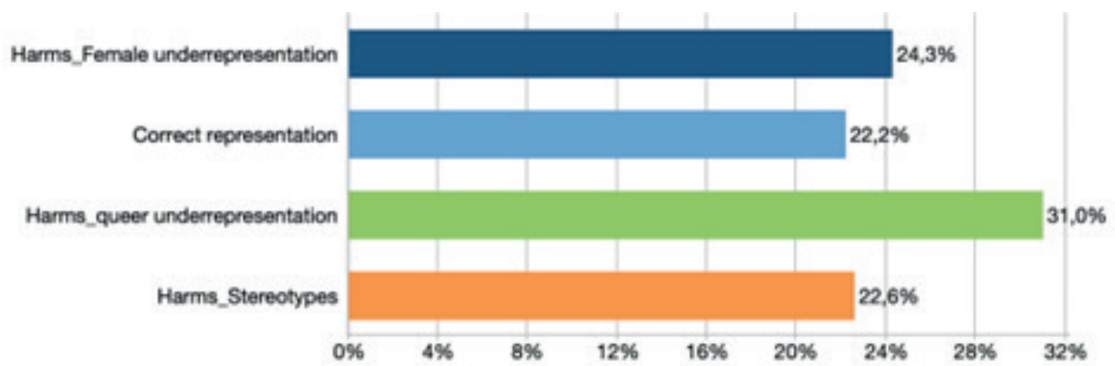

Figure 4: Representation harms in corpus

cupations were in most cases translated using masculine forms.

The translations show a strong bias towards treating the masculine grammatical gender as standard. This applies not only to occupational terms; most gender marked terms are translated into masculine forms with only a few exceptions. In Google Translate, these exceptions can be attributed to the linguistic characteristics of the language pair French/German.

Similarly, when a sentence containing a term referring to one or more person(s) in the French source text is translated into German by restructuring the syntax and introducing a different part of speech as the predicate, a gender-neutral expression may be created as part of the translation process without deliberately taking a gender-inclusive approach. For example, *Mais elles sont moins bien informé.es* (see (8)).

As we can see in the following example, the French sentence starts with the feminine third person plural pronoun (*ils* marks the 3. p. pl. for men). In German, however, the plural pronoun is ambiguous, because the plural pronoun *sie* indexes all genders. As such, the translation will be gender-indifferent.

(8) 12. *Mais **elles** sont moins bien **informé.es**, plus mal **suivies** que les hommes et s'occupent davantage de la santé de leurs proches que de leur santé.*[25]
Own translation: But they are less well-informed, have poorer follow-up than men, and are more concerned with the health of their loved ones than with their own health.
12G. *Aber **sie** sind weniger gut informiert, werden weniger gut betreut als Männer und kümmern sich mehr um die Gesundheit ihrer Lieben als um deren Gesundheit.*
12D. ***Sie** sind jedoch weniger gut informiert, werden schlechter betreut als Männer und kümmern sich mehr um die Gesundheit ihrer Angehörigen als um ihre eigene Gesundheit.*

In the example above, the original appears as an attempt at gender-inclusive phrasing using the feminine plural pronoun *elles* in a generic manner while also making use of the gender-inclusive strategy utilizing the point at the morpheme boundary to indicate the notion of including more than just the female gender. Alternatively, one could have used the gender-neutral neopronoun *iels.*

Another phenomenon specific to this language pair may arise due to the disparity in the gender marking of pronouns. While possessive pronouns are inflected according to the gender of both the possessor and the possessed entity in the third person in German, they are only inflected according to the gender of the possessed in French. Consequently, when there is only little context, the French source text may lack information needed to produce an adequate translation into German. Concerning this, there was a disparity between the translations provided by DeepL and Google Translate, as in the following example:

25 Germian, Isabelle. 2023. https://www.lesnouvellesnews.fr/le-coeur-des-femmes-moins-bien-soigne-que-celui-des-hommes/ (accessed July 2023).

(9) 11. ***Sa** conviction que "nous ne sommes pas mal adapté·e·s au monde dans lequel nous vivons".*[26]
Own translation: **Her** conviction that "we are not badly adapted to the world we live in".
11G. ***Seine** Überzeugung: „Wir sind für die Welt, in der wir leben, nicht schlecht geeignet."*
11D. ***Ihre** Überzeugung, dass „wir nicht schlecht an die Welt, in der wir leben, angepasst sind."*

Google Translate translates *sa* ('her' in this case) as *seine* ('his'), whereas DeepL produced *ihre* ('her'). In this case, the lack of information leads Google Translate to revert to a masculine generic target term for translation. However, it needs to be pointed out that DeepL, lacking the same information, produced a gendered translation. This may be attributed to the fact that further in the same sentence, a gender-inclusive participle indicates that a masculine term may not be adequate.

There is only one exception to the predominance of the generic use of the masculine grammatical gender in translations produced by Google Translate, namely the word *féministe* ('feminist', m./f.). The source term alone, having a *Differentialgenus*, does not indicate its grammatical gender, but is marked as gender-inclusive by the usage of the *point médian* in the preceding article and attributive adjective in French.

(10) 7. *Dites-nous qui vous a fait devenir **un·e meilleur·e féministe.***[27]
Own translation: Tell us who made you a better **feminist.**
7G. *Sagen Sie uns, wer Sie zu **einer besseren Feministin** gemacht hat.*
7D. *Sagen Sie uns, wer Sie zu **einer besseren Feministin/einem besseren Feministen** gemacht hat.*

DeepL therefore provided for this example a gender-inclusive translation that contains a binary mention of 'feminist' (m./f.) using the slash.

Google Translate on the other hand provided a translation solely in the feminine (7G): (*eine bessere Feministin*, 'a better feminist', f.). As seen with the predominating use of the masculine, the source text's gender-inclusive marking is omitted in the translation output here as well. In contrast, DeepL employed a gender-inclusive strategy for the target text including both women and men (*einer besseren Feministin/einem besseren Feministen* ['a better feminist', f./m.]), albeit not genderqueer people. In our view, this results from the frequency that *Feministin* appears within the corpora from which the translation machines draw.

The reproduction of such stereotypes in the translations, particularly those produced by Google Translate, offers an insight into gender bias which is present in the corpus data on which the MT software is based. Whereas Google Translate fails entirely to bypass this bias, to employ adequate target terms and improve on its translation quality, DeepL manages to apply gender-inclusive strategies at least in some cases. Nevertheless, the predominance of the generic use of the masculine reflects social practices where MG is used as the norm.

[26] Clement, Megan. 2023. "Impact: Qui de la pilule abortive?". https://lesglorieuses.fr/qui-a-peur-de-la-pilule-abortive (accessed July 2023)
[27] Clement, Megan. 2023. "Impact: Qui de la pilule abortive?". https://lesglorieuses.fr/qui-a-peur-de-la-pilule-abortive (accessed July 2023).

Therefore, it is of great importance not only to notice such disparities in translation but also to conduct further research into their causes, and to implement strategies to reduce them.

Interestingly, DeepL seems to apply the respective chosen strategy concerning the gender marking of terms referring to people consistently throughout a sentence, even in cases where the original terms differ in the morphological marking of gender-inclusive language. In an enumeration of three different terms for occupations, only the first term *électricien·ne* clearly employs a gender-inclusive strategy, while *paléontologue* as well as *prof' de techno*, having the *Differentialgenus*, could be read as female or male for the lack of a gender defining article. Nonetheless, DeepL applied the same strategy of gender-inclusive language for all three terms in German.

(11) 43. *On ne naît pas pour devenir **électricien·ne**, **paléontologue** ou **prof' de techno** au collège.*[28]
Own translation: You are not born to become an **electrician**, a **paleontologist** or a **technology teacher** at high school.
43D. *Man wird nicht geboren, um **Elektriker/in**, **Paläontologe/in** oder **Techniklehrer/in** am Gymnasium zu werden.*

Thus, it seems that the existence of the one term demonstrating the intention of using gender-inclusive language was sufficient for DeepL to translate using a gender-inclusive strategy, to produce a translation output which represents women and men, although it excludes genderqueer people.

6 Implications and conclusion

This study set out to analyze the treatment of grammatical gender in neural machine translation (NMT), considering the linguistic particularities present in German and French, as well as translational phenomena.

While grammatical gender itself does not consistently refer to biological or social gender, there is a strong correlation between the grammatical gender of a personal noun and the social gender it refers to. Thus, it is important to be conscious of the possible reciprocal influence that language use might have on the perception of reality, and vice versa. Consequently, it is not only the results of machine translation that perpetuate gender stereotypes; it is stereotypes in the first place that presuppose the creation of inadequate translation output through gender bias present in machine translation. Gender bias is developed when the data that NMT is built upon is flawed, in the way that gender norms and stereotypes exist in patriarchally shaped societies and are thus incorporated into machine learning without being considered and treated critically. Hence, a cycle develops in which gender bias is transported through the mediums of both human and machine.

The results of this study show clearly how gender bias is (re)produced in NMT translations. There is a strong tendency for translations to employ masculine forms in a generic, supposedly gender indistinct manner. This greatly reduces the visibility of women and, even more so, genderqueer people. Additionally, stereotypical gender roles and norms assigned to women and men reflect the notion

[28] Mouronval, Amélie. 2023. "Les Petits Glo: Et toi tu veux faire quoi?". https://lesglorieuses.fr/et-toi-tu-veux-faire-quoi/ (accessed July 5th 2023).

of a binary system with preferences for gender marking; this results in inadequate translations, particularly for prestigious professional and academic titles. Genderqueer people, however, are not subject to any kind of biased notions based on gender performance; rather, they are excluded completely in translations produced by the NMT software.

Considering the technical aspects of NMT, it is unclear how the various types of gendered language utilizing different characters affect the translation quality. However, it has become evident that employment of the point as a symbol for gender diversity between French morpheme boundaries may quite well be a difficulty to the NMT in struggling to differentiate from its common use as a punctuation mark. This is corroborated by the fact that none of the French words employing one or two points as a means of gender-inclusive language were translated as such to German: The masculine generic was mostly applied next to some words or word forms that are gender-neutral in German. Thus, it can be shown that technical aspects have an influence on the treatment of gendered language as well. The examples introduced above exhibit that although contemporary NMT systems like Google Translate and (partially) DeepL are built on vector-space representations rather than explicit statistical frequency tables, the representations themselves are trained on language data where certain forms – such as the masculine generic – occur significantly more often. In this way, the training process internalizes and reproduces dominant linguistic norms. The vector embeddings reflect the distributional regularities of the corpora, which means that more frequently occurring lexical items are statistically overrepresented. This therefore illustrates that vector-based NMT does not eliminate the influence of frequency. Consequently, inclusive forms that are less prevalent in the training corpora remain marginalized in machine translation outputs, reinforcing linguistic norms that privilege default MG.

This exhibits how important it is to develop legislation that regulates how AI is trained to deal with various disparities and inequalities. As was pointed out in this paper, the recommendations of UNESCO and the European Parliament need to become a set of rules that regulate the development of AI-tools such as NMT.

In order to improve upon the translation quality and reduce negative gender bias in NMT, it is necessary to start by becoming aware of the ways in which bias presents itself and where it stems from. This has been attempted at a limited capacity in this study. It has become evident that the source data used for NMT machine learning will need to be further evaluated regarding the presence of harmful stereotypes that will then need to be extracted from the data. Similarly, the process of data acquisition needs to be revised as well, seeing that diverse, bias-informed texts already exist and can be drawn upon. Furthermore, the algorithms employed in NMT need to be able to incorporate different strategies of gendered language, such as word formation according to genderqueer-inclusive standards.

To be able to further assess NMT's performance when translating languages with grammatical gender systems, translations from German to French will need to be considered as well. A comparison between the two directions of translation may offer valuable insight when investigating how the translation of gendered language may be improved.

References

Babych, Bogdab & Anthony Hartley. 2003. Improving machine translation quality with automatic named entity recognition. In Bente Maegaard (ed.), *Proceedings of the 7th International EAMT Workshop on MT and Other Language Technology Tools, Improving MT Through Other Language Technology Tools: Resources and Tools for Building MT,* pp. 1–8. Stroudsburg: Association for Computational Linguistics.

Becker, Kathrin, Miriam Brunnengräber, Christoph Gabriel, Silvia Hansen-Schirra, Maria Lau, Elena Lazic, Miriam Lind, Saskia Mahal, Klara Mladinic & Damaris Nübling. 2022. *Handreichung Diskriminierungsarme Sprache*. Mainz: Johannes Gutenberg-Universität. https://gleichstellung.uni-mainz.de/files/2022/03/Handreichung-Diskriminierungsarme-Sprache_JGU_DIGITAL.pdf (accessed June 2023).

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama & Adam Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in neural information processing systems* 29: 1–9.

Bordia, Shikha & Samuel Bowman. 2019. Identifying and reducing gender bias in word-level language models. *arXiv preprint arXiv*: 1904.03035. https://arxiv.org/abs/1904.03035 (accessed January 2024).

Braun, Friederike. 2000. *Geschlecht im Türkischen: Untersuchungen zum sprachlichen Umgang mit einer sozialen Kategorie.* Turcologica Series. Wiesbaden: Otto Harrassowitz Verlag.

Busch, Albert & Stenschke, Oliver. 2018. *Germanistische Linguistik. Eine Einführung.* Tübingen: Narr Francke Attempto.

Caliskan, Aylin, Joana J. Bryson & Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356.6334: 183–186. https://doi.org/10.1126/science.aal4230 (accessed August 2023).

Cho, Won Ik, Ji Won Kim, Seok Min Kim & Nam Soo Kim. 2019. On measuring gender bias in translation of gender-neutral pronouns. In Marta R. Costa-Jussà, Christian Hardmeier, Will Radford & Kellie Webster (eds.), *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pp. 173–181. Florence: Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-3824 (accessed August 2023).

Choi, Heelyoul, Kyunghyun Cho & Yoshoua Bengio. 2017. Context-dependent word representation for neural machine translation. *Computer Speech & Language* 45: 149–160. https://doi.org/10.1016/j.csl.2017.01.007 (accessed April 2025).

Crawford, Kate. 2017. The trouble with bias, *31st Conference on Neural Information Processing Systems (NIPS), keynote speech.* Long Beach, USA.

Di Sabato, Bruna & Antonio Perri. 2020. Grammatical gender and translation: A cross-linguistic overview. In Luise Flotow & Hala Kamal (eds.), *The Routledge Handbook of Translation, Feminism and Gender*, pp. 363–373. London, New York: Taylor & Francis Group.

Diewald, Gabriele & Damaris Nübling (eds.). 2022. *Genus – Sexus – Gender.* Berlin, Boston: De Gruyter.

Dupuy, Alexandra. 2020. L'écriture inclusive: la définir pour mieux la comprendre. *Correspondance.* https://correspo.ccdmd.qc.ca/wp-content/uploads/2020/11/correspondance-lecriture-inclusive-la-definir-pour-mieux-la-comprendre.pdf (accessed May 2023).

Ewels, Andrea-Eva (ed.). 2020. Themenheft "Geschlechtergerechte Sprache". *Der Sprachdienst* 64. Wiesbaden: Gesellschaft für deutsche Sprache e. V.

Farkas, Anna & Renáta Németh. 2022. How to measure gender bias in machine translation: Real-world oriented machine translators, multiple reference points. *Social Sciences & Humanieties Open* 5.1: 100239. https://doi.org/10.1016/j.ssaho.2021.100239 (accessed August 2023).

Hacker, Philipp. 2018. Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common market law review* 55.4: 1143–1185.

Hardmeier, Christian. 2012. Discourse in statistical machine translation. *Discours* 11 online, n.p. http://journals.openedition.org/discours/8726 (accessed August 2023).

Haut Conseil à l'Égalité entre les femmes et les hommes. 2023. *Pour une Communication Publique Sans Stéréotypes de Sexe. Guide Pratique.* Paris: Haut Conseil à l'Egalité entre les femmes et les hommes.

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

Koehn, Philipp. 2005. Europarl: A parallel corpus for statistical machine translation. *Proceedings of the tenth Machine Translation Summit*, pp. 79–86, Phuket: AAMT.

Kotthoff, Helga & Damaris Nübling. 2018. *Genderlinguistik.* Tübingen: Narr Francke Attempto Verlag.

Kuczmarski, James. 2018. Reducing gender bias in google translate, n.p. https://www.blog.google/products/translate/reducing-gender-bias-google-translate/ (accessed August 2023).

Lessinger, Enora. 2020. Le président est une femme: The challenges of translating gender in UN texts. In Luise Flotow & Hala Kamal (eds.), *The Routledge Handbook of Translation, Feminism and Gender*, pp. 374–389. London, New York: Taylor & Francis Group.

Ministère de l'Éducation nationale, de la Jeunesse et des Sports. 2021. Règles de féminisation dans les actes administratifs du ministère de l'Éducation nationale, de la Jeunesse et des Sports et les pratiques d’enseignement. *Bulletin Officiel de l’Éducation Nationale, de la Jeunesse et des Sports* 18. https://www.education.gouv.fr/sites/default/files/imported_files/documents/BO_18_MENJS_1405801.pdf (accessed June 2024).

Ministère de la Justice. 2017. Circulaire du 21 novembre 2017 relative aux règles de féminisation et de rédaction des textes publiés au Journal officiel de la République française (NOR : PRMX1732742C). *Journal Officiel de la République Française,*0272, texte 4. https://www.legifrance.gouv.fr/eli/circulaire/2017/11/21/PRMX1732742C/jo/texte (accessed June 2024).

Monti, Johanna. 2020. Gender issues in machine translation: An unsolved problem?. In Luise von Flotow & Hala Kamal (eds.), *The Routledge Handbook of Translation, Feminism and Gender*, pp. 457–468. London, New York: Routledge. https://doi.org/10.4324/9781315158938-39 (accessed August 2023).

Nübling, Damaris. 2020. Genus und Geschlecht: Zum Zusammenhang von grammatischer, biologischer und sozialer Kategorisierung. *Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse, Akademie der Wissenschaften und der Literatur Jahrgang 2020.* 1. Mainz, Stuttgart: Akademie der Wissenschaften und der Literatur, Franz Steiner Verlag.

O’Brien, Sharon & Silvia Rodriguez Vazquez. 2020. Translation and technology. In Sara Laviosa (ed.), *The Routledge Handbook of Translation and Education,* pp. 264–277. London, New York: Routledge.

Prates, Marcelo, Pedro H. Avelar & Luis C. Lamb. 2020. Assessing gender bias in machine translation: A case study with Google Translate. *Neural Computing & Applications* 32.10: 6363–6381. https://rdcu.be/dkyi1 (accessed August 2023).

Pusch, Luise. 1979. Der Mensch ist ein Gewohnheitstier, doch weiter kommt man ohne ihn. Eine Antwort auf Kalverkämpers Kritik an Trömel-Plötz‘ Artikel über Linguistik und Frauensprache. *Linguistische Berichte Braunschweig* 63: 84–102.

Rescigno, Argentina A., Eva Vanmassenhove, Johanna Monti & Andy Way. 2020. A case study of natural gender phenomena in translation: A comparison of Google Translate, Bing Microsoft Translator and DeepL for English to Italian, French and Spanish. In Sharon O’Brien & Michel Simard (eds.), *Proceeedings of the Workshop on the impact of machine translation (iMpacT 2020), Association for Machine Translation in the Americas,* pp. 62–90. Association for Machine Translation in the Americas. https://aclanthology.org/attachments/2020.amta-impact.4.Presentation.pdf (accessed June 2021).

Roberts, Nicholas, Davis Liang, Graham Neubig & Zachary Lipton. 2020. Decoding and diversity in machine translation. *Proceedings of the Resistance AI*

Workshop at 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Vancouver, CA.

Rosar, Anne. 2022: Mann und Frau, Damen und Herren, Mütter und Väter– Zur (Ir-) Reversibilität der Geschlechterordnung in Binomialen. In Gabriele Diewald & Damaris Nübling (eds.), *Genus – Sexus – Gender*, pp. 267–292. Berlin, Boston: De Gruyter.

Sarrasin, Oriane, Ute Gabriel & Pascal Gygax. 2012. Sexism and attitudes toward gender-neutral language. The case of English, French and German. *Swiss Journal of Psychology* 71.3: 113–124.

Saunders, Danielle & Bill Byrne. 2020. Reducing gender bias in neural machine translation as a domain adaptation problem. In Dan Jurafsky, Joyce Chai, Natalie Schluter & Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7724–7736. Stroudsburg, PA: Association for Computational Linguistics.

Savoldi, Beatrice, Marco Gaido, Luisa Bentivogli, Matteo Negri & Marco Turchi. 2021. Gender bias in machine translation. *Transactions of the Association for Computational Linguistics* 9: 845–874. https://doi.org/10.1162/tacl_a_00401 (accessed August 2023).

Schwarze, Brigitte. 2009. Genus im Sprachvergleich. *Tübinger Beiträge zur Linguistik (TBL)* 511. Tübingen: Narr Verlag.

Sczesny, Sabine. 2019. Wieso muss das sein?! Zum Nutzen geschlechtergerechter Sprache. *Blog Interdisziplinäre Geschlechterforschung.* https://www.gender-blog.de/beitrag/nutzen-geschlechtergerechter-sprache (accessed April 2025).

Sun, Tony, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai El Sherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang & William Yang Wang. 2019. Mitigating gender bias in natural language processing: Literature Review. In Anna Korhonen, David Traum & Lluís Màrquez (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,* pp. 1630–1640. Florence: Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1159 (accessed August 2023).

Sweeney, Latanya. 2013. Discrimination in online ad delivery. *Queue* 11.3: 10 –29.

Tomalin, Marcus, Bill Byrne, Shauna Concannon, Danielle Saunders & Stefanie Ullmann. 2021. The practical ethics of bias reduction in machine translation: why domain adaptation is better than data debiasing. *Ethics and Information Technology* 23.3: 419–433. https://doi.org/10.1007/s10676-021-09583-1 (accessed August 2023).

Tversky, Amos & Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology* 5.2: 207–232. https://doi.org/10.1016/0010-0285(73)90033-9 (accessed August 2023).

UNESCO. 2021. Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000381137 (accessed December 2023).

Vanmassenhove, Eva, Dimitar Shterionov & Andy Way. 2019. Lost in translation: Loss and decay of linguistic richness in machine translation. *Proceedings of Machine Translation Summit XVII Volume 1: Research Track*, pp. 222–232. Dublin: European Association for Machine Translation.

Vanmassenhove, Eva, Dimitar Shterionov & Matthew Gwilliamm. 2021. Machine translation: Effects of algorithmic bias on linguistic complexity in machine translation. In Paola Merlo, Jorg Tiedemann & Reut Tsarfaty (eds.), *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 2203–2213. Stroudsburg, PA: Association for Computational Linguistics.

Viennot, Éliane. 2020. Langage égalitaire : vers une rationalisation des procédés et des approches. Cahiers d'histoire. *Revue d'histoire critique* 146: 149–160. https://doi.org/10.4000/chrhc.14838 (accessed August 2023).

Voita, Elena, Pavel Serdyukov, Rico Sennrich & Ivan Titov. 2018. Context-aware neural machine translation learns anaphora resolution. In Iryna Gurevych & Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1264–1274. Melbourne: Association for Computational Linguistics. https://aclanthology.org/P18-1117/ (accessed August 2025).

Werthmann, Antonina & Andreas Witt. 2014. Maschinelle Übersetzung – Gegenwart und Perspektiven. In Herhard Stickel (ed.), *Translation and Interpretation in Europe. Contributions to the Annual Conference 2013 of EFNIL in Vilnius*, pp. 79–103. Frankfurt et al.: Lang.

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Le Quoc, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukas Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Gregory Corrado, Macduff Hughes & Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *ArXiv abs*/1609.08144: n. p.

Zhang, Xiaojun & Qun Liu. 2023. Machine translation: General. In Chan Sin-wai (ed.), *Routledge Encyclopedia of Translation Technology*, pp. 112–127. London, New York: Routledge.

Zhao, Jieyu, Yichao Zhou, Zeyu Li, Wei Wang & Kai-Wei Chang 2018. Learning gender-neutral word embeddings. arXiv preprint arXiv:1809.01496: n.p. https://arxiv.org/abs/1809.01496 (accessed January 2024).