



HAL
open science

Machine Translation and Gender biases in video game localisation: a corpus-based analysis

María Isabel Rivas Ginel, Sarah Theroine

► To cite this version:

María Isabel Rivas Ginel, Sarah Theroine. Machine Translation and Gender biases in video game localisation: a corpus-based analysis. 2022. hal-03540605v2

HAL Id: hal-03540605

<https://hal.science/hal-03540605v2>

Preprint submitted on 27 Oct 2022 (v2), last revised 9 Dec 2022 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Machine Translation and Gender biases in video game localisation: a corpus-based analysis

María Isabel Rivas Ginel¹, Sarah Theroine²

1 Universities of Bourgogne Franche-Comté & Valladolid, France & Spain

2 Université de Bourgogne, France

*Corresponding author: María Isabel Rivas Ginel isabelrivasginel@gmail.com

Sarah Theroine Sarah.Theroine@u-bourgogne.fr

Abstract

The video game industry has been a historically gender-biased terrain due to a higher number of male protagonists and hypersexualised representations [Dietz, 1998; Downs & Smith, 2010; Lynch et al., 2016]. Nowadays, echoing the debate on inclusive language, companies attempt to erase gender disparity by introducing more main female characters as well as non-binary characters. From a technological point of view, even though recent studies show that Machine Translation (MT) remains largely unadopted by individual video game localisers [Rivas Ginel, 2021, p. 171], multilanguage vendors are willing to invest in these tools to reduce costs [LIND, 2020, p. 50]. However, the predominance of the masculine in Natural Language Processing (NLP) and Machine Learning has created allocation and representation biases in Neural Machine Translation (NMT) [Crawford, 2017].

This paper aims to analyse the percentage of gender bias resulting from the use of Google Translate, DeepL, and SmartCat when translating raw in-game content from English into French. The games *DeltaRune*, *The Devil's Womb* and *The Faces of the Forest* were chosen due to the presence of non-binary characters, non-sexualized characters, and female protagonists. We compared the results in order to recount and analyse the differences between these tools' outputs in terms of errors related to gender. To this end, we created a parallel corpus to compare source documents and each translation to visualise the semantic and grammatical directions of the words embeddings [Zhou; Shi; Zhao and al, 2019] and extracted the collocations and concordance lines that represented gender identity by analysing the patterns in the source language.

keywords

Machine Translation, Neural Machine Translation, Video game Localisation, gender biases, corpora.

I CONTEXT

1.1 The video game industry

The increasing popularity of video games has, as a consequence, more than doubled the demands of the industry in terms of video game localisation to reach a larger public and increase the revenues generated by each game. Conversely, localisers' working conditions have not substantially improved over the last decade and, despite constant technological advances, they continue “to work in a double-blind process (no audiovisual context, no text linearity)” [Merino, 2013, p. 117]. In other words, localisers are not given access to the game itself and, even in the best conditions, struggle to identify the gender of the character or the addressee, among many other issues.

1.2 Motivation

The idea of focusing on the efficiency of Neural Machine Translation (NMT) by analysing gender-related issues in the field of video game localisation stems from three reasons. First, previous research on the degree of adoption of Machine Translation (MT) tools by video game localisers [Rivas Ginel, 2021, p. 171] has shown that even though MT remains largely unadopted (11,6% of regular users), “66% of agencies and 44% of in-house translation teams expect to invest in [MT] in 2020” [LIND, 2020, p. 50]. Furthermore, more than 70,37% of video game developers use Unity as their main game engine, which provides localisation management and instant MT options [Rivas Ginel, 2022, p. 320]. Therefore, it is likely that either developers or translation agencies might use MT and then send the text to video game localisers for post-editing to cut costs in a saturated market.

The second motivation behind this study is that the video game industry is moving with the times and increasing the number of female, gender-neutral and non-sexualised characters, leaving behind the image of a predominantly male domain. Although this evolution reflects the present trend in society towards diversity and inclusion, these changes entail new challenges for video game localisers, forcing them to redouble their efforts to avoid gender bias and search for neutral yet natural translations. Moreover, the mechanisms used for inclusivity vary from one language to another, notably with gendered languages such as those from Indo-European countries. Languages such as French, which tend to use masculine grammatical forms, can easily become ambiguous and cause the under-representation of women due to the fact that a pronoun in its masculine form can refer to a group of men, to a mixed-gender group of people, a neutral group of people, or to a person whose gender is not precise nor defined.

The third reason is the increasing number of papers that focus on the performance of NMT systems and explore solutions to reduce gender bias in MT. NMT’s performance can be affected by its ability to deal with Language Detection, Natural Language Processing and Natural Language Understanding; well-known notions in artificial intelligence used to detect, understand, and treat natural languages. Furthermore, the efficiency of these tools, especially if the source language is relatively neutral and the target language is grammatically gendered, is highly impacted by the methods used to train the artificial intelligence behind those systems. For example, an NMT system that translates a sentence without gender marks such as “the surgeon is very skilled” will automatically use the masculine form and produce the outcome “le chirurgien est très compétent”.

II METHODOLOGY

2.1 Choice of systems and games

2.1.1 Systems

We decided to analyse the output of Google Translate and DeepL—NMT-based—as well as SmartCat (to represent Statistical Machine Translation (SMT) tools). As the company explains in their blog [Google AI Blog], Google Neural Machine Translation (GNMT), which is integrated into Google Translate, was developed in 2016 and modified in 2020. It provides language detection and includes various technologies based on Artificial Intelligence that allow it to handle 100+ languages. GNMT uses a hybrid model architecture that replaces the previous RNN-based GNMT model—used to identify data correlations and patterns—with a transformer encoder (that reads the source text and searches for correlations) and an RNN

decoder for predictive typing [Google AI Blog]. These hybrid models improve the quality and lower the time of latency. Additionally, the 2020 update allows Google to web crawl using an embedding-based model rather than the previous phrase-based machine translation model of 2016 (dictionary-based crawler), thus improving the quantity and quality of the data and increasing the number of collected sentences [Google AI Blog]. Nevertheless, using parallel (or translated) texts as the linguistic data used in its training phase subjects the availability and quantity of paired texts and languages to their presence on the web, affecting the translation quality. Additionally, GNMT will be more accurate at translating terms or words from a source language to a target language rather than long sentences, entire texts, or complex idioms [Yang, Wang & Chu, 2020].

Once again using the company's own description [DeepL Blog], DeepL uses Deep Learning and Neural Networks and draws the linguistic data used for its training from Linguee, a massive corpus built around manually translated sentences, idioms, and texts. The quantity and diversity of the data utilised account for DeepL's degree of accuracy and performance in specialised domains and its capacity of offering translation alternatives such as synonyms and rephrased options [DeepL Blog]. Additionally, DeepL includes predictive typing which, combined with the translation alternatives function of DeepL, works similarly to translation memories and means that DeepL might be able to memorise the user's translation alternative when the latter changes a term or rephrases the entire sentence. In consequence, the user's alternative would increase the amount of data exploited by DeepL, which already uses Deep Learning to constantly train paired languages, allowing it to train with it and increasing the number of translation alternatives. In terms of MT, DeepL uses an API to provide restricted access to a glossary of terms which increases the accuracy. In addition, this API also limits the length of documents uploaded into DeepL meaning that the text is split into sentences, then translated and finally put back together [DeepL Blog].

Finally, as a representative of MT integrated into Computer Assisted Translation (CAT) tools and an example of SMT we chose SmartCat, which uses Yandex Free by default. This system is not based on translated texts or language rules but upon statistics, which means that the tool uses comparable corpora. SMT analyses source documents in one language and statistically compares them with related documents in another language first; then it gathers, relates, and aligns the data with parallel texts [Yandex Blog]. Another difference is that the system can easily and quickly be adapted when a language faces modifications (spelling, usage, new terms or idioms), which also means that it must be regularly updated to improve its quality. Yandex's language training is built around three models: a translation model, a language model, and a decoder. Those models are used together to sort all terms, idioms, and sentences extracted by language from the comparable corpora to find and match correlations of words, terms or sentences extracted from the parallel corpora [Yandex Blog]. The models draw statistical patterns from a first set of paired sentences and then use another set to constantly recalculate, compare, and match terms to find the more accurate equivalent in the target language. The translation is done by the decoder, which gathers all the translation alternatives for a given sentence or sequence, combines phrases from the translation model and statically classifies them in order of probability [Yandex Blog]. In this sense, the decoder's function can be assimilated to a phrased-based model or a Seq2Seq model. The decoder also uses the translation model to determine the highest probability of equivalence and frequency of use to improve the accuracy and idiomaticity.

2.1.2 Games

The games *The Faces of the Forest*, *The Devil's Womb* and *DeltaRune* were chosen due to the presence of female protagonists, non-binary characters, and non-sexualized characters. All of them are free to play and the .json documents containing the strings were either online or we were granted permission to use them for educational purposes. Given the difference in the number of words between the three documents, we decided to use the entirety of the two first games and their sum's equivalent in the case of *DeltaRune*. *The Faces of the Forest* has a little over 2350 words in total and 11 characters: 4 females, 6 males and 1 non-binary; the main character is Ana. *The Devil's Womb*, an RPG game with almost 14900 words and 4 possible endings, has over 30 characters, the protagonist is a female demon named Luna, the characters with the bigger roles are predominantly female (13 females), and there are many demons and minions whose gender was not specified. Finally, *DeltaRune* has more than 36000 words (although only the first 17900 were selected), it is an RPG as well, the main character is non-binary and has two companions, a female, and a male. The selected part contains 40 characters, 18 of them (the enemies mostly) are non-binary or simply referred to as "it"; there are 15 male characters and 7 female characters.

2.2 Process

After establishing the MT systems and the games, the translatable strings were automatically extracted from the .json files using an *ad-hoc* custom computer programme. We decided to work with this file format since previous research showed that .json files were used by 65,17% of the 379 developers who isolated gameplay text using external files [Rivas Ginel, 2022, p. 307]. Then, the text was copied and pasted into Google Translate and DeepL and, in the case of SmartCat, the file was uploaded and automatically translated without any other interference in order to automate the process as much as possible. Once again, to reproduce video game localisation practices [Rivas Ginel, 2022, p. 307], we used Excel files to display the source text and the translations. Afterwards, we created so-called "character bibles" of every game to catalogue characters, display their gender (or lack thereof), and include a screenshot for better reference. The bibles were used when the text was ambiguous and we needed to use walkthroughs to determine the identity of the speaker or the addressee.

				Nom	Description	Genre	Image
Source (EN)	Google (FR)	DeepL (FR)	SmartCat (FR)	Alphys	She is the teacher of Kris and other students	Female	
The Faces of the Forest	Les visages de la forêt	Les visages de la forêt	Les Visages de la Forêt	A Man	The man is a hidden NPC the party can encounter in Forest.	Male	none
Start new game	Commencer une nouvelle partie	Commencer une nouvelle partie	Commencer une nouvelle partie	Asggore Dreemurr	He lives alone in the flower shop "Flower King" in Hometown, and is Kris and Asriel's father	Male	
Continue saved game	Continuer la partie sauvegardée	Continuer la partie sauvegardée	Continuer la partie sauvegardée	Asriel	He is mentioned by numerous characters in Hometown, but is away at university and does not appear in the game	Male	none
Credits	Crédits	Crédits	Crédits	Azzy	One mention	Male	none
You sure?	Vous êtes sûr?	Vous êtes sûr ?	- Vous sûr?	Berdly	He sits in the front row of Alphys's class, with Termyie to his right, Noelle Holiday to his left, and an empty seat behind him.	Male	
Click here to confirm	Cliquez ici pour confirmer	Cliquez ici pour confirmer	Cliquez ici pour confirmer	Blooper	An enemy that appears in the Forest.	it	
Castellano	Castellano	Castellano	Castellano	Blue Blocks	The Blue Blocks are Darkners the party can encounter in Forest.	it	
English	Anglais	Anglais	Anglais		An employee of ICE-E'S P'E'ZZA along with Burgerpants, The		
Français	Français	Français	Français				
Mapudungun	Mapudungun	Mapudungun	Mapudungun				
Created in 3 weeks in September and October 2020 for the [b]Hispanic Heritage Month Jan[b] by the [b]Custro Assets [b] team	Créé en 3 semaines en septembre et octobre 2020 pour le [b]Hispanic Heritage Month Jan [b] par l'équipe [b] Custro Assets [b]	Créé en 3 semaines en septembre et octobre 2020 pour le [b]Hispanic Heritage Month Jan[b] par l'équipe de [b]Custro Assets [b]	Créé en 3 semaines en septembre et octobre 2020 pour le [b]Hispanic Heritage Month Jan[b] par l'équipe [b] Custro Assets [b]				
Source (EN)	GOOGLE	DEEPL	SMARTCAT				
ARE YOU THERE?	ES-TU LÀ ?	ÊTES-VOUS LÀ ?	T'ES LÀ ?				
ARE WE CONNECTED?	SOMMES-NOUS CONNECTÉS ?	SOMMES-NOUS CONNECTÉS ?	SOMMES-NOUS CONNECTÉS ?				
...				
EXCELLENT.	EXCELLENT.	EXCELLENT.	EXCELLENT.				
TRULY EXCELLENT.	VRAIMENT EXCELLENT.	VRAIMENT EXCELLENT.	VRAIMENT EXCELLENT.				
NOW.	À PRÉSENT.	MAINTENANT.	MAINTENANT.				
WE MAY BEGIN.	NOUS POUVONS COMMENCER.	NOUS POUVONS COMMENCER.	NOUS POUVONS COMMENCER.				
FIRST.	PREMIER.	D'ABORD.	PREMIER.				
YOU MUST CREATE A VESSEL.	VOUS DEVEZ CRÉER UN NAVIRE.	VOUS DEVEZ CRÉER UN VAISSEAU.	VOUS DEVEZ CRÉER D'UN NAVIRE.				
SELECT THE HEAD THAT YOU PREFER.	SÉLECTIONNEZ LA TÊTE QUE VOUS PRÉFÉREZ.	CHOISISSEZ LA TÊTE QUE VOUS PRÉFÉREZ.	SÉLECTIONNEZ LA TÊTE QUE VOUS PRÉFÉREZ.				
SELECT THE TORSO THAT YOU PREFER.	SÉLECTIONNEZ LE TORSO QUE VOUS PRÉFÉREZ.	SÉLECTIONNEZ LE TORSO QUE VOUS PRÉFÉREZ.	SÉLECTIONNEZ LE TORSO QUE VOUS PRÉFÉREZ.				
SELECT THE LEGS THAT YOU PREFER.	SÉLECTIONNEZ LES JAMBES QUE VOUS PRÉFÉREZ.	SÉLECTIONNEZ LES JAMBES QUE VOUS PRÉFÉREZ.	SÉLECTIONNEZ LES JAMBES QUE VOUS PRÉFÉREZ.				
THIS IS YOUR BODY.	CECI EST VOTRE CORPS.	CECI EST VOTRE CORPS.	C'EST VOTRE CORPS.				
DO YOU ACCEPT IT?	L'ACCÉPTEZ-VOUS ?	L'ACCÉPTEZ-VOUS ?	AVEZ-VOUS L'ACCÉPTER ?				
EXCELLENT.	EXCELLENT.	EXCELLENT.	EXCELLENT.				
YOU HAVE CREATED A WONDERFUL FORM.	VOUS AVEZ CRÉÉ UNE FORMULE MERVEILLEUSE.	VOUS AVEZ CRÉÉ UNE FORME MERVEILLEUSE.	VOUS AVEZ CRÉÉ UNE MERVEILLEUSE FORME.				
NOW.	À PRÉSENT.	MAINTENANT.	MAINTENANT.				
LET US SHAPE ITS MIND AS YOUR OWN.	LAISSONS-NOUS FAIRE FABRIQUER SON ESPRIT COMME VÔTRE	FORMONS SON ESPRIT COMME LE VÔTRE.	LAISSEZ-NOUS FAÇONNER SON ESPRIT COMME LE VÔTRE.				
WHAT IS ITS FAVORITE FOOD?			QUELLE EST SA NOURRITURE				

Image 1. Excel files and character bibles.

Then, a colour code was created to manually annotate gender-related issues by highlighting the affected cells to provide a visual representation of the errors:

- **Blue:** use of the masculine for a reference to a female character.
- **Red:** use of the feminine for a reference to a masculine character.
- **Violet:** references to non-binary or non-gendered characters.
- **Green:** wrong use of the masculine for a reference to an object.
- **Orange:** wrong use of the feminine for a reference to an object.

Finally, we proofread, annotated, and reviewed individually at least once every document and then went together through them once more before removing all the lines that did not contain gender-related issues. Unfortunately, this included excluding segments where the translation was so poor that it was impossible to determine the type of error, a fairly common occurrence. Once the non-problematic lines were removed and a final check had been performed, we created a single document per colour for better visualisation and then proceeded to count all the errors per tool and type thrice (to make sure that the resulting number was consistent).

Image 2. Excel files per type of error.

III RESULTS

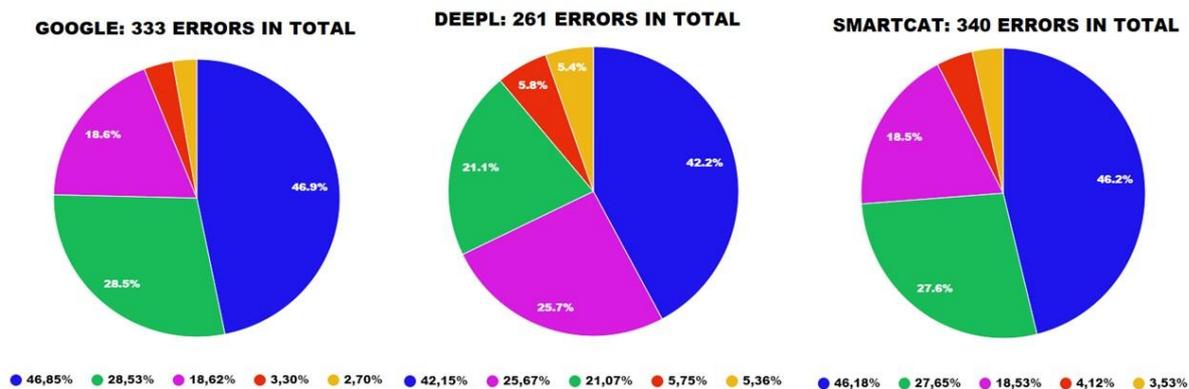
3.1 Colour code

The previous image (Image 2) already shows patterns regarding the type of error and tool: in the case of blue and green errors, we can easily observe a higher number of cells without issues in the column containing DeepL’s translation. Conversely, the same tool appears to err more than the rest in the other three categories. If we look at the total number of errors per tool all categories confounded (Table 1), Google made 333 gender-related errors, DeepL made 261, and SmartCat 340. These figures represent 3.03%, 2.38% and 3.1% of the 10980 segments, respectively. However, it is necessary to point out that SmartCat’s translation quality was remarkably low and, thus, we were not able to distinguish the type of mistake made on numerous occasions. Therefore, the percentage of Yandex’s errors should be much higher than the one listed above.

TOOL	ALL ERRORS	BLUE	GREEN	VIOLET	RED	ORANGE
GOOGLE	333 3,03%	156 46,85%	95 28,53%	62 18,62%	9 2,7%	11 3,3%
DEEPL	261 2,38%	110 42,15%	55 21,07%	67 25,67%	15 5,75%	14 5,36%
SMARTCAT	340 3,1%	157 46,18%	94 27,65%	63 18,53%	14 4,12%	12 3,53%
TOTAL OF SEGMENTS	10980					

Table 1. Percentage of errors per tool and category.

Figures 1 to 3 show the results arranged in pie charts and allow us to observe that all three tools had their highest number of errors when dealing with references to female characters (blue) even though DeepL performs better. DeepL shows a relatively higher tendency to favour feminine over masculine and has the highest percentages in the red and orange categories—and the lowest for blue and green. The differences in terms of percentages between the performance of Google and SmartCat are almost negligible although it would seem that SmartCat has a slightly higher tendency to overcompensate in the case of red and



orange.

Figures 1 to 3. Distribution of errors per tool.

3.2 Corpus-based analysis

Subsequently, we performed a corpus-based analysis to provide concrete examples of the issues encountered in each category. First, all the Excel documents were exported into a text file, one for each tool. Then, they were aligned using Aligner and converted into a single multilingual tmx file with all the segments containing errors. Finally, the file was uploaded into SketchEngine to create a parallel corpus that allowed us to analyse the translations and gender-related errors from a qualitative and a quantitative approach. In order to find relevant tokens for the analysis, a Wordlist search was carried out first to assert the frequency of appearance and use of all the words present in the corpus. After sorting the results, we decided to focus on “Luna”, “Ana”, “kid and kiddo” and finally on the pronouns “it”, “they” and “you”. These terms became the starting point of a more in-depth analysis with advanced parameters such as the parallel concordance function using the “simple” or the “lemma”

options. Furthermore, the images have been modified *post hoc* to include the colour code to frame the sentences as a visual aid to highlight the previously discussed categories.

3.2.1. Nouns

The token “Luna” (the protagonist of *The Devil’s Womb*) refers to a feminine character and thus, in French, the pronouns and adjectives must agree with it if it’s the subject of the sentence. However, even if Google Translate and SmartCat both detected that Luna was the subject, they didn’t apply the rule as there wasn’t any other gender reference in the original and, by default, transformed feminine into masculine. On the other hand, DeepL did not make mistakes in this particular case (Image 3).



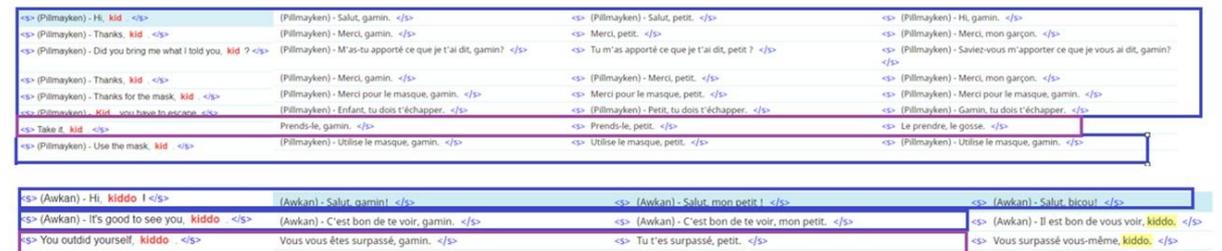
Image 3. Token “Luna”.

The token “Ana” (the protagonist of *The Faces of the Forest*) predictably shows the same results. In this case, Google Translate made three mistakes, SmartCat had it all wrong, and DeepL proved to be more efficient with only one mistake. In these sentences, as both the name of the protagonist and its coreference—the pronoun “I”—were present, the probability of gender-related errors should have been reduced. However, the results show the contrary and indicate that DeepL seems to be more performant than its competitors in terms of analysing the context and translating accordingly (Image 4).



Image 4. Token “Ana”.

By using an advanced lemma parallel concordance for “kid”, SketchEngine sorted all the results for the tokens “kid” and “kiddo”, which serve as coreferences to various characters in *DeltaRune* and “Ana” in *The Faces of the Forest* (said by Ana’s grandmother and a neighbour). These terms were problematic for every tool and were automatically translated using the masculine, being easily misled when processing English due to its gender-neutral form. Incidentally, SmartCat’s results for “kiddo” were too corrupted and cannot be compared to the other tools since the system did not even translate the term into French (Image 5). Additionally, the lines framed in violet, from *DeltaRune*, illustrate the intentional use of non-binary English. This decision, clear for a human translator, would have entailed using a neutralisation strategy or inclusive writing in French since “gamin”, “petit” and “le gosse” are



gendered.
Image 5. Tokens “Kid & Kiddo”.

3.2.2. Pronouns

In the following screenshot (Image 6), we can see that DeepL was the only tool that wrongly translated the pronoun “you”. In the first example, even though it handled “Luna” properly, it made a mistake while translating “Luna tried talking to you, but you were so nervous”. In this sentence, “you” is a coreference for the character Curly—a witch—and the adjective “nervous” should be feminine. In the final line, there is no explicit grammatical gender in the sentence although, if we analyse the context, we can observe that the pronoun “you” refers to an unnamed male demon. Nevertheless, as both Google Translate and SmartCat translated it by “vous” it is not possible to define whether or not they would have been misled too.

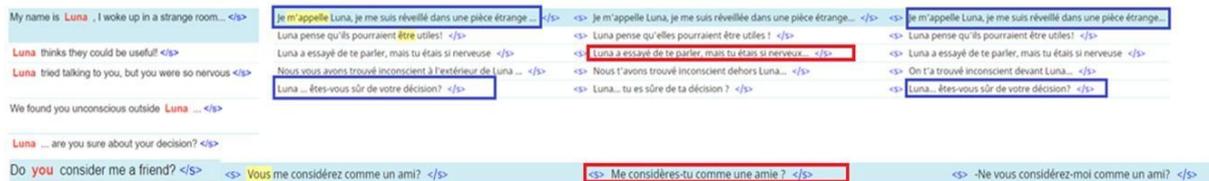


Image 6. Token “You”.

The pronoun “it” was consistently mistranslated by Google and SmartCat. The first sentence, “it got me a little worried”, shows that Google and DeepL analysed the sentence perfectly and produced idiomatic solutions: “Cela m’a un peu inquiété” or “ça m’a un peu inquiété” respectively. In the second sentence, “I hope you find it useful”, we can observe that the only system that took the context into account correctly was DeepL since the pronoun “it” refers to the company of a dog (without directly referring to the dog) and using the masculine article “le” is not appropriate. The third sentence, even though similar in construction to the first one, was correctly translated by both DeepL and SmartCat but not by Google. The seven next sentences, where “it” refers to a “key” and a “flower” (both feminine in French) show that Google and SmartCat translated the pronoun almost every time using the male article “il”. To sum up, Google and SmartCat made respectively 9 and 8 gender mistakes on 10 randomly chosen sentences while DeepL did not make any mistakes. In addition, we have to highlight that SmartCat translated “it cannot be told” into “Cela ne peut pas être indicible” which does not make sense, making the sentence potentially irrelevant to this analysis.



Image 7. Token “It”.

The translation of the pronoun “they” with little or no context is a recurrent issue in human translation as it can convey the notion of a group of people, objects, or a neutral form. In the following sentences (Image 8), “they” was the coreference for a group of female demons in *The Devil’s Womb*. Both Google and SmartCat applied the grammatical rule of male as a generic and only DeepL succeed to translate correctly all the sentences, showing once more that it is more efficient in analysing and memorising the context and the subject of the sentence. Finally, the last line shows that Google Translate and SmartCat failed to recognise that “they” referred to the Demon Kings and used the feminine.

<p>Luna thinks they could be useful </s></p> <p>dreamed the Nymphecoras and they surrendered to their deceptions.</p> <p>they died and their souls were absorbed by the </s></p>	<p>Luna pense qu'ils pourraient être utiles! </s></p> <p>révalent les Nymphécoras et ils se livraient à leurs déceptions. </s></p> <p>ils sont morts et leurs âmes ont été absorbées par le </s></p>	<p>Luna pense qu'elles pourraient être utiles ! </s></p> <p>révalent des Nymphécoras et qu'elles s'abandonnaient à leurs tromperies ; </s></p> <p>elles sont mortes et leurs âmes ont été absorbées par les </s></p>	<p>Luna pense qu'ils pourraient être utiles! </s></p> <p>révalent les Nymphécoras et ils se rendirent à leurs tromperies.</p> <p>ils sont morts et leurs âmes ont été absorbées par le </s></p>
<p></s> WHEN THEY PLEASE </s></p> <p></s> They can be used liked scissors but their </s></p> <p></s> I don't think they were randomly put in that position. </s></p> <p></s> cursed, therefore they were removed and destroyed. </s></p> <p></s> Pommes're scary, but all's they 'e don't is following the King. </s></p> <p></s> They ... </s></p> <p></s> They look really nice on you! </s></p>	<p></s> QUAND IL S'IL, VOUS PLAÎT </s></p> <p></s> Ils peuvent être utilisés comme des ciseaux mais leur </s></p> <p></s> Je ne pense pas qu'ils aient été placés au hasard dans cette position. </s></p> <p></s> maudits; c'est pourquoi ils ont été enlevés et détruits. </s></p> <p></s> Les Pommes sont effrayants, mais tout ce qu'ils font, c'est suivre le </s></p> <p></s> Ils... </s></p> <p></s> Ils vous vont vraiment bien! </s></p>	<p></s> QUAND ELLES LE VEULENT </s></p> <p></s> Elles peuvent être utilisées comme des ciseaux mais leur </s></p> <p></s> Je ne pense pas qu'elles aient été mises dans cette position par hasard. </s></p> <p></s> maudites, elles étaient donc enlevées et détruites. </s></p> <p></s> Les Pommes sont effrayants, mais ils ne font que suivre le Roi. </s></p> <p></s> Elles... </s></p> <p></s> Ils sont vraiment beaux sur toi ! </s></p>	<p></s> QUAND ILS S'IL, VOUS PLAÎT </s></p> <p></s> Ils peuvent être utilisés comme des ciseaux mais leur </s></p> <p></s> Je ne pense pas qu'ils aient été mis au hasard dans cette position. </s></p> <p></s> maudits; par conséquent, ils ont été enlevés et détruits. </s></p> <p></s> Les Pommes sont effrayants, mais ils ne font que suivre le Roi. </s></p> <p></s> Ils... </s></p> <p></s> Ils ont l'air vraiment joli sur vous! </s></p>
<p></s> linked by the fact that they took place in the same </s></p> <p></s> Where are they ? </s></p>	<p></s> liées par le fait qu'elles ont eu lieu dans le même </s></p> <p></s> Où sont-elles? </s></p>	<p></s> liés par le fait qu'ils ont eu lieu dans les mêmes </s></p> <p></s> Où sont-ils ? </s></p>	<p></s> liés par le fait qu'elles ont eu lieu dans le même </s></p> <p></s> Ou sont-ils? </s></p>

Image 8. Token “They”.

Conclusion

Creating this corpus allowed us to compare the use of gender in the source texts with the raw output of various MT systems to visualise the semantics, syntax, and linguistic differences between a neutral language such as English and a gendered language. By cross-referencing the information extracted from this corpus, we were able to determine the efficiency of each tool, how they analyse the context, their accuracy, and the quality of their translation. As such, DeepL has proved to avoid many gender biases while being more efficient than Google and SmartCat. In addition, even if Google uses NMT technology, our analysis shows that its architecture and models are closer to SmartCat’s. On the other hand, DeepL seems more accurate and fluent, surpassing Google thanks to its use of Deep Learning, as we have seen across the paper. Recent approaches in Neural Machine Translation show that integrating explicit gender inflexion tags into NMT systems reduces the number of gender biases, especially when translating sentences where the gender referent (pronoun) is identified [Sanders et al., 2020]. A similar method, incorporating a “speaker gender” tag into the training phase of MT systems, has also been developed to convey gender at a sentence level [Vanmassenhove et al. 2018]. Finally, another approach consists of adjusting the word embedding of NMT systems to reduce gender bias.

Future research will focus on asserting the tools’ bias in the context of gender-neutral choices by analysing the results for the token “Kris” (a non-binary character) in-depth as well as all the other errors in the violet category. Even though nowadays we can observe a higher demand for inclusivity and non-binary solutions, players are less familiar with inclusive writing, neutral phrasing or even gender-neutral pronouns, which could negatively affect the flow of the text and the player's immersion. The key factors of the use of inclusive language in video games are the client’s intent, their perception of diversity and inclusivity along with the localisers’ decisions and skills. The use of inclusive language techniques and constructions requires high awareness from developers and localisers, however, the audience’s reception must be taken into account and the topic should be researched further. Furthermore, training a specific MT system applying gender-neutral tagging to indicate neutral referents in the field of video game localisation remains an interesting avenue for future research. Finally, reproducing the experiment with the latest upgrades of the same tools to observe their evolution compared to the April 2021’s versions would allow us to examine the results of Deep Learning.

References

- Bernal-Merino, M. Á. *The localisation of Video Games*. PhD thesis, Imperial College (London), 2013.
- Crawford, K. The Trouble With Bias. Technical report, Neural Information Processing Systems Keynote at (NIPS’17)- 2017. https://www.youtube.com/watch?v=fMym_BKWOzk&ab_channel=TheArtificialIntelligenceChannel

DeepL Blog. <https://www.deepl.com/en/blog/how-does-deepl-work#:~:text=Like%20most%20translation%20systems%2C%20DeepL,methodology%2C%20mainly%20in%20four%20areas.> Accessed 2022-03-07

DeltaRune. <https://store.steampowered.com/app/1671210/DELTARUNE/> Accessed 2022-03-07

Dietz, T. L. An examination of violence and gender role portrayals in video games: Implications for gender socialization and aggressive behavior. *Sex Roles*. 1998;38(5-6):425-442.

Downs, E., & Smith, S. L. Keeping abreast of hypersexuality: A video game character content analysis. *Sex Roles*, 2010;62(11-12):721-733.

Google AI Blog. <http://ai.googleblog.com/2020/06/recent-advances-in-google-translate.html> Accessed 2022-03-07

Language industry platform (LIND). Report: 2020 Language Industry Survey. 2020. https://ec.europa.eu/info/departments/translation/language-industry-platform-lind_en Accessed 2022-03-07

Lynch, T.; Tompkins, J. E.; van Driel, I. & Fritz, N., Sexy, Strong, and Secondary: A Content Analysis of Female Characters in Video Games across 31 Years. *Journal of Communication*. 2016;66(4):564-84.

Rivas Ginel, M. I. La traducción automática en la localización de videojuegos. *Traduction automatique et usages sociaux des langues quelles conséquences pour la diversité linguistique*. 2021:167-178

Rivas Ginel, M. I. Video Game localisation tools: a user survey. *DE LA HIPÓTESIS A LA TESIS: TRADUCTOLOGÍA Y LINGÜÍSTICA APLICADA*. 2022: 295-324

Saunders, D.; Byrne, B. . Reducing Gender Bias in Neural Machine Translation as a Domain Adaptation Problem. 2020. arXiv:2004.04498 [cs]. <http://arxiv.org/abs/2004.04498>

The devil's womb. https://store.steampowered.com/app/913320/The_Devils_Womb/ Accessed 2022-03-07

The Faces of the Forest. <https://gamaverse.com/the-faces-of-the-forest-game/> Accessed 2022-03-07

Vanmassenhove E., Hardmeier, C., and Way, A. Getting gender right in neural machine translation. *In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pages 3003-3008, Brussels, Belgium, October-November. Association for Computational Linguistics

Yandex Blog. <https://yandex.com/company/technologies/translation/#:~:text=When%20a%20user%20enters%20a,rigth%20before%20the%20user's%20eyes.> Accessed 2022-03-07

Yang, S., Wang, Y., & Chu, X. (2020). A Survey of Deep Learning Techniques for Neural Machine Translation. *arXiv:2002.07526 [cs]*. <http://arxiv.org/abs/2002.07526>

Zhou, P.; Shi, W.; Zhao, J. and al (2019). Examining Gender Bias in Languages with Grammatical Gender. *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5276-5284, Hong Kong (China), January. Association for Computational Linguistics.