Introduction:

Google is much more than a search engine. In 2018 Google Translate was one of the largest publicly available machine translation tools in existence with 200 million users a day.[1] Google Translate stands apart from other online translation services with its ability to translate email messages, documents, YouTube video captions, instant messages and chats, and anything else you can access in the Google family of Web and cloud applications.[2] Google is a powerful member of the digital technology oligopoly commonly referred to as GAFAM (Google, Apple, Facebook, Amazon, and Microsoft), which touches every corner of the world and by default the plethora of languages spoken and written including 108 languages translatable with Google Translate. Inspired by the paper *Assessing gender bias in machine translation: a case study with Google Translate* by Prates, Avelar and Lamb, which claims that automated translation tools can be exploited through gender neutral languages to yield a window into the phenomenon of gender bias in AI, this paper seeks to verify the paper's findings three years later with analysis of two gender neutral languages: Japanese and Vietnamese. Since the paper by Prates, Avelar and Lamb was written in 2018, Google Translate has updated its software in an attempt to mitigate gender bias in its translations, therefore this paper also serves as an audit of Google's adjustments to its translation algorithm.

While this paper focuses on gender bias specifically, the focus on Google Translate as an AI tool allows for the broader assessment of machine bias, as defined in the paper by Prates, Avelar and Lamb as how "trained statistical models—unbeknownst to their creators—grow to reflect controversial societal asymmetries, such as gender or racial bias." Examples from the paper by Prates, Avelar and Lamb include Apple's Iphone X failing to differentiate between two distinct Asian people and Google photos classifying black people as gorillas. The research question highlights the Google Translate machine bias with respect to gender by exploring the gender Google Translate assigns to professions from translations of the two selected gender neutral languages: Vietnamese and Japanese in 2021. In addition to addressing this question, this paper highlights the evolution of Google Translate as a tool as well as the historical, political, and cultural contexts of the Vietnamese and Japanese languages.

Method:

The research is based on the paper *Assessing gender bias in machine translation: a case study with Google Translate* by Marcelo Prates, Pedro Avelar and Luis C. Lamb which was published in 2018 and examined the occurrence of potential gender bias in Google Translate.[3] The authors mainly focused on biases in professions and certain selected adjectives attributed to specific genders. They compiled a dataset utilizing a list of job positions from the U.S. Bureau of Labor Statistics (BLS) to build sentences in constructions like "He/She is an Engineer" (where "Engineer" is replaced by the job position of interest) in 12 different gender neutral languages such as Hungarian, Chinese, Yoruba, and Japanese. For example, they took sentences such as "He/she/it is a nurse", which in Hungarian is translated as "ő egy ápolónő", whereby "ő" is a gender-neutral pronoun replacing he/she/it. Using the Google Translate API, this was then translated back to English to see whether the algorithm would replace the pronoun by a male, female or neutral version in English. The authors then compiled the

---

[1] Prates, M.O.R., Avelar, P.H. & Lamb, L.C. Assessing gender bias in machine translation: a case study with Google Translate. Neural Comput & Applic 32, 6363–6381 (2020). https://doi.org.acces-distant.sciencespo.fr/10.1007/s00521-019-04144-6

[2] Johnson, G. (2012) Google Translate http://translate.google.com. Technical services quarterly. [Online] 29 (2), 165–165.

[3] Ibid

percentage of female and male translations over the various occupation categories in the dataset to show the bias. These numbers were also compared to the percentages of female and male employment participation in each sector as shown by the BLS, to verify whether the bias in the translations was a reproduction of existing biases in society. The paper by Prates, Avelar and Lamb found that Google Translate yielded male defaults much more frequently than what would be expected from demographic data alone. Additionally, despite a noticeable level of variation among languages and categories, the null hypothesis that male pronouns are not significantly more frequent than female ones was consistently rejected for all languages and all categories examined in the paper. While the paper acknowledges that the distribution of translated gender pronouns may deviate from 50:50, it should not deviate to the extent of misrepresenting the demographics of job positions and that statistics of gender pronouns in Google Translate outputs could reflect the demographics of male-domninated fields, the findings show that Google Translate outputs failed to follow the real-world distribution of female workers across a comprehensive set of job positions.

As the goal of this research is to analyze the results of Prates, Avelar and Lamb's study in greater detail and to verify whether or not Google, as publicly announced, managed to fix the problem of gender bias, we selected two gender-neutral languages to examine and compare. We chose Vietnamese as a gender-neutral language that had not been included in the paper by Prates, Avelar and Lamb, and Japanese, as this was one of the most extremely biased languages in the original study and also provided an interesting contrast to Vietnamese in terms of their capitalist and communist historical contexts. We researched these contexts, as well as the linguistic notions of gender in these presumably neutral languages to see whether they might account for potential differences in the results. Using the same occupations included in the original study, as published on the Github Repository https://github.com/marceloprates/Gender-Bias we created a dataset with the same method. However, we focused on occupations only and left out the list of adjectives. Furthermore, we also summarised the datasets for artistic, dance, theatre and writing jobs into the one category of artistic jobs. The resulting translations were then compared to labour statistics retrieved from the Statistics Bureau at the Ministry of Internal Affairs and Communications of Japan and the Labor Force Survey on Vietnam published by the International Labour Organization to further contextualize the results.

Evolution of Google Translate

Google Translate has adapted and improved overtime, particularly following several high profile cases of discrimination. In 2015, users discovered that the word 'гей', the Russian equivalent of 'gay', produced suggestions including faggot, fag, fairy, queen, sodomite and pansy boy while other homophobic terms appeared in translations from English into French, Spanish and Portuguese.[4] A Google spokesperson responded to this incident by explaining "Our systems produce translations automatically based on existing translations on the web, so we appreciate when users point out issues such as this." This explanation echoes an important reflection from the paper by Prates, Avelar and Lamb, which emphasizes that it is not feasible to train these models on unbiased texts, since such texts are scarce. However, it is possible to engineer solutions to remove bias from the system after an initial training.

[4]Woollacott, Emma. "Google Apologizes, Fixes Homophobic Slurs In Translator." Forbes, Forbes Magazine, 28 Jan. 2015, www.forbes.com/sites/emmawoollacott/2015/01/28/google-apologizes-fixes-homophobic-slurs-in-translator/?sh=473724036424.

In 2016, Google Translate began to transition from a purely phrase-based statistical machine translation system into a system that uses deep learning and artificial neural networks that adapt and learn.[5] This new artificial neural approach attempts to read, understand, and translate sentences as units, guessing through statistical evidence, context, and previous experience what the words might mean. However, the system still could not understand the original text nor the text it produced, leaving it intelligent but not conscious.

In 2018, shortly after the publication of the paper by Prates, Avelar and Lamb, Google began to offer both feminine and masculine translations for single words when translating from English to French, Italian, Portuguese or Spanish, as well as when translating from Turkish to English.[6] Notably, only one of these languages is gender neutral and was treated in the paper by Prates, Avelar and Lamb, this language is Turkish. Google also expressed ambitions to address non-binary gender in translations and integrate updates to iOS and Android apps, and address gender biases in auto-complete.

Historical comparison of Vietnam and Japan' gender equality conceptions

Comparing two east Asian nations sheds light on the specific ideologies which explain the evolution of gender equality over time. The equal treatment in terms of gender can be seen through four main narratives : economic, organizational, political and cultural.[7]

Both Japan and Vietnam established equal rights, namely a legal equality in their constitutions. Adopted in 1949, the vietnamese constitution stipulated that *'women are equal to men in all respects'* while the japanese supreme law considered men and women *'separate but equal'*. In Japan, the housewife role was not considered as oppressive but as an autonomous role helping women increase their skills to manage the functioning of their families. Both genders were viewed as contributing to economic growth based on extensive work for men and domestic tasks for women.

Both the Vietnam war of reunification, and World War II (WWII) in Japan provided opportunities which led to a labor market feminization. In Vietnam, the Women's Union - a loyal opposition within the political bureaucracy which aimed to safeguard and promote women's interests since 1930 - implemented job quotas, which led to women embodied 35% of all jobs in education, medicine and light industry. Additionally, females were promoted to leading positions and stood for 40.9% in people's council in 1972.[8] In Japan, the end of the U.S occupation authorities in 1952, and the loss of men during the conflict favoured women. They were involved in rebuilding the country in several spheres. This commitment allowed a tremendous economic growth[9] two years after the end of WWII.

---

[5] Constantine, P. (2019) Google Translate Gets Voltaire: Literary Translation and the Age of Artificial Intelligence. Contemporary French and francophone studies. [Online] 23 (4), 471–479.

[6] Dickey, Megan Rose. "Google Translate Gets Rid of Some Gender Biases." TechCrunch, TechCrunch, 7 Dec. 2018, techcrunch.com/2018/12/07/google-translate-gets-rid-of-some-gender-biases/?guccounter=1.

[7] Blau, Brinton, and Grusky (2006), The declining significance of gender ? New York : Russell Sage Foundation 296p

[8] Goodkind, Daniel, (1995) "Rising Gender Inequality in Vietnam Since Reunification," Pacific Affairs , Autumn, 1995, Vol. 68, No. 3 (Autumn, 1995), pp. 342-359

[9] Kristen Schultz Lee et al. (2010) Separate Spheres or Increasing Equality? Changing Gender Beliefs in Postwar Japan. *Journal of marriage and family*. [Online] 72 (1), 184–201

Nevertheless, the conception of gender equality in those countries remained different, even divergent. The war of reunification in Vietnamese symbolised a socialist trend during the Cold War, through the leader Ho Chi Minh, whereas Japan was a capitalist country willing to tolerate some inequalities to promote productivity. Due to marxism, it seems that women benefited from a definition of equality that improved their status, even though Vietnam experienced extreme poverty. The end of the ownership system coupled with the proletarian revolution questioned the place of women in the transition to socialism. A great example of women integration is the state investment in education during the 1970s, where gender equality in terms of language has been set up in order to attract females in universities.[10] Literacy programs during the socialist era were thus a major way to implement gender equality in education and thus in professional positions. A neutral language was a path to equal treatment.

However, *Doi Moi* during the 1980s in Vietnam was a several-year-plan policy which aimed to allow capitalistic aperture to foreign investment, thanks to a transition to a free market. Although the Vietnam Women's Union tried to secure women's welfare, the economical cut in maternity leave in 1994 by the state and companies increased gender segregation in the Vietnamese labor market. Japan was confronted in the same period with an economic crisis. While it had abolished the family system after World War II, the state discouraged women to work outside home. Women benefited from tax deductions if they earned less than 1.35 millions of yen per year, and there was a lack of affordable child care for children under the age of three.[11]

The history of gender equality has a significant influence in the social structure of Vietnam and Japan. While the World Economic Forum ranked Vietnam #31 in the world for gender equality in 2020, Japan places far behind at #91.[12]

Notions of gender and gender neutrality in the Vietnamese and Japanese languages

In addition to the historical notion of gender-equality, it is important to further account for the role gender plays in a country's language. The feminist movement has long criticised androcentric grammatical structures in many languages, which disadvantage girls and women in their social and professional lives.[13] These ideas are based on theorizations stating that grammatical gender constructions can influence our cognitive understanding of our surroundings, thus impacting the relative standing of different genders in society.[14] Currently, there are no existing languages which do not distinguish between different genders at all. However, they do so to varying degrees. Consequently, in their quantitative study Prewitt-Freilino and Caswell have found that countries in which gendered languages are dominant (which refers to languages that assign a masculine, feminine or sometimes neutral gender to nouns and adjectives and pronouns are adapted accordingly) correlate

---

[10] Goodkind, Daniel, (1995) "Rising Gender Inequality in Vietnam Since Reunification," Pacific Affairs , Autumn, 1995, Vol. 68, No. 3 (Autumn, 1995), pp. 342-359

[11] Kristen Schultz Lee et al. in *Separate Spheres or Increasing Equality? Changing Gender Beliefs in Postwar Japan. Journal of marriage and family*

[12] "Global Gender Gap Report 2020," World Economic Forum accessed at http://www3.weforum.org/docs/WEF_GGGR_2020.pdf

[13] Stahlberg, D., Braun, F., Irmen, L., and Sczesny, S., (2007), "Representation of the sexes in language", in K. Fiedler (Ed.), *Social communication: 170.* New York: Psychology.

[14] Prewitt-Freilino, Jennifer L., and Caswell, T. Andrew, (2012), "The Gendering of Language: A Comparison of Gender Equality in Countries with Gendered, Natural Gender, and Genderless Languages", *Sex Roles*, 66: 268-269. DOI 10.1007/s11199-011-0083-5

with a lower degree of societal gender equality than countries where a natural gender language (languages that distinguish between different gendered pronouns but don't assign grammatical gender to nouns) or genderless language (languages with a complete lack of gender in their grammatical structure) is spoken.[15]

According to the categorization by Prewitt-Freilino and Caswell, English classifies as a natural gender language, as it shows gender differentiations through pronouns such as he or she but doesn't grammatically mark nouns as gendered. Vietnamese on the other hand is generally considered to be a genderless language.[16] However, gender still plays a role in connection with a person's age, as different pronouns are used for referring to older and younger people, also dependent on their gender.[17] For example, the pronouns *em* (for I) and *chi* (you) are used to address a female person older than the speaker but applied vice versa when the speaker (then referred to as *chi*) is older than the addressed female person (then referred to *em*). To address male persons, the pronouns *em* and *anh* would be used accordingly. Similar constructs are also applied to refer to a third person (he or she), but mostly used in the context of kinship. Similarly, according to Prates, Avelar and Lamb, Japanese can be considered as a gender-neutral language.[18] However, similar to Vietnamese, Japanese also includes a limited amount of gendered speech patterns, such as gendered personal pronouns and sentence-final particles.[19] The usage of gender in these grammatical forms is dependent on the degree of formality. For example, the word *watashi* in Japanese can be used both for male and female speakers, but is less formal for women than for men. The pronoun with corresponding formality to the female *watashi* would be the male pronoun *boku*. Further, there are several sentence-final particles that can express notions of masculine or feminine gender or also be neutral.

Results

Our analysis has shown that the Google API consistently provided translations with a 50/50 distribution of male and female pronouns across all occupation categories for both Vietnamese and Japanese, as is shown in Table 1. The only exceptions are some minor deviations for Vietnamese in the categories of artistic occupations (51% female and 49% male), corporate occupations (47% female and 53% male) and industrial occupations (51% female and 49% male). This proves that the changes that were enacted by Google to remove the bias in their translator API after the study of Prates, Avelar and Lamb was published, were effective. Google Translate now consistently provides a male and a female version for each translation in the dataset and thus adheres to societal expectations of gender equality. If we compare these results to the labour statistics for all categories available for both Japan and Vietnam, we can see that contrary to the results in the initial study, in terms of occupations Google Translate is now even less biased than the actual distribution in society, as shown in Figures 1

---

[15] Ibid
[16] Ibid
[17] "Say Pronouns in Vietnamese: I/You/We and My/Your/Our", n.d., yourvietnamese.com, accessed at https://yourvietnamese.com/learn-vietnamese/vietnamese-pronouns/#:~:text=The%20common%20canonical%20Vietnamese%20words,to%20me%20or%20of%20me.
[18] Prates, M., Avelar, P., and Lamb, L. C., (2020), "Assessing Gender Bias in Machine Translation – A Case Study with Google Translate", Neural Computing and Applications, 32: 6368. Accessed at https://link.springer.com/article/10.1007/s00521-019-04144-6
[19] Tompowsky, R. (2014). "An Exploration of Gender-specific Language in Japanese Popular Culture", *University of Gothenborg*: 11-13, accessed at https://gupea.ub.gu.se/bitstream/2077/35154/1/gupea_2077_35154_1.pdf

and 2 and Table 2 accordingly. Both Japan and Vietnam still show an unequal distribution of women participation for most occupation categories, for example they still make up the majority of the workforce in the healthcare sector (75.3% in Japan and 60.8% in Vietnam) whereas the service sector, which in the dataset by Prates, Avelar and Lamb mainly consists of jobs in transportation such as taxi or bus drivers, is still largely dominated by men (78.7% in Japan and 90.9% in Vietnam).

Job Category	Japanese		Vietnamese	
	Female	Male	Female	Male
Artistic	50%	50%	51%	49%
Computer	50%	50%	50%	50%
Corporate	50%	50%	47%	53%
Healthcare	50%	50%	50%	50%
Industrial	50%	50%	51%	49%
Science	50%	50%	50%	50%
Service	50%	50%	50%	50%

Table 1: Percentage of female and male pronouns obtained for each occupation category for the two tested languages Japanese and Vietnamese.

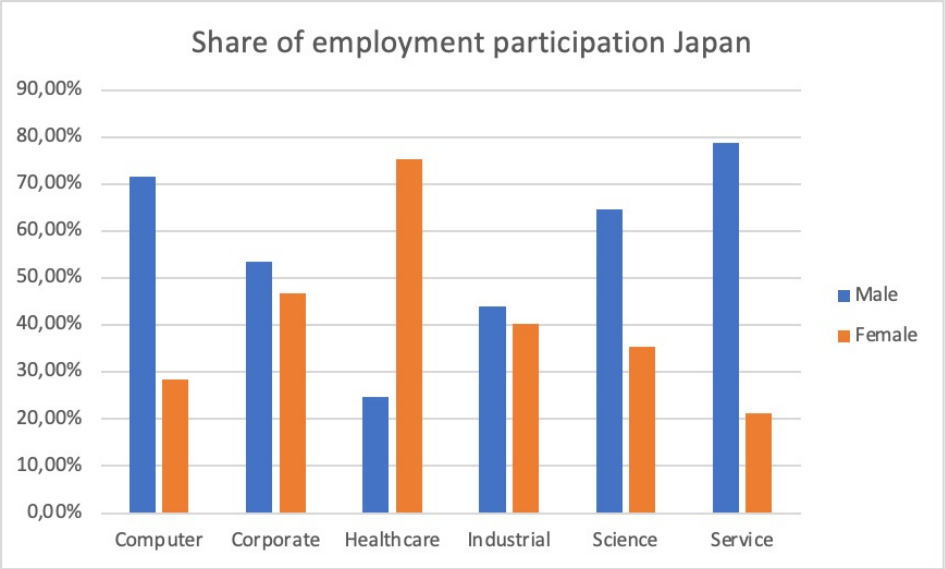

Figure 1: Share of employment participation for males and females per occupation category in Japan. Note: To arrive at comparable results, several categories from the statistics obtained from the Japanese Statistics Bureau had to be combined by taking their average in order to generate categories including similar jobs as in the dataset used in the original study. No data was available for the original category of artistic occupations. For more detailed information refer to Table 2.[20]

[20] Statistics Bureau. (2020). "Statistical Handbook of Japan". *Ministry of Internal Affairs and Communications Japan*: 128. Accessed at: https://www.catalyst.org/research/women-in-the-workforce-japan/

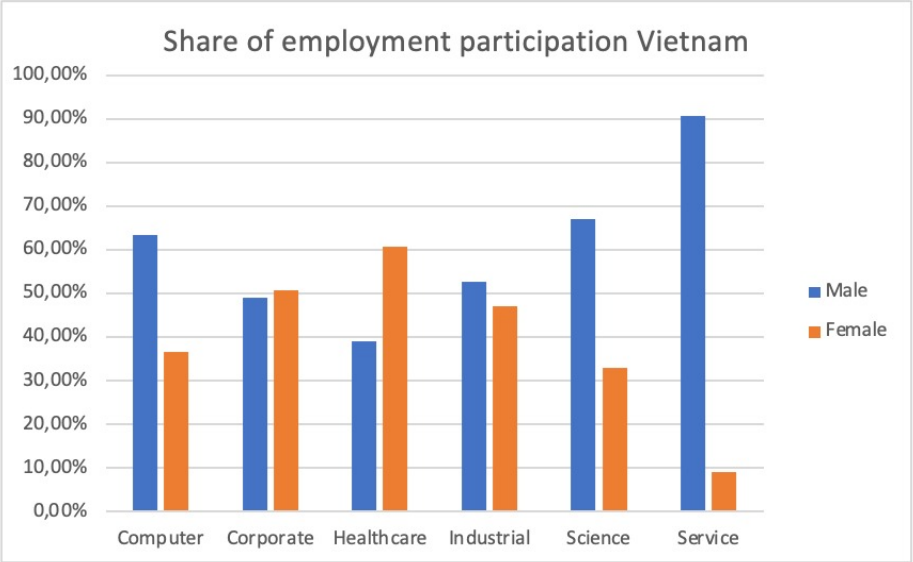

Figure 2: Share of employment participation for males and females per occupation category in Vietnam. Note: To arrive at comparable results, several categories from the statistics obtained from the International Labour Organization had to be combined by taking their average in order to generate categories including similar jobs as in the dataset used in the original study. No data was available for the original category of artistic occupations. For more detailed information refer to Table 2.[21]

Occupation category	Male share of employment Japan	Female share of employment Japan	Original categories Japan	Male share of employment Vietna	Female share of employment Viet	Original categories Vietnam
Artistic (incl. Dance, film, theatre, writing)	/	/	/	/	/	/
Computer	71,60%	28,40%	Information & communications	63,40%	36,60%	Information and communication
Corporate	53,40%	46,70%	Finance and insurance; real estate and goods rental and leasing *(1) (2)*	49,20%	50,80%	Financial intermediation, banking and insurance; real estate activities *(5) (6)*
Healthcare	24,70%	75,30%	Medical, health care and welfare	39,20%	60,80%	Health and social work
Industrial	43,90%	40,30%	Manufacturing; Construction; acommodations, eating and drinking services; wholesale and retail trade *(3) (4)*	52,80%	47,20%	Manufacturing; Construction; wholesale and retail trade, repair of cars, motorcycles and other motor vehicles; hotels and restaurants *(7) (8)*
Science	64,60%	35,40%	Scientific research, professional and technical services	67,10%	32,90%	Technological, scientific, and specialized activities
Service	78,70%	21,30%	Transport and postal activities	90,90%	9,10%	Transport and storage

Remarks:
(1) Average value male. Finance and insurance: 47%; real estate and goods rental and leasing: 59.7%
(2) Average value female. Finance and insurance: 53%; real estate and goods rental and leasing: 40.3%
(3) Average value male. Manufacturing: 7%; construction: 83.2%; acommodations, eating and drinking services: 37.6%; wholesale and retail trade: 47.9%
(4) Average value female. Manufacturing: 30%, construction: 16.8%; acommodations, eating and drinking services: 62.4%; wholesale and retail trade: 52.1%
(5) Average value. Financial intermediation, banking and insurance: 44.2%; real estate activities: 54.2%
(6) Average value. Financial intermediation, banking and insurance: 55.8%; real estate activities: 45.8%
(7) Average value. Manufacturing: 45.3%; construction: 90%; wolesale and retail trade, repair of cars, motorcycles and other motor vehicles: 43.3%; hotels and restaurants: 32.7%
(8) Average value. Manufacturing: 54.7%; construction: 10%; wolesale and retail trade, repair of cars, motorcycles and other motor vehicles: 56.7%; hotels and restaurants: 67.3%

Table 2: Aggregated labour statistics on the share of employment participation of males and females per category in Japan and Vietnam. Note: To arrive at comparable results, several categories from the statistics obtained from the Japanese Statistics Bureau and the International Labour Organization had to be combined by taking their average in order to generate categories including similar jobs as in the dataset used in the original study. No data was available for the original category of artistic occupations.[22]

As we have elaborated in the previous sections, the historical and economic context of each country has had a significant impact on their notions of gender equality, with Vietnam as a formerly communist country currently ranking higher than Japan, which is still considered to be a largely unequal country in terms of gender. Whilst we cannot verify this due to the now equal distribution of

[21] General Statistics Office of Vietnam. (2016). "Labour Force Survey 2016". *International Labour Organization*: 29. Accessed at https://www.ilo.org/surveyLib/index.php/catalog/1837/related-materials
[22] Ibid.

male and female translations with the Google API, we strongly assume that this historical context is the reason why Japanese was one of the languages with the most biased outcomes in the original study. As Vietnamese was not originally analyzed by Prates, Avelar and Lamb, there is no possibility for comparison. However, it is likely that the translations from Vietnamese would have yielded a more equal distribution between male and female pronouns even before Google adapted its API, since Vietnam is a more gender-equal country overall. The influence of gender notions in society on a country's language and vice versa have also been briefly mentioned in the previous section on gender-neutral languages.

This becomes even more clear regarding languages generally considered to be gender-neutral, which can still include grammatically gendered constructs, even though they might play out differently than in the languages most researchers are familiar with. For example, in contrast to English, gender is usually expressed through sentence-final particles in Japanese and highly dependent upon the level of formality. In studies as the one conducted by Prates, Avelar and Lamb, such differences should be taken into account, at the minimum as a control variable in the following regression, as they might heavily influence the results. For example, while Prates, Avelar and Lamb tested the translations for gender bias mainly in regard to the use of gendered pronouns in relation with occupations, it is possible that the presumed gender-neutral base sentence in Japanese that was used for the translation did actually include gendered expressions in the form of sentence-final particles. If this is the case, then the highly biased outcomes for Japanese in the original study might not only be the result of the biased API, but also of other gendered expressions in the original sentences that were overlooked. This is something that we encountered in our analysis: When initially translating the basic sentences "He is a + occupation"/ "She is a + occupation" from English to Japanese to create the gender-neutral dataset, we noticed a difference in the Japanese sentences for the same occupation between the male and the female version with different signs at the end of the sentence. Thus, the sentences used as a base for the analysis by Prates, Avelar and Lamb might not have been as gender-neutral as they thought. However, they failed to take this into account as the template they used for Japanese (あの人は "occupation" です), contrary to most other languages in the study, doesn't translate to the form of "He/She is a + occupation" but rather as "This person/this man is a + occupation" (depending again on the translator used) and might thus have had an impact on the final results.

This leads to the conclusion that when analyzing gender bias in AI algorithms, it is important to consider the language's context to arrive at meaningful explanations of why these biases might exist. After all, algorithms are always a reflection of the existing biases in the data, in this case the language they are fed with. Whilst Prates, Avelar, and Lamb provided a larger-n qualitative study with a good initial overview of gender-bias in the translation of several gender-neutral languages, they failed to take into account the individual contexts of each language beyond the labour participation statistics. Our study has shown that further research is necessary to investigate the potential links between languages' historical, socioeconomic and linguistic notions of gender and the corresponding biases in translator algorithms.