Social Bias in Machine Translation – with a Focus on Gender Bias

Table of Contents

1 Introduction

Neural Machine Translation (NMT) has significantly facilitated our daily life and changed the translator's job profile. NMT systems are rich in information because they are trained on large amounts of data available online – data which has mostly been created by society. This data reflects societal ways of thinking and patterns of behaviour, affecting the choices NMT systems make when translating a text from one language into another. At times, this can result in skewed or wrong translations, such as the current German Chancellor Olaf Scholz often being translated as a feminine chancellor (DE: *Kanzlerin*) since, before Olaf Scholz took office, Angela Merkel had been the German Chancellor for 16 years, and from those 16 years. NMT systems trained on data from these 16 years naturally learned to translate the English *chancellor* as *Kanzlerin* in German (UEPO.de 2022). These kinds of translation phenomena are the focus of this paper, which examines typical biases, such as gender, racial or social bias, in machine translation (specifically in NMT). The aim is to find instances of such biases in NMT and to illustrate these instances in a video so that students get a better understanding of typical occurrences of bias in machine translation and know how to look out for and handles these biases.

2 Machine Translation

Key to understanding how NMT systems work is that they are *data-driven*, meaning that they are trained on large translation corpora (Vanmassenhove et al. 2021:2204; Rama/Vanmassenhove 2021:68). To process this data, NMT systems create numerical representations of the source text that include all its grammatical and semantic information (Peters et al. 2018:1). An exact translation of a sequence in the target language would have a similar representation that includes the same information as the sequence in the source text (McCain et al. 2022:2). For a given source sentence, an NMT model produces the most probable translation based on what it has learned from its training data (Saunders/Byrne 2020:7724; Bahdanau/Cho/Bengio 2014:2).[1] One aspect included in the numerical representation of language, which we analyse further in this paper, is the information about the gender marked in the words (Gonen/Goldberg 2019:1).

3 Gendered Language

This section aims to define the term *gendered language* to better contextualise and understand the present analysis. In general, the term *gender* describes the social distinction of humans present in many societies which is based on their body (Kotthoff/Nübling 2018:14). It also describes which gender a person feels they belong to, and, in most cases, the gender identity (*gender*) corresponds to the sex assignment (*sexus*) made at birth (ibid.). Conventionally, people fall into two categories: male and female. However, current academic discussions and studies refer to the existence of more than these conventional two genders. Traditionally, people behave according to their gender and are even expected to do so. In a broader sense, gender not only refers to a biological distinction but also to culturally and historically evolved ways of how to dress, speak, behave, consume or act (ibid.). The idea of gender manifests itself in society not only through behaviour or attire but also in our language.

[1] A very good illustration of the operating principle of modern NMT systems for non-technical audiences can be found in Forcada (2017) and in Pérez-Ortiz et al. (2022).

Another recent issue closely related to gendered language is the generic masculine, which is used in many grammatical languages, such as Spanish or German, to refer to people across genders (Kotthoff/Nübling 2018:91). Traditionally, the generic masculine, such as the plural German term *Lehrer* for *teachers*, can refer simply to a group of male teachers or it can refer to a group of both male and female teachers. This is ambiguous and if the context is unclear, a reader cannot be certain whether only men or both men and women are referred to (Kotthoff/Nübling 2018:96f.). One could also say *Lehrer und Lehrerinnen*, explicitly referring to both male and female teachers. In a few cases, there is also a generic feminine (such as *Kosmetikerinnen* for *cosmeticians*), often associated with professions still significantly dominated by women, such as florists or cosmeticians, where men are a minority (Kotthoff/Nübling 2018:121).

In recent years, using the generic masculine has become somewhat controversial because it only *implicitly* includes all genders. Women or non-binary people are not *explicitly* referred to, meaning that the generic masculine can indirectly place the focus on men. To be inclusive of all gender groups, new spellings have been adopted in many languages to achieve a higher degree of equality. Examples in German include the asterisk (*Kanzler*innen*), the underscore (*Kanzler_innen*), the colon (*Kanzler:innen*) or using a capital *I* (*KanzlerInnen*) (Dudenredaktion 2022). These spellings are meant to explicitly refer to all genders. However, there is not yet a universally accepted 'gold standard', meaning that these different spellings coexist as of now and are still debated (sometimes fiercely so) in society. It is also worth mentioning that the topic has attracted the attention of experts in the field of specialised communication, where various ideas and approaches are currently being presented and discussed (e.g. Evers et al. 2022 from the perspective of terminology).

In the context of the present paper, it is important to note that most of today's training data does not yet contain gendered language. Therefore, two questions concerning machine translation are: Will NMT systems trained on current data be able to recognise patterns in gendered language and correctly render these patterns in the respective target language? And what would be the correct gender-inclusive spelling, given that society is still debating this topic?

4 Social Bias in Machine Translation

In neutral terms, biases can be thought of as inclinations towards specific ways of thinking or behaving. Biases are not negative per se (Shah/Schwartz/Hovy 2020:5248) but they can be a result of prejudices and stereotypes leading to faulty and hasty judgments and shaping or even distorting individual and social patterns of perception (Wondrak 2014). Biases inherent in our society, such as prejudices, are reflected in the language we use (Czihlarz 2019). In natural language processing (NLP), machines that are trained on the language we as a society produce can amplify prejudices and biases reflected in this language (Vanmassenhove et al. 2021:2). NMT as a specific NLP technology is no exception, as the engines 'learn' to translate by matching and adopting patterns found in their training data. This training data will automatically include certain patterns in language and of society that an NLP model may not be able to generalise to other demographics. This means that bias is an inherent property of NLP systems that, if not addressed, can have negative consequences (Shah/Schwartz/Hovy 2020:5248). Bias research in NMT is therefore not only a technical issue but also an ethical one.

4.1 Social Bias

Biases are a result of psychological heuristics. They are mental “shortcuts” that help us react faster in certain situations (Shah/Schwartz/Hovy 2020:5262). In the past, looking at people and judging them based on their appearance, skin colour, gender, age, clothing, language or accent was part of our survival instinct. In today’s modern society, however, it is a rather flawed and problematic characteristic as people tend to jump to (sometimes wrong) conclusions based on these judgements. Social biases are present in the personal characteristics and preferences of individuals, including their beliefs, attitudes, and choices that are shared within the group they belong to.

4.2 Gender Bias in MT

Gender bias or gender-related bias is one specific type of social bias. Gender bias results from formulations, mental assumptions or statistical errors that result in misrepresentations of actual gender relations (Stangl 2022). Gender manifests itself both in the agreement with other words in a sentence and the choice of context-based words or at the level of syntactic constructions, and this information is integrated into NMT systems (Vanmassenhove/Hardmeier/Way 2018:3003f.). NMT systems are often trained on unbalanced data that naturally refer more to men than to women, leading to translation outputs unknowingly perpetuating social biases (Costa-jussà/de Jorge 2020:26; Saunders/Byrne 2020:7724). Furthermore, NMT systems have traditionally translated on a sentence-by-sentence basis, and this lack of document-context information can lead to professions being translated with stereotyped genders (Basta et al. 2020:99). Figure 1 shows how gender stereotypes inherent in society (and therefore in NMT training data) lead to professions being translated predominantly as either male or female.

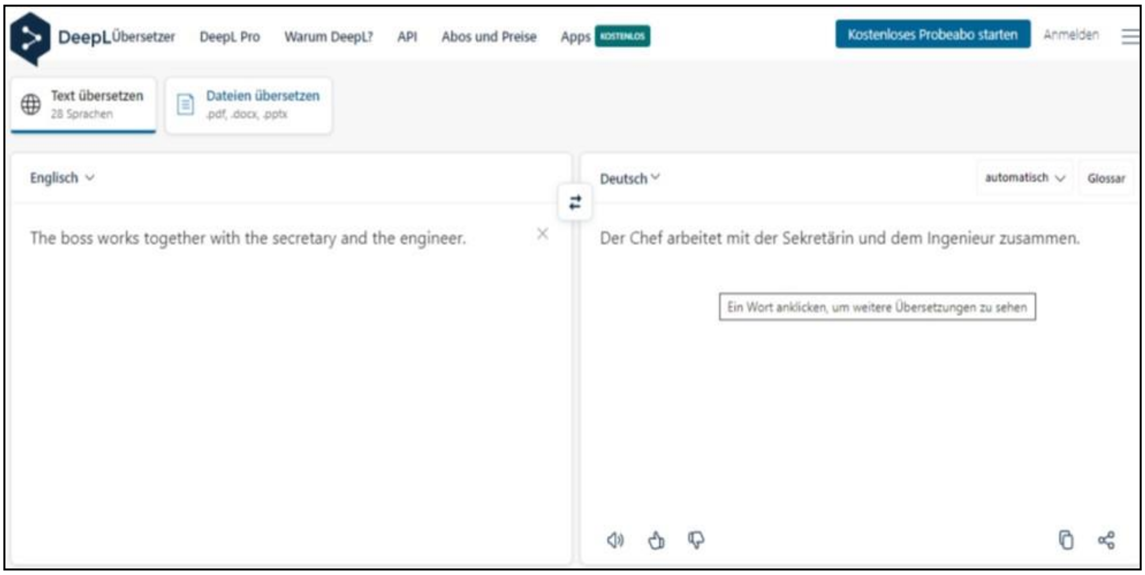

Fig. 1: DeepL translation from English into German exhibiting gender bias (15 November 2022)

In the example above, there are no contextual cues as to the gender of the persons mentioned in the sentence. In the German translation, DeepL therefore assigned gender based on societal stereotypes (which are present in the system’s training data): The *boss* is stereotypically male (*der Chef*), the *secretary* is stereotypically female (*die Sekretärin*) and the *engineer* is again stereotypically male (*der Ingenieur*).

Gender is not specified in all languages, but in case that it is specified in both the source and target languages, the gender should ideally be (machine-)translated correctly as expressed in the source text. However, there are genderless languages (e.g. Turkish/Finish) or languages that only display natural but no grammatical gender (e.g. English) (Savoldi et al. 2021:847f.). At the same time, there is a tendency to use gender-neutral vocabularies (e.g. *chairperson* instead of *chairman*/*chairwoman*). Yet, in light of all this diversity, there is a lack of established standards, both from a monolingual and multilingual perspective. In gender-ambiguous cases, NMT systems can only assume which gender should be used in the translation. These assumptions are based on statistical likelihood resulting from the data with which the systems were trained (Forcada 2017:295).

The architecture of an NMT system is a fundamental part of how it works. However, the training data is equally important. Usually, NMT systems are trained on substantial amounts of high-quality parallel translation data. Through this training, NMT systems learn to translate data they have not seen before (Koehn 2020:37). The better (e.g. larger, of higher quality, more balanced) the training data, the better the translations will be. But this training data can also be flawed or unbalanced, which can lead to equally flawed translations. A simplified example: If the training data for the translation direction English→German shows that *doctor* was translated more frequently as *Arzt* (male) than as *Ärtztin* (female), the system is more likely to use, even overuse, the male noun *Arzt* when translating the (gender-neutral= source noun *doctor*.

The type of data NMT systems are trained on will directly affect the MT output. Therefore, training datasets need to be adapted to help reduce bias in MT (Hovy et al. 2020:1689). There is a growing body of initiatives fostering the creation of more gender-balanced texts, such as the *DeBiasByUs* project (Daems/Hackenbuchner 2022), which aims to collect cases of gender-biased machine translations in order to create a research database. But as our data analysis in chapter 5 will show, this move towards gender-balanced language is not yet represented in any significant way in today's training datasets. Hence, many of the texts include biases, stereotypes and generalisations that are reflected in the translations of the MT systems trained on these texts. For more balanced training data, we have to take a closer look at how representative the data is and adapt it to better represent gender in language.

5 Methodology and Data Analysis

Having considered social bias in machine translation from a theoretical perspective, this section presents the methodology and the data analysis on which the findings and results discussed in the later sections of this paper are based.

Firstly, in order to create a test dataset, we wrote various texts containing at least one of the biases explained in previous chapters in English. These texts were specifically written to test MT systems and are supposed to provoke gender bias to some extent. Such datasets are often called *challenge sets* in MT research (Isabelle et al. 2017). For example, sample text 1 uses the gender-neutral pronouns *they* and *their* somewhat interchangeably with the pronouns *she*, *he*, *his* and *her*. Though this is not necessarily grammatically incorrect, this sample text is more inconsistent in its use of pronouns than a text would usually be if it wasn't written for this purpose. Even though the texts contain different biases, such as social, gender or racial bias, the focus of this paper lies on gender bias.

Subsequently, we translated these English texts into German using DeepL or Google Translate. In this way, we can observe how these commercial systems deal with existing bias in the source text and how they translate this bias into the target text. We then summarised and visualised the results in a tutorial video. You can find the texts in our DataLit^{MT} GitHub Repository and the video on our DataLit^{MT} Website.

6 Possible Solutions to Gender Bias in MT

When looking into the different biases that can arise when translating using an MT system, it becomes clear that some kind of effort should be made to find a solution to gender bias in MT. There are different ways to mitigate social biases in MT. However, as we discuss in this chapter, there is no one-size-fits-all approach.

6.1 User Solutions

Perhaps one of the easiest ways to help prevent bias in MT is to write unbiased source texts or debias them in pre-editing. Examples in English include using plural nouns and pronouns (*they*/*their*), replacing (singular) pronouns with generic nouns or removing unnecessary pronouns (ProEdit n.d.). Writing unambiguously can also lead to more bias-free MT outputs.

In practice, however, translators are rarely the authors of their source texts, nor do they often have the time and/or permission to pre-edit the source texts. And if a source text is written in a (genderless) language such as English, translators simply might not know or be able to disclose if a person is female, male or non-binary. Ideally, NMT systems should produce unbiased or gender-correct translations, thus decreasing the necessity of gender-related pre-editing (or post-editing).

6.2 Machine Solutions

A way to get a step closer to debias an NMT system is to train it on annotated datasets. In gender-annotated datasets, the speaker is tagged with gender information at the start of a sentence (Vanmassenhove et al. 2018:3004).

Furthermore, it might be helpful to use datasets exhibiting a higher frequency of female plural terms in order to create a more varied and balanced dataset. This balance should then also be reflected in the MT output (Vanmassenhove et al. 2018:3005). In this context, Saunders and Byrne, who addressed gender bias as a domain adaptation problem, suggest that using a "small, trusted gender-balanced set" to train an NMT system could reduce gender bias in machine translation as it "allow[s] more efficient and effective gender debiasing than a larger, noisier set" (2020:7724).

If the gender of a person in a source text is ambiguous (i.e. cannot be known, such as in "My neighbour brought me flour."), an alternative is showing the user this ambiguity and letting them choose between different options (different genders), as shown in figure 2.

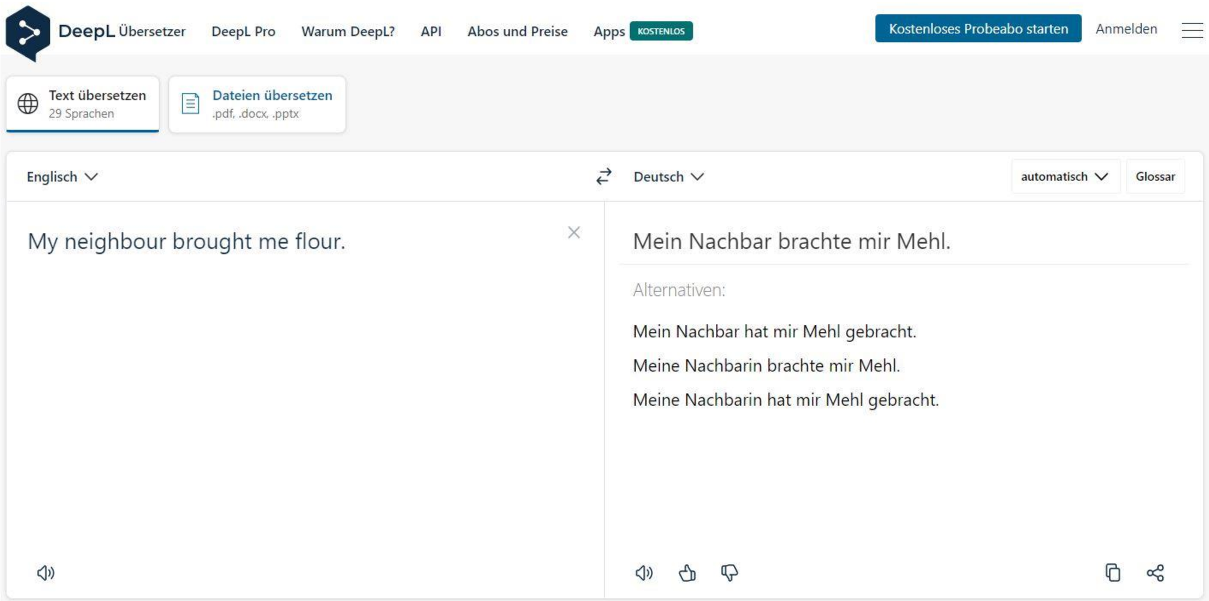

Fig. 2: Example of DeepL showing different translations for an ambiguous source sentence (24 November 2022)

In cases of ambiguity, systems like *Fairslator* (Fairslator 2022) allow the user to manually select the gender of the person, and adjust the output accordingly. The user/translator can choose between different options and has the opportunity to provide more information about the source text so that the resulting translation is as bias-free as possible (Měchura 2022). This feature is demonstrated in the tutorial video.

Another pilot project aimed at analysing how susceptible a source text might be to bias, such as gender or racial bias, was developed by Word2Vec (Mouroum 2022). The Word2Vec team developed a (prototype) computational solution to analyse text, such as the sentence:

“As a nurse Peter had to face long working hours, forcing him to quit his job and spend more time with his children at home. Now he is considering becoming a monk or a police officer.”

How susceptible certain words in this sentence are to gender and racial bias, can be analysed with Word2Vec (see figure 3).

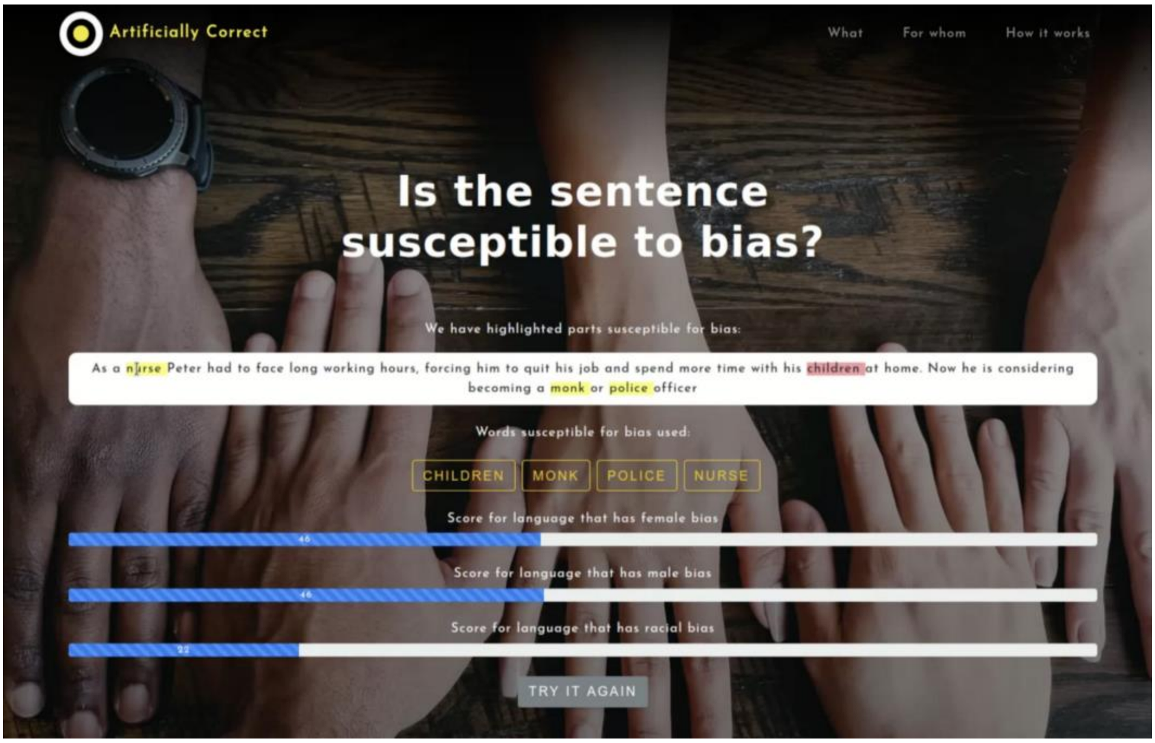

Fig. 3: Word2Vec analysing what words in a sentence are susceptible to bias (Mouroum 2022)

A growing body of research in the field of gender bias in MT is investigating potential solutions for analysing or disambiguating texts from a bias perspective and for training NMT systems with the aim of debiasing these systems during the training stage. Savoldi et al. (2021) summarise research conducted on the topic of gender bias reduction in MT and provide a unified framework for further research in this field.

6.3 Issues with Debiasing

Even though it is important to debias MT systems and MT output, translators might face several types of issues when it comes to debiasing translations. Debiasing, particularly in the case of gendered language, may unnecessarily lengthen the text or even make it less readable for some target groups (see example: target sentence options 1 & 2).

Example:

- Source sentence: *The* ***receptionists*** *were kind.*
- Target sentence:
  - Option 1: *Die* ***Empfangsmitarbeiter:innen*** *sind nett.*
  - Option 2: *Die* ***Empfangsmitarbeiterinnen und Empfangsmitarbeiter*** *sind nett.*

If the source text is written in Simple or Easy Language and is translated into gendered German, it may fail in its task of being easy and understandable. For this style of writing, it is especially important to write as simply as possible, which is why Maaß (2020), for example, advocates avoiding gender-sensitive language in Simple or Easy Language texts. However, the use of gender-sensitive variants may be permissible “in contexts where they [d]o not disproportionately burden the sentence and are concordant with the text functionality” (ibid.:247).

It may also be the case that gender is overemphasised in translation and thus more emphasis is placed on this aspect than is desirable (see, e.g., target sentence option 2 in our example above). The main message of the source sentence in the example is not *who* is working there but *how* they are working. However, in the gendered German translation, the emphasis appears to be on *who* is working at the reception.

This issue is linked to overcomplicating language that is meant to be easy to understand. If a target text is more precise than its more ambiguous source text, this may also lead to errors and misunderstandings finding their way into the text or may impact the readability of the target text. Nevertheless, it is important to be aware of aspects of gender bias and other social biases in translation, since translation is an important means of communication in today's globalised world and therefore plays and important role in developing more bias-free language and fairer societies.

7 Conclusion

In this paper, we discussed gender bias as a specific form of social bias that can be present in machine translation. We showed how gender bias can be introduced in MT systems, illustrated some examples of gender bias in the output of MT systems and discussed several solutions to gender bias in MT. We hope to have shown that bias in NLP systems is usually correlated with bias in the data used to train these systems. In the context of DataLit^{MT}, the phenomenon of such data-derived bias is linked, e.g., to our ability to critically think about applying data and data-driven systems and to recognise the ethical dimension of working with potentially biased data and systems, our ability to evaluate data with a focus on identifying and removing bias from such data, and to our ability to make informed decisions with regard to potential biases in the data and data-driven systems that we use as members of society or experts in a particular professional field.

References

Bahdanau, Dzmitry/Cho, Kyunghyun/Bengio, Yoshua (2014): Neural machine translation by jointly learning to align and translate. In: Bengio, Yoshua/LeCunn, Yann (Eds.): Proceedings of the 3rd international conference on learning representations (ICLR 2015), 1–15. Published on arXiv. https://doi.org/10.48550/arXiv.1409.0473.

Basta, Christine/Costa-Jussà, Marta R./Follonosa, José A. R. (2020): Towards mitigating gender bias in a decoder-based neural machine translation model by adding contextual information. In: Cunha, Rossana/Shaikh, Samira/Varis, Erika/Georgi, Ryan/Tsai, Alicia/Anastasopoulos, Antonios/Raghavi Chandu, Khyathi (Eds.): *Proceedings of the fourth widening natural language processing workshop.* Association for Computational Linguistics, 99–102. https://aclanthology.org/2020.winlp-1.25 (25 November 2022).

Costa-Jussà, Marta R./de Jorge, Adrià (2020): Fine-tuning neural machine translation on gender-balanced datasets. In: Costa-jussà, Marta R./Hardmeier, Christian/Radford, Will/Webster, Kellie (Eds.): *Proceedings of the second workshop on gender bias in natural language processing*. Association for Computational Linguistics, 26–34. https://aclanthology.org/2020.gebnlp-1.3 (25 November 2022).

Czihlarz, Jochen (2019): Kann gendergerechte Sprache Vorurteile verringern? *ANTI-BIAS*. https://www.anti-bias.eu/anti-bias-strategien/praxisbeispiele/gendergerechte-sprache/ (25 November 2022).

Daems, Joke/Hackenbuchner, Janiça (2022): DeBiasByUs: Raising awareness and creating a database of MT bias. In: Macken, Lieve/Rufener, Andrew/Van den Bogaert, Joachim/Daems, Joke/Tezcan, Arda/Vanroy, Bram/Fonteyne, Margot/Barrault, Loïc/ Costa-jussà, Marta R./Kemp, Ellie/Pilos, Spyridon/Declercq, Christophe/Koponen, Maarit/Forcada, Mikel L./Scarton, Carolina/Moniz, Helena (Eds.): *Proceedings of the 23rd annual conference of the European Association for Machine Translation*. European Association for Machine Translation, 289f. https://biblio.ugent.be/publication/8761022 (25 November 2022).

Dudenredaktion (2022): Geschlechtergerechter Sprachgebrauch. *Duden*. https://www.duden.de/sprachwissen/sprachratgeber/Geschlechtergerechter-Sprachgebrauch (29 November 2022).

Evers, Elisabeth/Früh, Beate/Schmitz, Klaus-Dirk (2022): Gendern terminologisch betrachtet. In: *technische kommunikation* 44(6), 34–38. https://technischekommunikation.info/ (29 November 2022).

Fairslator (2022): Fairslator.com. https://www.fairslator.com/ (29 November 2022).

Forcada, Mikel L. (2017): Making sense of neural machine translation. In: *Translation Spaces* 6(2), 291–309. https://doi.org/10.1075/ts.6.2.06for.

Gonen, Hila/Goldberg, Yoav (2019): Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. *arXiv*. https://doi.org/10.48550/arXiv.1903.03862.

Hovy, Dirk/Bianchi, Federico/Fornaciari, Tommaso (2020): "You sound just like your father". Commercial machine translation systems include stylistic biases. In: Jurafsky, Dan/Chai, Joyce/Schluter, Natalie/Tetreault, Joel (Eds.): *Proceedings of the 58th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 1686–1690. https://doi.org/10.18653/v1/2020.acl-main.154.

Isabelle, Pierre/Cherry, Collin/Foster, George (2017): A challenge set approach to evaluating machine translation. In: Palma, Martha/Hwa, Rebecca/Riedel, Sebastian (Eds.): *Proceedings of the 2017 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2486–2496. http://dx.doi.org/10.18653/v1/D17-1263.

Koehn, Phillip (2020): *Neural machine translation*. Cambridge: University Press.

Kotthoff, Helga/Nübling, Damaris (2018): *Genderlinguistik: Eine Einführung in Sprache, Gespräch und Geschlecht*. Tübingen. Narr Francke Attempto.

Maaß, Christiane (2020): *Easy language – plain language – easy language plus: Balancing comprehensibility and acceptability*. Berlin: Frank & Timme. https://doi.org/10.26530/20.500.12657/42089.

McCain, Miles/Garg, Rhythm/Barakaiev, Igor (2022): Lasts4Ever: Language-agnostic semantic text similarity for "every" language. Stanford CS224N custom project. https://web.stanford.edu/class/cs224n/reports/custom_116626702.pdf (29 November 2022).

Měchura, Michal (2022): About this project. *Fairslator*. https://www.fairslator.com (29 November 2022).

Mouroum, Marvin (2022): How we fight bias, sexism and racism with AI… *Medium*. https://medium.com/@mouroum.m/how-we-fight-bias-sexism-and-racism-with-ai-3f228478aba3 (29 November 2022).

Pérez-Ortiz, Juan Antonio/Forcada, Mikel L./Sánchez-Martínez, Felipe (2022): How neural machine translation works. In: Kenny, Dorothy (Ed.): *Machine translation for everyone. Empowering users in the age of artificial intelligence*. Berlin: Language Science Press, 141–164. https://doi.org/10.5281/zenodo.6653406.

Peters, Matthew E./Neumann, Mark/Iyyer, Mohit/Gardner, Matt/Clark, Christopher/Lee, Kenton/Zettlemoyer, Luke (2018): Deep contextualized word representations. *arXiv*. https://doi.org/10.48550/arXiv.1802.05365.

ProEdit (n.d.): Tips for avoiding gender bias in writing. *ProEdit*. https://proedit.com/tips-for-avoiding-gender-bias-in-writing/ (29 November 2022).

Rama, Arbnor/Vanmassenhove, Eva (2021): A comparison of different NMT approaches to low-resource Dutch-Albanian machine translation. In: Ortega, John/Ojha, Atul Kr./Kann, Katharina/Liu, Chao-Hong (Eds.): *Proceedings of the 4th workshop on technologies for MT of low resource languages (LoResMT2021)*. Association for Machine Translation in the Americas, 68–77. https://aclanthology.org/2021.mtsummit-loresmt.7/ (25 November 2022).

Saunders, Danielle/Byrne, Bill (2020): Reducing gender bias in neural machine translation as a domain adaptation problem. *arXiv*. https://doi.org/10.48550/arXiv.2004.04498.

Savoldi, Beatrice/Gaido, Marco/Bentivogli, Luisa/Negri, Matteo/Turchi, Marco (2021): Gender bias in machine translation. In: Roark, Brian/Nenova, Ani (Eds.): *Transactions of the Association for Computational Linguistics* 9. Cambridge: MIT Press, 845–874. https://doi.org/10.1162/tacl_a_00401.

Shah, Deven/Schwartz, H. Andrew/Hovy, Dirk (2020): Predictive biases in natural language processing models: A conceptual framework and overview. In: Jurafsky, Dan/Chai, Joyce/Schluter, Natalie/Tetreault, Joel (Eds.): *Proceedings of the 58th annual meeting of the Association for Computational Linguistics.* Association for Computational Linguistics, 5248–5284. https://aclanthology.org/2020.acl-main.468 (25 November 2022).

Stangl, Werner (2022): Gender Bias. *Online-Lexikon für Psychologie und Pädagogik*. https://lexikon.stangl.eu/30369/gender-bias (29 November 2022).

UEPO.de (2022): Gender Bias: Google Translate und DeepL übersetzen „el canciller“ mit „die Bundeskanzlerin“. *Uepo.de*. https://uepo.de/2022/01/17/gender-bias-google-translate-und-deepl-uebersetzen-el-canciller-mit-die-bundeskanzlerin/ (29 November 2022).

Vanmassenhove, Eva/Hardmeier, Christian/Way, Andy (2018): Getting gender right in neural machine translation. In: Riloff, Ellen/Chiang, David/Hockenmaier, Julia/Tsujii, Jun'ichi (Eds.): *Proceedings of the 2018 conference on empirical methods in natural language processing*. Brussels: Association for Computational Linguistics, 3003–3008. https://doi.org/10.18653/v1/D18-1334.

Vanmassenhove, Eva/Shterionov, Dimitar/Gwilliam, Matthew (2021): Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. *arXiv*. https://doi.org/10.48550/arXiv.2102.00287.

Wondrak, Manfred (2014): Wie lautet die Unconscious Bias Definition? Wie unterscheiden sich Stereotype von Vorurteilen? *ANTI-BIAS*. https://www.anti-bias.eu/wissen/definitionen/unconsciousbias-definition/ (29 November 2022).