ISSN 1974-4382

GETTING GENDER RIGHT IN VIDEO GAMES, SIMPLIFYING FOR INCLUSIVITY

MARIA ISABEL RIVAS GINEL
SARAH THEROINE
UNIVERSITÉ DE BOURGOGNE FRANCHE-COMTÉ

isabelrivasginel@gmail.com; sarah.theroine@u-bourgogne.fr

Citation: Rivas Ginel, Maria Isabel and Sarah Theroine (2025) "Getting Gender Right in Video Games, Simplifying for Inclusivity", in Vecchiato, Sara, Sonia Gerolimich, Iris Jammernegg, Fabio Regattin and Deborah Saidero (eds.) *Semplicità e semplificazione in redazione, traduzione, terminologia e didattica*, *mediAzioni* 47: A143-A156, https://doi.org/10.6092/issn.1974-4382/22591, ISSN 1974-4382.

Abstract: The video game industry's meteoric rise since its birth in the 1970s has steadily increased the demand for professional localisers capable of dealing with "game crunch" (Brogan 2022) in what can only be described as "a double-blind process (no audiovisual context, no text linearity)" (Bernal-Merino 2013). Due to these working conditions, previous research in the field (Rivas Ginel 2021) has shown that the second most common error found by linguistic testers is mistranslations, a phenomenon that results from the above-mentioned constraints as well as the presence of variables. Variables, in the context of video games, are used to create the illusion of immersion by, for example, allowing the player to change the name (or gender) of the character. Therefore, translators are forced to find creative yet natural solutions to accommodate potential changes in the resulting sentence through simplification, neutralisation, and the use of controlled language. This article, the last of a series of papers analysing gender bias caused by machine translation in video games, aims to provide an in-depth analysis of the above-mentioned techniques used by translators to deal with gender when translating from English into French. We begin by providing a brief overview of the characteristics of video game localisation and their impact on the translators' work. Afterwards, we discuss some examples of mistranslations found in other games and the output generated by state-of-the-art baseline machine translation systems (Rivas Ginel and Theroine 2022). Subsequently, we will analyse neutralisation and simplification techniques extracted from a parallel corpus compiled from already translated video games that include non-binary characters. Finally, we present some of the results of an ongoing project that aims at conceiving a neural machine translation (NMT) tool for English into French video game localisation specialised in neutralisation techniques.

Keywords: video games; localisation; corpora; gender-neutral; inclusivity; neural machine translation (NMT).

1. Introduction

The video game industry has undergone a remarkable evolution since its inception in the 1970s, becoming one of the main players within the entertainment sector. Following the trend established by the software industry and media in general, video game developers took notice very early of the benefits of releasing multilingual products in terms of return on investment. Thus, since the early 90s, the main game-developing companies began translating their games into what was established by the emerging localisation industry as the main target languages: French, Italian, German, and Spanish (also known as FIGS). This practice, still maintained and recommended today by experts in the field (Chandler 2020), has had a clear impact on shaping the market and professional localisers' profiles (Table 1). However, recent studies and reports (Newzoo 2023) show the increasing importance of other target languages and markets.

Table 1. Top ten translators' native languages vs. main target languages (Rivas Ginel 2023: 153).

NATIVE LANGUAGE		MAIN TARGET LANGUAGE	
Spanish	158	Spanish	149
French	121	French	111
Italian	72	Italian	68
English	63	English	52
German	49	German	47
Portuguese	47	Portuguese	46
Russian	36	Russian	33
Turkish	25	Turkish	22
Catalan	24	Mandarin	14
Mandarin	22	Polish	14

This article focuses on providing an in-depth examination of the techniques employed by professional video game localisers to address gender-related considerations when translating from English into French, which stands as the second most represented target language in Table 1. To achieve this, we will commence with a brief overview of the framework surrounding this paper and the preceding related work conducted by the authors. Subsequently, we will explore the defining attributes of video game localisation and how they significantly impact the translator's task, particularly in terms of using inclusive or neutral language.

https://doi.org/10.6092/issn.1974-4382/22591

Following this, we will showcase the most representative neutralisation and simplification techniques extracted from a parallel corpus compiled from previously translated video games featuring non-binary characters. Finally, we present and discuss some examples of the neutralisation techniques currently being used to create a fourth – and final – parallel corpus. This dataset, which is the result of manually rewriting its content to completely remove gender references, is to be used as the culmination of the All-inGMT ongoing initiative. The ultimate goal of this venture is the development of a neural machine translation (NMT) suite of tools specifically tailored for English to French video game localisation, and capable of implementing neutralisation techniques and non-binary language to support translators in their decision-making processes.

2. Context: The All-inGMT project

The present paper represents a step in a series of articles devoted to analysing gender bias caused by the use of Machine Translation (MT) and Neural Machine translation (NMT) systems in video games with female and non-binary main characters; exploring non-binary language and neutralisation techniques used by professional translators; and conceiving NMT tools specially designed to deal with these issues. The project's catalyst was an initial study of Google Translate, DeepL, and SmartCat's performance when translating from English into Spanish (Rivas Ginel and Theroine 2023), and from English into French (Rivas Ginel and Theroine 2022c). Besides the authors' background, these languages were chosen due to the fact that they are at the top of the list presented in Table 1 and, contrary to English, require grammatical modifications related to gender when dealing with variables, which will be further explained in section 3.

Consequently, a corpus was assembled by leveraging pre-existing translations of video games with non-binary characters that were accessible online, either through open access or with explicit permission from the developers. Thus, the instances showcasing the techniques used in terms of transcreation and the use of non-binary language were extracted and analysed (Rivas Ginel e Theroine 2022b) and subsequently used to rewrite manually a series of parallel corpora (Theroine *et al.* 2022a) in the framework of the All-inGMT project (Rivas Ginel e Theroine 2022d; 2022a; Theroine *et al.* 2022b). The project, which stands for 'All-inclusive games machine translation', officially started at the beginning of 2022 and was funded throughout its first year by the ISITE BFC and the programme 'Investissement d'Avenir'. Still ongoing, it aims to develop a series of NMT systems to support video game localisers when dealing with non-binary characters by means of providing the appropriate neologisms or by offering an alternative solution that uses paraphrasing to neutralise gender-related references.

2.1. Current state of the project

The All-inGMT suite presently counts four distinct features: automatic annotation for English and French non-binary language, an English into French non-binary NMT tool, and an intralingual French into non-binary French

https://doi.org/10.6092/issn.1974-4382/22591

translator. Initially, the first two features were created as a means to improve the non-binary NMT systems' performance by automatically annotating artificially generated data added to our corpora in order to augment their size, which only consisted of 7,523 paired sentences – or a total of 135,700 (EN) and 152,276 (FR) words. The tools were obtained by using roBERTa's (Liu *et al.* 2019) token classification feature to detect and then automatically annotate new sentences – or inferences – and annotated samples of our non-binary corpus were used to retrain it. The F1 scores obtained, which result from the combination of the system's performance in terms of precision and recall, were 0.97 for French and 0.91 for English (Table 2). A close analysis of the output generated showed problems related to ambiguity in English – mostly in the case of the personal pronoun 'they' – and wrong tokenisations in French (Rivas Ginel and Theroine 2022d). In an attempt to enhance the English score, a list of terms to be excluded in the annotation process was drafted and implemented, albeit without yielding the desired improvement. Further examination of the degree of co-reference with NeuralCoref[1] showed that, despite the list, the neural network was unable to establish links between the non-binary instance and the corresponding pronoun in long sentences (Rivas Ginel and Theroine 2022d).

Table 2. RoBERTa's results for automatic annotation.

Model	**F1**	**Precision**	**Recall**
French	0.97190	0.9668	0.97698
English	0.91041	0.9084	0.912408

In the case of the NMT systems, we decided to use M2M100, a multilingual encoder-decoder (seq-to-seq) model primarily designed for translation tasks and initially pre-trained on a dataset covering thousands of language directions, using supervised data generated through large-scale mining (Tikhonov and Malykh 2022). Three distinct models were trained in this part of the study: one using non-annotated data, a second using automatically annotated data, and a third using a manually annotated corpus. In order to assess their performance, we employed three different automatic evaluation metrics, BLEU, METEOR and ChrF (Popović 2015). As depicted in Table 3, the results are low compared to other state-of-the-art NMT systems. The best scores were obtained using manually annotated data, a process that becomes less effective when dealing with larger datasets, which was the initial motivation for training RoBERTa. However, despite RoBERTa's commendable performance considering the size of the training set, it still made enough annotation errors in English to impact the performance and hinder the quality of the translation of the NMT tool trained with it.

[1] https://github.com/huggingface/neuralcoref.

https://doi.org/10.6092/issn.1974-4382/22591

Table 3. English into French non-binary systems' performance by use of annotations.

	M2M non-annotated	**M2M automatic annotation**	**M2M manual annotation**
BLEU	0.258	0.251	0.264
METOR	0.578	0.570	0.574
CHRF	0.539	0.527	0.542

Subsequently, we incorporated 6,352 paired sentences from our neutralised corpus and introduced an additional 6,500 back-translated sentences (without any annotations) to the baseline training set, reaching 20,375. Furthermore, we also decided to use SwitchOut to further increment the corpus size. SwitchOut is an algorithm used for data augmentation in NMT that creates content by "randomly replacing words in both the source sentence and the target sentence with other random words from their corresponding vocabularies" (Wang *et al.* 2018: 856). This is done automatically during each epoch and provides constant 'on the fly' data augmentation that takes less time than back translation (around twenty minutes to generate the sentences). These augmentations resulted in a slight improvement in scores: BLEU 0.269, METEOR 0.582, and ChrF 0.547. Future endeavours related to this specific tool will focus on enhancing its performance by exploring new possibilities facilitated by the emergence of large language models and their capability to generate substantial amounts of data.

Finally, the project has a fully functional tool (also based on a pre-trained M2M100 model) that adapts French into non-binary French created by aligning the French corpus and the non-binary French corpus. Hunalign[2], a sentence aligner, facilitated this process by taking into account sentence length and a dictionary that catalogued all entries with words in both French and their corresponding non-binary forms, encompassing variations like masculine, feminine, plural, and neutral forms. The dictionary was further supplemented with words from the corpus, specifying that identical entries were to be treated as a form of 'translation'. As Table 4 shows, the automatic evaluation metrics used to analyse its performance showcase considerably high scores for an NMT system. Given the size of our training dataset, these results stem from the closeness between French and non-binary French as the tool only has to adapt a reduced number of elements that are subjected to changes related to gender while many other remain the same such as verbs, adverbs, prepositions, etc. Future work will entail a human evaluation phase focused solely on compliance with the guidelines employed (Alpheratz 2018).

Table 4. Scores for the intralingual NMT model.

BLEU	0.9254
METOR	0.9642
CHRF	0.9715

[2] http://mokk.bme.hu/resources/hunalign/.

https://doi.org/10.6092/issn.1974-4382/22591

3. The video game industry

One of the critical dimensions characterising this specialised field is that it can be considered what Bernal-Merino (2013:119) aptly describes as a "double-blind process", marked by the absence of audiovisual context and the non-linearity of the text. Within this intricate operation, localisers are entrusted with the translation of specific units of text referred to as 'strings', which are typically provided in Excel files, severing the link with any type of visual context. Additionally, to streamline the development of multilingual products, game developers frequently organise the strings based on characters, areas, or other criteria, which causes the aforementioned lack of linearity. Moreover, localisers seldom have direct access to the actual game, as the localisation phase begins while the game is still being developed following a simultaneous shipment model – also known as sim-ship (Mangiron Hevia and O'Hagan 2013).

As previously mentioned, another distinctive aspect of video game localisation is the pervasive use of variables, which play a pivotal role in augmenting the player's immersion by allowing dynamic alterations and the possibility of customisation, giving them the chance to modify character names or choose their gender. Variables – short fragments of code interspersed within sentences – offer the possibility to create a single string that would fit every situation, as the system will automatically replace the code with the player's choice. However, if we go back to the data presented in Table 1, we can see that this technique is not as straightforward for the majority of the target languages in the top ten. In fact, the most common Spanish or French translation for the sentence 'Welcome to the game, /nplayer_name/n!' where '/nplayer_name/n' will be automatically replaced by a name created by the user, does not work without referring to the character's gender and requires reformulation.

Consequently, video game localisers face the challenge of devising creative yet natural solutions to harmonise potential alterations within the resulting sentence and would need to favour, in the case of the previous example, expressions such as 'c'est un plaisir' (Muñoz Sánchez 2017). This translates as a need to employ simplification, neutralisation, and controlled language to achieve optimal outcomes. Furthermore, the unique combination of the aforementioned working conditions and the presence of variables introduces formidable challenges, the predominant one consisting of accurately determining the gender of the character speaking or the intended addressee within the available context. Research in this domain (Rivas Ginel 2022) and professionals' testimonials (Techoueyres 2020), underscore the prevalence of mistranslations – particularly related to gender – as the second most common error identified by linguistic testers. Such errors, largely uncommon in traditional translation spheres, further emphasise the intricate nature of video game localisation.

Exploring the nuances of gender-related translations becomes increasingly vital in today's context. As non-binary characters become more prevalent in the media and society embraces inclusivity, a notable shift away from traditional binary gender concepts is evident. This transformative movement has also made its mark on the video game industry, prompting developers to integrate non-

https://doi.org/10.6092/issn.1974-4382/22591

binary characters into their game narratives. The translation of games such as Temtem, carried out by the language service provider Native Prime, gained notoriety due to the obstacles they encountered when dealing with French, German, Spanish and Brazilian Portuguese non-binary pronouns[3] due to the novelty of the topic. Nevertheless, "[w]hile it might be the case that there is good non-binary gender representation in some other games, especially games developed by queer indie game developers [...], this does not always appear in AAA games" (Heritage 2021: 227-228). Regardless of the size of the developing company and the presence of non-binary characters, translators end up more often than not basing their decisions on probabilities whenever they lack information or context about the gender, a strategy that may cause problems down the production pipeline:

> Personally, I worked on a project a couple of years ago where one of the main characters was a lady that wrote letters to the person she loved, and during six months we were wondering whether it was a man or a woman. So we interpreted it as a man since 99% of the time it is the case. However, two months before the release of the game, during the testing phase, the studio wrote back and told us that it was actually a non-binary character. (Techoueyres 2020, min. 27:11 – 28:47)

4. Translators' techniques for working around gender

Given the current context in terms of inclusivity, business practices in the domain of video game localisation, and the increasing use of MT systems, we decided to initiate the All-inGMT project after noticing the lack of tools dealing with this issue specifically and the mistakes generated by NMT state-of-the-art generic models. In fact, at the time of the study (2021) and using three games where the text was completely linear and there were no variables, we observed that GoogleTranslate and DeepL (and Yandex, a statistical machine translation system) made a significant number of errors with gender references in general when translating from English into French, whether it was in the case of characters or objects. Table 5, a reduced version of the original that solely focuses on the two NMT models, shows the number and percentages of errors by type made in the 10,980 segments. The results showed that the highest number of mistakes were related to the translation of pronouns such as "you" and "they" and that DeepL was slightly better at dealing with them. Presumably due to Google's and DeepL's training on extensive corpora, we also observed an inconsistent but definitely present use of a series of neutralisation techniques typically employed by professionals in the field of video game localisation.

[3] https://www.linkedin.com/posts/native-prime_thenativetouch-activity-6993145915230302208-lOw1/?trk=public_profile_like_view&originalSubdomain=es.

https://doi.org/10.6092/issn.1974-4382/22591

Table 5. Reduced version of the percentage of errors per tool and category (Rivas Ginel and Theroine 2021: 6).

	ALL	Female as male	Feminine as masculine	Non-binary	Male as female	Masculine as feminine
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%

Consequently, we chose to compile an initial parallel corpus in both English and French, in order to conduct a more in-depth analysis of the techniques employed by professional video game localisers in terms of non-binary language and neutralisation. Following this study and after realising the potential applications of our dataset, the methodology extracted was used to manually modify the original corpus twice to form the foundation for our NMT suite of tools. The source corpus draws from a diverse collection of video game scripts translated by professionals, totalling over 600,000 English words and 700,000 French words. These scripts were sourced from games either available online through open access or generously provided by independent developers such as Accidental Queens and Brace Yourself Games. Notable titles within this corpus include “A Normal Lost Phone”, “Alt-Frequencies”, “Crypt of The Necrodancer”, and “Industries of Titan”.

As expected, the majority of the instances that required attention when translating from English into French and provided the largest array of techniques were in the case of the pronouns ‘you’ and ‘they’. Given the nature of the corpus, ‘they’ could be used to refer to a non-binary character, a group of male characters, a group of female characters, a mix of males and females, or include all three genders. French, and most of the languages present in Table 1, employ a variety of pronouns to differentiate between all these options and thus, constitutes and challenge in terms of localisation. The most common solution found across our corpus was the systematic reformulation involving the use of ‘on’. This also was applied to deal with the pronoun ‘you’, which can either refer to the second-person singular or plural or any of the previous cases used as a politeness formula. In the case of the latter, another salient option was the systematic use of ‘vous’, which covers all four options (all uses of ‘you’ and the non-binary pronoun ‘they’), even though is less favoured in the video game sphere.

4.1. Techniques specifically used for non-binary characters

As an example, some of the most representative techniques used to specifically work around gender-related issues in the case of two non-binary characters, Kris from Deltarune and Charlie from Alt-Frequencies, involved: modifying verb tenses, replacing adjectives, omitting articles, restructuring, deleting full

https://doi.org/10.6092/issn.1974-4382/22591

sentences, or rephrasing. The original French translation for the sentence “Kris! Show up earlier next time”, “Yo, Kris! Sois plus matinal la prochaine fois !”, became “Kris! Tu devrais te lever plus tôt.”. Besides employing a more natural sentence, using “être” (“sois”) in French requires a grammatical agreement between the adjective and the subject of the sentence. Therefore, depending on the gender, the adjective will differ and become, for instance, “matinal” in the case of masculine and “matinale” for feminine. Thus, favouring the use of the conditional tense solves this issue by creating an impersonal structure, fostering a more generalised and gender-neutral statement. Secondly, in “Kris! Don’t act shocked. You know it’s true”, the French new translation reads “Kris, ne prends pas cet air surpris. Tu sais que c’est vrai”, where the adjective ‘shocked’ is replaced with the noun ‘air’, maintaining the intended meaning while avoiding gender-specific references. The third technique, the omission of part of the information, is illustrated in the phrase “Let’s go freak”, which yields the French translation “Dépêche-toi, minable”. The original French translation used the adjective “gros” before “minable” to highlight the harshness of the statement made by the speaker, emphasising the contempt towards the recipient. Yet once again, in French the adjective must match the character’s gender and this omission strategically sidesteps explicit gender references, contributing to a more inclusive tone (Rivas Ginel and Theroine 2022b).

In the case of examples referring to Charlie, rephrasing sentences to omit gender-specific terms is showcased in “Just to correct you, I am not an expert on the time loop”, translated to “Précision, je ne pense pas pouvoir m’attribuer une quelconque expertise en boucle temporelle”. This rendering ensures clarity without relying on gendered language. The restructuring of sentences is illustrated in “I’m not giving out advice on how to vote if that’s what you’re looking for”, resulting in the French version “Si vous me demandez de donner une consigne de vote, je m’y refuse. Mon domaine c’est la science, pas la politique”. This reorganisation facilitates an equivalent pragmatic effect in an inclusive yet natural manner. Finally, the option of deleting a sentence that does not add essential information to the dialogue remains a valid strategy when other alternatives become too unnatural or lengthy. This can be seen in the translated version of the sentence “Let me put it this way. Think of our report as a medication information leaflet”, which becomes “Il faut le voir comme une brochure d’information médicale. Une brochure provisoire, parce que…”.

5. Neutralisation and simplification techniques used in All-inGMT

The methods extracted from the original project are currently being employed in the second – and final – rewriting phase. In this instance, rather than eliminating strings lacking direct references to gender, as was done during the creation of the non-binary corpus, all segments are retained and rephrased to remove the use of gendered language. Consequently, the size of the corpus will not be drastically reduced, diminishing the need for data augmentation techniques that might compromise content quality. The original scripts already exhibited conscientious efforts towards simplification and neutralisation,

https://doi.org/10.6092/issn.1974-4382/22591

especially in response to variables. However, aligned with the approach taken for the non-binary tool, we have opted to artificially increase the number of potentially gender-neutral characters by systematically replacing their names with unisex alternatives. This endeavour was undertaken in both languages and the full neutralisation process is almost over in the case of English and halfway through in the case of French. Table 6 shows concrete examples extracted from this corpus in particular, to better illustrate the extent of the modifications currently being applied and the size of the task at hand.

Table 6. Comparison between the original and the neutralised parallel corpus.

GAME	ORIGINAL ENGLISH	NEUTRAL ENGLISH	ORIGINAL FRENCH	NEUTRAL FRENCH
Alt-Frequencies	(Bob) and hot old Bob.	(Joe) and hot old Joe.	(Bob) et **le bon vieux** sexy Bob.	(Joe) et Joe, **votre fossile** sexy.
Alt-Frequencies	(Ennis B.) - Looks like we got a proper question for our dear friend A. None.	(Elie B.) - Looks like we got a proper question for our dear friend A. None.	(Ennis B.) - On dirait qu'on a une question pour notre **ami** A. None.	(Elie B.) - On dirait qu'on a une question pour notre **camarade** A. None.
Alt-Frequencies	(Ennis B.) – The station managers are going to be so happy that we've just encouraged people to switch over to empty frequencies.	(Elie B.) - The station managers are going to be so happy that we've just encouraged people to switch over to empty frequencies.	(Ennis B.) - **Les directeurs des stations seront ravis** d'apprendre qu'on envoie **les auditeurs** sur des fréquences à l'abandon.	(Elie B.) - **Les personnes en charge des stations seront ravies** d'apprendre qu'on envoie **leur public** sur des fréquences à l'abandon.
Crypt of the NecroDancer	Cadence, I'm sorry! I was under his spell…	Clarke, I'm sorry! I was under a spell…	Cadence, **je suis désolé !** Je suis sous l'emprise d'un sort…	Clarke, **pardonne-moi !** Je suis sous l'emprise d'un sort…
Deltarune	Visitors, visitors! Now we can play, play!	Visitors, visitors! Now we can play, play!	**Des visiteurs, Des visiteurs !** C'est l'heure de jouer, de jouer !	**De la visite, de la visite !** C'est l'heure de jouer, de jouer !

https://doi.org/10.6092/issn.1974-4382/22591

In the table presented above, the original and new names of the characters are underlined, while segments requiring neutralisation are signalled in bold. Notably, upon initial observation, it becomes evident that the original English corpus required significantly fewer modifications compared to its French counterpart. This discrepancy is a key factor contributing to the near completion of the rewriting process for the English corpus. Therefore, with a primary focus on the French original and adapted columns, the examples above illustrate three main techniques employed during the rewriting process. The first technique, introduced in the initial row, involves restructuring the sentence by moving the character's name from its original end position to the beginning. These changes, among the most common ones, aim to enhance fluency and eliminate the potential use of the gendered male article 'le' by favouring the neutral possessive 'votre' while preserving naturalness.

The second technique involves substitution, a pivotal strategy for neutralisation that favours the use of epicene words. Examining the first row, the adjective 'old', initially translated as 'vieux', a gendered term in French, has been substituted with the noun 'fossile'. This substitution maintains the concept of ageing while sidestepping the use of adjectives, which can be sensitive to gender in the French language. Likewise, in the second row, a comparable scenario arises with the term 'friend/ami'. Given that the spelling of 'ami' varies based on gender ('ami' for males and 'amie' for females), the term 'camarade' (comrade) emerges as a neutral alternative despite political connotations in some contexts. Here, the only gendered element would be the article preceding the word, but this is only applicable in the case of 'the' or indefinite articles 'a/an', not in the case of 'our'. These examples demonstrate that such terms can be effectively employed when referring to individuals, encompassing all genders in certain contexts. This flexibility is contingent on the prior modification of pronouns or syntax to align with the desired neutralisation.

The third technique involves rephrasing specific segments of the sentence to meticulously eliminate any gender references. This technique has been applied in rows three, four, and five of the linguistic adaptation process, as detailed below. In row number three, the English phrase "The station managers are going to be so happy..." and its French counterpart "Les directeurs des stations seront ravis..." had to be modified to achieve a neutral French sentence, but no changes were necessary in the case of English. Indeed, the French term 'directeurs' is subject to gender agreements, which means it has a feminine equivalent, 'directrices', based on grammatical gender rules. Furthermore, in the same segment, it is crucial to consider that the adjective 'happy/ravi' is linked to the masculine subject 'directeur', posing a potential grammatical mistake if not addressed properly. To handle gender-related concerns, a pragmatic solution was implemented, substituting the plural subject 'directeurs' with the neutral term 'personnes' (in English, 'people'). This effectively averts gender references as the word is feminine regardless of the gender of the subjects, and the agreement automatically requires using the feminine form for the adjective as well. Finally, the sentence was rephrased even further by incorporating 'en charge de', as 'personnes' alone does not convey the nuanced idea of management and hierarchy introduced by 'directeurs'.

https://doi.org/10.6092/issn.1974-4382/22591

A parallel scenario unfolds in the fourth row, where the auxiliary used in the original French translation of 'I'm sorry' necessitates agreement in gender between the subject and the adjective once more. In this specific example and in pursuit of a neutral syntax, this segment had to be rephrased to remove 'désolé' (gender-dependent) from the equation; we chose 'pardonne-moi' among the different options for emphasis only. Incidentally, the use of 'I'm sorry' was very common throughout the entire corpus, particularly in the longest games, and we had to consider each case individually to adapt it to the context in question and convey the intended meaning. In the last row, we can observe a similar strategic application of rephrasing. The original sentence structure, influenced by the word 'visitors' and its French counterpart 'visiteurs', undergoes gender modifications based on the group's composition. The rewrite transitioned to a non-personal structure, facilitated by the use of 'de la visite'.

6. Conclusions and future work

The examples presented above illustrate the current process that is taking place by ensuring the application of rephrasing as a technique to ensure gender neutrality, incrementing the occurrences in our corpus. They also portray the complexity of a task that involves rewriting large amounts of text and deciding between all possible translations that would fit the project's purpose while considering register and idiomaticity. Once the process is over, and after further research of the new venues opened to us by the arrival of large language models and generative artificial intelligence — or genAI — we will create the final NMT system in our suite of tools and make them available to professional localisers and the public alike.

In summary, through this comprehensive exploration of neutralisation techniques in the field of video game localisation, we endeavour to provide insight into the intricate landscape surrounding the translation of gender-related variables and how professionals deal with the inclusion of non-binary characters in terms of linguistic choices. Additionally, our efforts to conceive a suite of neural machine translation tools trained to offer alternatives to localisers facing these types of challenges could facilitate their task, ultimately fostering more inclusive and accurate localisation practices.

REFERENCES

Alpheratz (2018) *Grammaire Du Français Inclusif*, Paris: Vent Solars Linguistique.

Bernal-Merino, Miguel Ángel (2013) *The Localisation of Video Games*, PhD thesis, London: Imperial College.

Chandler, Heather M. (2020) *The Game Production Toolbox*, Boca Raton: CRC Press Taylor & Francis Group.

Heritage, Frazer (2021) *Language, Gender and Video Games: Using Corpora to Analyse the Representation of Gender in Fantasy Video Games*, Cham: Palgrave Macmillan.

https://doi.org/10.6092/issn.1974-4382/22591

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer and Veselin Stoyanov (2019) "RoBERTa: A Robustly Optimized BERT Pretraining Approach", *arXiv preprint arXiv*:1907.11692.

Mangiron Hevia, Carme and M Minako O'Hagan (2013) *Game Localization Translating for the Global Digital Entertainment Industry*, Amsterdam-Philadelphia: John Benjamins Publishing Company.

Muñoz Sánchez, Pablo (2017) *Localización de Videojuegos*, Seville: Síntesis, S. A.

Newzoo (2023) *Global Games Market Report*, available online at https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2023-free-version.

Popović, Maja (2015) "ChrF: Character n-Gram F-Score for Automatic MT Evaluation", *10th Workshop on Statistical Machine Translation*, WMT 2015 at the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015 - Proceedings, 392–395.

Rivas Ginel, Maria Isabel (2022) "Video Game Localisation Tools a User Survey", in Miguel Ibáñez Rodríguez, Carmen Cuéllar Lázaro, Paola Masseau (eds.) *De la hipótesis a la tesis: Traductología y lingüística aplicada*, Granada: Editorial Comares, 295-324.

Rivas Ginel, Maria Isabel (2023) *The Ergonomics of CAT Tools for Video Game Localisation*, PhD Thesis in Linguistics, Université Bourgogne Franche-Comté, English, https://theses.hal.science/tel-04351436.

Rivas Ginel, Maria Isabel and Sarah Theroine (2021) "Machine Translation and Gender Biases in Video Game Localisation: A Corpus-Based Analysis", Presentation at the conference *Colloque interdisciplinaire « Vers une robotique du traduire ? »*, Strasbourg: France.

Rivas Ginel, Maria Isabel and Sarah Theroine (2022a) "All-Inclusive Games Machine Translation: All-InGMT", Presentation at the conference *Interculturality and Sexuality in Videogames*, Mons, Blegium.

Rivas Ginel, Maria Isabel and Sarah Theroine (2022b) "Gender in Video Games: Creativity for Inclusivity", *Traducerea-Act Creativ: Între Știință Și Artă* in honorem Ludmila ZBANŢ, profesor universitar, doctor habilitat cu ocazia aniversării a 65-a de la naștere, Chișinău, Republica Moldova. https://doi.org/10.5281/zenodo.6523928.

Rivas Ginel, Maria Isabel and Sarah Theroine (2022c) "Machine Translation and Gender Biases in Video Game Localisation: A Corpus-Based Analysis", *Journal of Data Mining and Digital Humanities*, 15 décembre 2022, *Vers une robotique du traduire*, https://doi.org/10.46298/jdmdh.9065.

Rivas Ginel, Maria Isabel and Sarah Theroine (2022d) "Neutralising for Equality All-Inclusive Games Machine Translation: The All-InGMT Project", in Sheila Castilho, Rocío Caro Quintana, Maria Stasimioti and Vilelmini Sosoni Sosoni (eds.) *Proceedings of the New Trends in Translation and Technology Conference - NeTTT 2022. 4-6 July 2022, Rhodes Island, Greece*, 125-133.

Rivas Ginel, Maria Isabel and Sarah Theroine (2023) "Jeux Vidéo, Localisation, Traduction Automatique et Problèmes de Genre", *L'Âge d'or* 16, en ligne, https://doi.org/10.4000/agedor.6931.

Techoueyres, P. (2020) "LocJam Masterclass." min. 27:11 – 28:47.

Theroine, Sarah, María Isabel Rivas Ginel, Eva Vanmassenhove, Éric Poirier, Loïc Barrault, Felix Hoberg, Izabella Thomas, Lucie Bernard, Oliver Czulo and Sebastian Vincent (2022a) “The Corpora Games: All-Inclusive Games Machine Translation”, Poster presentation at *The 32nd Meeting of Computational Linguistics in the Netherlands* (CLIN 32), Tilburg, The Netherlands.

Theroine, Sarah, María Isabel Rivas Ginel, Eva Vanmassenhove, Éric Poirier, Izabella Thomas, Felix Hoberg, Loïc Barrault, Sebastian Vincent, Lucie Bernard and Oliver Czulo (2022b) “Playing with Gender All-Inclusive Games Machine Translation: The All-InGMT Project”, Presentation at the conference *Translation in Transition 6*, Charles University, Prague.

Tikhonov, Pavel and Valentin Malykh (2022) “WikiMulti: A Corpus for Cross-Lingual Summarization”, in *Conference on Artificial Intelligence and Natural Language*, Cham: Springer Nature Switzerland, 60-69.

Wang, Xinyi, Hieu Pham, Zihang Dai and Graham Neubig (2018) “SwitchOut: An Efficient Data Augmentation Algorithm for Neural Machine Translation”, *arXiv preprint arXiv*:1808.07512.

https://doi.org/10.6092/issn.1974-4382/22591