A Snapshot into the Possibility of Video Game Machine Translation

Damien Hansen [1,2,3]
damien.hansen@uliege.be

Pierre-Yves Houlmont [1,2]
pyhoulmont@uliege.be

[1] University of Liège, CIRTI*, 4020 Liège, Belgium
[2] University of Liège, Liège Game Lab, 4000 Liège, Belgium
[3] Univ. Grenoble Alpes, CNRS, Grenoble INP**, LIG, 38000 Grenoble, France

Abstract

We present in this article what we believe to be one of the first attempts at video game machine translation. Our study shows that models trained only with limited in-domain data surpass publicly available systems by a significant margin, and a subsequent human evaluation reveals interesting findings in the final translation. The first part of the article introduces some of the challenges of video game translation, some of the existing literature, as well as the systems and data sets used in this experiment. The last sections discuss our analysis of the resulting translation and the potential benefits of such an automated system. One such finding highlights the model's ability to learn typical rules and patterns of video game translations from English into French. Our conclusions therefore indicate that the specific case of video game machine translation could prove very much useful given the encouraging results, the highly repetitive nature of the work, and the often poor working conditions that translators face in this field. As with other use cases of MT in cultural sectors, however, we believe this is heavily dependent on the proper implementation of the tool, which should be used interactively by human translators to stimulate creativity instead of raw post-editing for the sake of productivity.

1 Introduction

Since the apparition of recurrent neural networks with attention mechanisms (Bahdanau et al., 2014), neural machine translation (NMT) has improved to the point of becoming the default paradigm for this task. With new architectures such as the Transformer (Vaswani et al., 2017) and a similarly growing number of domain adaptation techniques (Chu and Wang, 2018), NMT has also started being used in increasingly more complex domains, and even tailored to the production of specific translators and companies.

The video game market similarly seems to have been ever growing in the last decades, becoming one of the fastest growing and highest grossing industries today within the cultural and entertainment sectors. The recent global health crisis further reinforced this trend and showed that games have an important role to play beyond just entertainment, as can be seen in a recent EU report on the cultural and creative sectors (CCS) in Europe (IDEA Consult et al., 2021): "Few winners can be found in the CCS during the COVID-19 global crisis. One of those, together with streaming platforms, is the video games sub-sector. With a turnover of EUR 21.6 billion and a 3% year-on-year growth from 2018, the gaming sub-sector has proven to be strong also in hard economic times."

---

*Centre Interdisciplinaire de Recherche en Traduction et en Interprétation.
**Institute of Engineering Univ. Grenoble Alpes.

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas
Orlando, USA, September 12-16, 2022. Volume 2: Users and Providers Track and Government Track

Page 257

As part of this success, it has now been expected for a while that video games be translated not just into the Western E-FIGS (English to French, Italian, German and Spanish), but into at least eight or ten languages in order to be considered a profitable and successful enterprise (Bernal-Merino, 2015). This creates, in turn, a huge demand for translation and a challenge in its own right, especially for smaller or newly formed studios.

It should furthermore be noted that, nowadays, video game localization is mainly considered from a market perspective. Indeed, the main purpose of this process is to make a given title available to a larger audience. To do so, several attempts have been made to produce game localization manifestos encouraging best practices (Chandler and Deming, 2012; Honeywood and Fung, 2012).

Those practices — and scientific studies in the field of video game localization in general — are said to be market-driven (O'Hagan, 2013). Even though video games represent an exceptional place of intercultural contact, cultural traces are quasi systematically neutralized in an attempt to ensure sales, to the point where they would be considered as localization errors were they to remain in the target version (Mandiberg, 2015), despite evidence from reception studies showing that current localization practices do not correspond to what players expect from a cultural point of view (Ellefsen and Bernal-Merino, 2018; Fernández Costales, 2016; Geurts, 2015).

On the other hand, the video game industry, as is the case with other digital productions[1], is characterized by a particularly high level of fan-made content (Barnabé, 2015; Hurel, in press), with undertakings varying from patches[2], mods[3], or even (re)creations that far outgrow the original games themselves. It is also very common for fans and volunteers to look into the original translations, sometimes correcting minor mistakes in these very large works, sometimes coming up with entirely new localization projects if the title was never translated or if the translation was not done by professionals and its quality deemed too poor (Díaz Montón, 2007; Muñoz Sánchez, 2009; Vazquez-Calvo et al., 2019).

Indeed, translation has traditionally been seen in the market as a trivial and disposable process that could be targeted to cut costs, leading up to the numerous examples of bad video game translations that careless uses of machine translation (MT) definitely did not help with (Mandelin and Kuchar, 2017). This brings several issues, not only for the players, but for the companies as well. As Bernal-Merino (2008) explains: "One of the most common complaints we can read about in internet forums is the lack of translation quality, and how sometimes they have to go back to the original version to find out what to do and how. In many cases, even though the game might have been translated and it is playable, fans fail to be impressed because the poor quality of the localisation defeats its own purpose: to thrill and engage the gamer. [...] Language and culture are ever-present elements in us and the things we do, players cannot help but notice continuous serious mistakes in the game, and it will erode their trust in some developer and publisher brands."

But what makes video game translation such an arduous task and how could machines play a role in it if it already proves challenging for humans? In addition to the common pitfalls of translation and creative works, translators in the field are typically presented with spreadsheets in which text segments are not ordered chronologically and do not offer any other kind of contextual information, concerning for instance the communication medium, the type of discourse,

---

[1] As the Web gave rise to the development of a "user-participatory culture" (Jenkins, 2006) and "user-generated content" (van Dijck, 2009), the translation sector similarly witnessed the emergence of "user" or "amateur" crowd-sourced translations (Lavault-Olléon, 2011; Gambier, 2016; Doherty, 2016).

[2] A patch, or fix, is a light modification made to a program in order to correct bugs, for example, to add a translation or to improve its usability.

[3] Mods affect the game more deeply and generally aim at customizing or expanding rather than fixing it. These modifications can affect its appearance, its story or even its gameplay (Barnabé, 2015).

the person speaking, etc. (Díaz Montón, 2007). What is more, a survey conducted with translators working in this sector found that only 30% of them had access to the game they were asked to translate (Theroine et al., 2021). The localization of video games, however, is dependent on more than just text, as it relies on a precise balancing of intersemiotic dynamics. Indeed, the different semiotic systems of a given game work together to provide a complete and coherent media experience, and may present "tensions" or a form of "interdependency" that can limit the spectrum of possibilities for the translator (Houlmont, 2022). This lack of audiovisual or contextual cues and text linearity is so prevalent that it has been theorized by Bernal-Merino (2015) as a double-blind process. In a sense, machines would therefore have as much (or rather as little) information at their disposal as humans do when translating video games. This observation thus makes us wonder what their performance would be in this scenario, but more importantly how useful they could be to translators.

2 State of the Art

As we mentioned, the advances in NMT architectures and domain adaptation techniques allowed for better performances in various domains. In the cultural sector, machine translation has mostly been used in the process of film subtitling, but a growing number of studies has focused on the challenges of its development, for instance, with literary texts (Sakamoto, 2020; Hansen, 2021).

Concerning video games, it is hard to find scientific sources that go further than acknowledging the existence of MT within this sector, as in O'Hagan and Mangiron (2013). This use of machine translation is further confirmed in various online posts and articles, but also in sessions dedicated to localization at the Game Developers Conference (Bartlet et al., 2014; IGDA, 2018). Some online translation service providers also put it forward, alongside with the traditional arguments of cost and time savings that typically accompany MT, and it would appear that some studios and translators are sometimes forced to turn to this technology in order to meet the strict and tight deadlines of the industry. It remains nevertheless agreed that MT would never be suitable for the translation of the more creative in-game contents, such as dialogues (O'Hagan and Mangiron, 2013). We can once again find confirmation for this observation in online interviews (Ruete, 2021), but also in the deployment and user feedback of open source tools like *RetroArch* (Libretro, 2019), that leverage Google's translation service to make retro games accessible to a larger public.

Machine translation, however, might benefit the video game localization industry, whether it is to cope with the heavy requirements of a simultaneous multilingual game delivery, to help fans make their content available to more people, or to assist professionals during this long-term and demanding task. Yet, we have found no evidence of MT tools being specifically adapted for the translation of video games, and the existing use of MT as well as the advances in NMT make it the perfect time to try and see how effective such a custom engine could be. To our knowledge, this is thus the first article to dive deeply into the possibility of video game machine translation, or, at the very least, that an MT system was specifically designed for such a task.

3 Methodology

Video game translation is an eminently difficult task that involves linguistic skills and cultural awareness from the two languages at hand, diegetic knowledge about the world in which the game takes place, as well as an understanding of how the game is to be played in real life and how the translation fits into its development. To have an idea of whether MT could be of help to professional translators, we built a first system tailored to the translation of video games for the English–French language pair.

This system was trained with the 2nd edition of the OpenNMT framework (Klein et al., 2017) on a custom-made corpus of video game translations. Our bespoke MT engine was then evaluated on two role-playing games (RPG), namely *The Elder Scrolls V: Skyrim* (Bethesda Game Studios, 2011) and *Fallout 4* (Bethesda Game Studios, 2015), taking place in a high fantasy and post-apocalyptic setting respectively.

3.1 Data set

For this task, we used a relatively small in-domain data set (of about 1 million sentences), that combines the official translations for 22 games, mainly from the fantasy and post-apocalyptic RPG genre.[4] The same corpus was used to train both models, but we extracted validation and test samples without replacement for each game being evaluated.

	Sentences	EN tokens	FR tokens
Training	951,373	16,942,551	18,008,101
Validation	4,785	70,576	74,189
Test	501	8,574	9,254

Table 1: Data sets for *The Elder Scrolls V: Skyrim* (Bethesda Game Studios, 2011).

	Sentences	EN tokens	FR tokens
Training	951,861	16,958,940	18,020,052
Validation	4,401	56,486	64,738
Test	397	6,275	6,754

Table 2: Data sets for *Fallout 4* (Bethesda Game Studios, 2015).

These titles were chosen not only for their wide popularity, but also because they are representative of the multiple difficulties professionals might have to face in this industry. One of them is the sheer volume of in-game content to be translated, which often results in multiple translators working on the same project (Díaz Montón, 2007). On top of this, both games make use of a very specific terminology that draws from their distinct fictional universe. This abundance of lexical creations, or *irrealia* (Loponen, 2009), is a challenge in its own right, especially if the work is collaborative in nature or if professionals are not familiar with it, but this aspect is all the more important as this terminology builds on previous titles.

The Elder Scrolls and the Fallout series have indeed distinguished themselves through their rich mythopoetic universe, which partly explains why fans have taken such an interest in creating content in and around the games. Proof of this is the breadth of the fictional literature found in *Skyrim* for instance. The number of these metadiegetic works and their size — several hundred pages for some — actually lead us to delete all books from the Elder Scrolls games in our corpus, as it would have needed an additional and separate alignment process, but they were the only material removed from the original files.

The acquisition method for this corpus was inspired by the work of hobbyist modders. This type of content is a useful way to foster player engagement and to extend the lifespan of a game, which is why developers and distributors facilitate the creation and sharing of content through the release of toolkits or dedicated interfaces such as the workshop section in *Steam* (Valve, 2003). Today, platforms such as the mod repository *Nexus Mods*[5] or the amateur French translation team *La Confrérie des Traducteurs*[6] serve as a testament to the dedication of some of the fans. Among these mods, a few pieces of software are built from scratch to help with the translation process.[7] They rely on custom translation memories (TMs) and allow the user to load translation files from existing games to help speed up and improve the quality of the translation.

---

[4] For a more detailed overview of the corpus, see appendix A.

[5] https://www.nexusmods.com/.

[6] https://www.confrerie-des-traducteurs.fr/.

[7] It might be important to keep in mind here that such practices are not at all intended to replace professional translators, but only to extend the reach of works that would never be translated otherwise.

And although most of these users are probably not even aware of this, the process is exactly similar to that of computer-assisted translation tools, which have long been used by professionals.[8]

In a similar fashion, we were able to retrieve the translation files for a number of games, convert them from their native format into spreadsheet files, and align each pair in a standard translation memory format though the *Heartsome TMX Editor 8.0* (Heartsome Technologies Ltd, 2014). The resulting bitexts were cleaned semi-manually with the same program and a few regex manipulations to discard duplicate, empty and untranslated segments, to normalize the extremely varied typographic conventions and to remove game-specific formatting — most tags and variables were still left as such. We then extracted and compiled the remaining 956,661 segments, many of which contain more than one sentence.

Since 1 million bilingual segments is scarce for NMT — 6 million is even considered "frugal" by today's standards (Blin, 2021) — we also built models which we fine-tuned on our gaming data set. The initial training was performed on common generic data sets for the same language pair: *Books*, *Europarl-v8*, *GlobalVoices-2018*, *News-Commentary-v16*, and *TED2020* (Tiedemann, 2012), for a total of 2,923,826 additional aligned sentences.[9]

3.2 Architecture

In each case, data sets were tokenized with Moses (Koehn et al., 2007) and further segmented with sentencepiece's unigram model (Kudo, 2018), using a vocabulary size of 32,000 tokens. We then trained two Transformer models for each scenario (fine-tuned and in-domain only) with OpenNMT's default parameters and the "base" architecture described in Vaswani et al. (2017): 6 encoder and decoder layers, 8 attention heads, a dimension of 512 and feed-forward layers of size 2048, with dropout 0.1. We stopped training models after convergence (200,000 steps) and translated all files with the default parameters of OpenNMT again, which uses a beam size of 5 at inference.

4 Results

For evaluation, we detokenized the output with the same Moses script and computed a BLEU score with sacreBLEU (Post, 2018). We first report in Table 3 the score obtained on *The Elder Scrolls V: Skyrim* and *Fallout 4*, two games for which we could compare the output with the official translation. These test sets contained mostly dialogues and, to a lesser extent, quest descriptions, action choices and instructions.

Our first comparison indicated that adding data did not necessarily increase performance. In the first case, the added 3 million segments resulted in a negligible increase in BLEU, whereas that score even decreased in the second case. For this reason, and since our aim was to evaluate systems trained on in-domain data, we discarded the tuned models. This decision was further motivated by our evaluation of the resulting translations, which showed that the in-

	In-domain	Fine-tuned
Skyrim	37.14	37.38
Fallout 4	31.18	30.52

Table 3: Score given by sacreBLEU for the in-domain only and fine-tuned models on both games.

domain models seemed to learn inherent rules of the domain which the fine-tuned models did not, such as the fact that gendered variables tend to disappear when translating into French:

---

[8] Apart from in-house teams however, the use of TMs remains relatively uncommon in the video game industry.

[9] An experiment was also carried out with the dozens of millions of sentences from the WMT 2014 translation task (Bojar et al., 2014), but the final score dropped systematically due to the quality and dissimilarity of these data sets.

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas
Orlando, USA, September 12-16, 2022. Volume 2: Users and Providers Track and Government Track

Page 261

SRC: Is it true what they say? There was a dragon held captive in Whiterun, and you... you released it? **By the gods, woman, why?**

HYP (fine-tuned): C'est vrai ce qu'ils disent ? Il y avait un dragon captif à Blancherive, et vous... vous l'avez libéré ? **Par les dieux, femme, pourquoi ?**

HYP (in-domain): C'est vrai ce qu'on raconte ? Il y avait un dragon emprisonné à Blancherive et vous... vous l'avez libéré ? **Par les dieux, pourquoi ?**

REF: C'est vrai, ce qu'on raconte ? Qu'il y avait un dragon captif à Blancherive et que vous l'avez délivré ? **Mais pourquoi, par les dieux ?**

In doing so, we could assess more accurately the performance of systems trained only on data from the video game domain, which are already very encouraging considering that we used such a small training data set. To illustrate these results more clearly, we provide in Table 4 and Table 5 a comparison with publicly available systems using three measures provided by sacreBLEU. For each, an arrow indicates if the improvement is reflected by a higher or a lower score.[10]

	BLEU ⇑	chrF2++ ⇑	TER ⇓
Google Translate	27.75	48.25	66.75
DeepL	29.27	50.04	61.26
Custom	37.14	55.80	53.32

Table 4: Scores given by sacreBLEU for the custom and publicly available systems on *Skyrim*.

	BLEU ⇑	chrF2++ ⇑	TER ⇓
Google Translate	26.05	45.35	72.39
DeepL	27.60	47.04	67.94
Custom	31.18	48.80	62.96

Table 5: Scores given by sacreBLEU for the custom and publicly available systems on *Fallout 4*.

These automatic metrics are heavily dependent on form and processing, but this evolution gives an idea of the improvement, which according to Toral and Way (2015) should achieve at the very least a BLEU score of 20 to be useful in a post-editing workflow. To better describe the results, however, this last section offers a more qualitative analysis.

5 Analysis

As noted by Marie et al. (2021), MT evaluation has become less comprehensive and reliable over the last years, but one way to overcome this pitfall is to support automatic evaluation with human analysis. For this reason, we present two sets of evaluations. The first conveys the broad trends observed in our analysis of the two games presented previously, which we cannot publish for copyright reasons. We will further illustrate these remarks with a second set offering concrete examples from the translation of a fan-made mod. We believe the best way to transparently convey and judge the result of an automatically translated text is to provide a full and continuous example of this translation, so we have made it available online and we report here the index of the quoted sentence where appropriate.[11]

---

[10] Metric signatures for sacreBLEU (publicly available systems tested on 23/12/2020):

    BLEU   #:1|c:mixed|e:no|tok:13a|s:exp|v:2.0.0

    chrF2++ #:1|c:mixed|e:yes|nc:6|nw:2|s:no|v:2.0.0

    TER   #:1|c:lc|t:tercom|nr:no|pn:yes|a:no|v:2.0.0.

[11] https://gitlab.uliege.be/dhansen/VGMT-article/.

We provide 323 segments, broken down into four illustrative excerpts. They are mainly made up of the dialogues that are often raised as criticism against MT and that were chosen because they formed mostly coherent pieces of conversation, with the exception of disrupting segments that are typical of video game files. The first excerpt is singular, as it contains segments from the 2002 game, which therefore also appear in our original training data. We included it nonetheless,

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas
Orlando, USA, September 12-16, 2022. Volume 2: Users and Providers Track and Government Track

Page 262

The mod, named *Beyond Skyrim*, comes from the fifth instalment of the Elder Scrolls series and has been in development for multiple years, with the aim of adding seven more provinces faithful to the game universe in addition to the only one featured in the original title. This includes a province featured in a former opus of the franchise, *The Elder Scrolls III: Morrowind* (Bethesda Softworks, 2002), on which we focused our attention. While there remains typical MT errors that require the intervention of a trained translator, the actual texts produced by our adapted engine for this mod made it also clear that such tools could be useful to the translation of games developed by studios and fan-made content.

5.1   General observations

As a note, we should clarify that these are not exhaustive analyses, but rather the main trends observed when assessing the resulting translations and subsequently using this material with students to initiate discussions about MT. We also provide a few examples, but we strongly invite readers to have their own look at the output. The main and somewhat expected observation is that although *Google Translate* and *DeepL* reach a respectable BLEU score and produce mostly understandable solutions, the concrete results seem hardly adapted to a video game translation, in particular when compared to an adapted system.

Indeed, both online systems have a hard time when it comes to **register**, for instance. Whereas our system was generally able to render the high register and medieval feel of *Skyrim*'s original French translation as well as the particularly colloquial, even crude register of *Fallout*, both were neutralized with the online tools (cf. in the provided translation to segments 7, 26). We noticed nonetheless that the training data from one game contaminated translations for the other, as crude language appeared on rare occasions in *Skyrim* and *Fallout*'s dialogues were sometimes too polite, or the French distinction between the informal/formal 2nd person singular/plural was not observed (97, 98, 115). This suggests that even with an in-domain data set, it could be useful to go beyond this simple video game adaptation and give different weights to the training data in order to further adapt the system on specific video games or franchises.

Another evident advantage of a custom system for which we do not even need to point to a given example is its ability to translate virtually all of the fictional content and terms from the games, be it names of people, institutions, places, spells, monsters, items, mythological concepts, expressions from constructed languages... This is all the more salient as terms such as "Dragonborn" (*Enfant de dragon*), "Greybeards" (*Grises-Barbes*) or "Skyrim" (*Bordeciel*) are at the very core of the story, and this **vocabulary** is always left untranslated by *Google Translate* or *DeepL*. A particularly representative example comes from *Fallout 4*, where "Dogmeat", the name of a dog companion, has been translated by *la viande de chien* in *DeepL* (literally "the canine flesh") instead of the meat brand for dogs (*Canigou*) that is used in the game. Closely related is the rules for the **capitalization** of words that are much more restrictive in French and observed by the personalized engine, whereas the online tools systematically copy the English case.

We have also noticed that our MT system has learned translation strategies used to anticipate **gender** variation. Indeed, players can often choose their gender in video games, but this is even more visible in *Skyrim*, where they can also choose between various fantasy races. While this is not a problem in English, many words referring to the player character change according to gender in French. As such, the common translation strategy is to neutralize the potential gender pitfall by deleting every direct mention of gender (as illustrated above), not using tenses that require a feminine or masculine form (cf. 28), omitting gendered words (218) or

---

as these intertextual references can be frequent and we will see that the system sometimes takes liberties with these. There follows three other excerpts for which the quality can be judged respectively as good, average or less adequate. For each of these, we also provide another machine translation suggested by the generic tool *DeepL* on 25/02/2022.

using generic terms/paraphrases in their stead (16). Our custom engine has applied this strategy systematically, interestingly showing that NMT can learn not only specific vocabulary but also translation strategies that can anticipate common sources of error in video game translations.

Lastly, the adapted MT engine offers surprising results as a whole with the translation of these **dialogues**. This if even true of khajiit and argonian languages, two races in the game speaking from a specific third-person point of view. There remains, however, notable difficulties, such as with characters speaking in a way that is depicted as drunk (250–252), with some — not all — idioms (208), colloquial speech (117, 212), or oral language that is ripe with pauses or incomplete sentences (128, 303) that the system tries to complete by itself.

A localization-related difficulty shared by all systems lies in the **ambiguities** between imperatives and infinitives, that take the same form in English but not in French. Paired with the lack of context, it is thus very hard for MT to differentiate between dialogue choices, orders given to the player or quest objectives and maintain a coherent translation. Hence, both French forms *-er* / *-ez* appear somewhat randomly. This is a challenge that human translators can face if they are not given any context, but that could be alleviated for both human and machine by using tags. A final obstacle faced by MT which makes human intervention imperative is its tendency to **translate literally** (61), especially when idioms are concerned (300), although we should perhaps remind that our model was trained with extremely limited data. More resources could improve this last point, as well as those that follow.

A last and interesting observation is that MT happens to correct human errors, which can be due to either time pressure or a lack of intersemiotic context, such as when the text describes a geographically situated object in the game and the translator cannot rely on any information. These scenarios reinforce our idea that MT could improve the quality of the final translation, by offering alternative solutions or encouraging translators to reflect on ambiguities.

5.2   Machine translation of fan-made content

Leaving behind questions of reference or comparison between generic and adapted models, we now want to delve into more a more language-specific analysis. For this, we focus exclusively on the translation of our mod and we continue to give references to specific examples.

The main observation, which is a known characteristic of NMT, is that while the system does not have many issues with form or fluency, there are many problems with meaning or adequacy. The most obvious and problematic are instances where the sentence is grammatically correct but has an **opposite meaning** (97, 112, 146, 263). In other cases, words are correctly translated but not in the given context, creating a **shift in meaning** (67, 114). On rarer occasions, words simply have a **wrong translation** (114, 132), or there is an **omission** (70) or **hallucination** (318) in the target text.

Our review has also highlighted other minor issues, for instance with **determiners**, especially if they are omitted in English (55, 75, 272). **Errors in the source text** can be problematic (289), even though this is not always the case (17). Finally, truecasing must be ignored seeing as case serves to distinguish most fantasy-related words, but terms in **all caps** usually confuse the machine (108, 244).

As a final note, we found that despite the presence of some segments in the training data for the first excerpt, the machine took some liberties with the translation. Some of these could arguably be said to work better than the original translation (36), but other dialogues show that the translation can simply vary while being equivalent to the reference (30), or introduce an error in the text (29). It is therefore necessary to remain mindful of these weaknesses. And while MT can bring more coherence between titles of the same franchise, these liberties can also be problematic, for instance with in-game books for which the translation should not be changed. With dialogues, however, these might lessen the feeling of *déjà vu* for players and rejuvenate the game experience through retranslation.

6   Discussion

As we have established in previous sections, we were able to achieve very encouraging results on a video game translation experiment by training an MT system on a surprisingly small in-domain data set. This was both expected and unexpected, in the sense that using relevant training data would logically boost performance and allow the system to assimilate specific terms or previously translated phrases, but we had not foreseen that this custom engine would learn abstract translation strategies that are particular to this domain, such as neutralizing variables when translating from a neutral into a gender-inflected language. We therefore think that, if implemented the right way, MT might be susceptible of helping with terminology and formal or standardized expressions, to offer relevant suggestions and maybe speed up the process, becoming a useful tool for either professional translators or amateurs working on fan translations.

This requires nevertheless that studios consider the translation process and resources as an integral part of the game development, which is not often the case and a source of problems even with human translation. Indeed, localization should be planned early in the development by rigorously following internationalization steps to accommodate translations into any language (Chandler and Deming, 2012). The use of MT further emphasizes this need, which could mean ensuring that texts are not hard-coded and mixed with lines of code, using tags that could be useful to human and machine to alleviate the ambiguities and lack of context, or refraining from imposing formal constraints to account for differences with languages other than English.

On the other hand, the video game industry benefits from a significant advantage, which is that all translations are already aligned and easily convertible into language resources. Yet, our study also highlighted the parallel between the decontextualized working conditions of the translator and the machine translation system. This lack of contextual information and resources is pointed as one of the main challenges translators have to face in the field (Bernal-Merino, 2008; Theroine et al., 2021), and one might wonder in these circumstances why translators so rarely have access to some kind of translation memory, or at least an official glossary. This alignment incidentally makes it easy to train MT systems tailored to the particular video game genre, or even specific works, which our study has shown to be able to effectively reproduce the expected terminology and offer relevant suggestions.

This might even come as a welcome change, if we consider that some professionals in the field (around 12%) already use MT in their workflow (Rivas Ginel and Theroine, 2021). With the convenience of a customized engine, we could even expect this number to rise, as such a tool help deliver faster translations, provide otherwise lacking information to translators and, most important of all, boost creativity. This last point might be the biggest asset of human translators in this domain, given that it is required to overcome the numerous technical and linguistic challenges that are typical of video game translation (Díaz Montón, 2007), without mentioning the very creative nature of the content itself. To this end, we plan to delve into a closer reading of these *ad hoc* MT solutions from the perspective of inter-semiotic dynamics (Houlmont, 2022).

In this respect, MT could free up time for creative thinking and particularly challenging segments or simply offer relevant suggestions, especially if it is used in combination with other tools and material such as TMs, reference corpora or termbases. However, we think this would happen if MT is used not as a first draft that may constrain the translation, but as a suggested translation akin to a TM match that would help correct points of interpretation in the source text, maintain stylistic and lexical cohesion, and even spot eventual mistakes in the target text.

On the amateur side, this technology could prove even more useful to fans for the translation of their user-generated content, thus expanding the success and replayability of their favourite games, or allowing games that were never intended to be translated to reach a wider audience. This, in turn, could promote cultural exchanges and sensitivity through a medium that otherwise tends to erase such influences against players' own expectations (Mandiberg, 2015).

All of these positive aspects of MT conversely depend on various ethical issues, the main one being the reason for its introduction. Relying on machine translation only for the sake of productivity, without human intervention, or blindly forcing its deployment even when it is not appropriate is sure to have a drastic impact on quality, creativity and the overall appeal of the game. Such habits evidently tie into much broader issues, some of which MT might even reinforce as in the case of amateur translations being used by companies seeking to cut costs. Finally, we should not forget that all of the resources used to train these systems depends on the quality of the material that is provided by human translators, whose rights seem to become less and less apparent as translation technologies progress (Bowker, 2021). We therefore hope that this exploratory article and its conclusions will hopefully start some discussions in the video game, or the cultural sector as a whole.

References

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2014). Neural Machine Translation by Jointly Learning to Align and Translate. In *3rd International Conference on Learning Representations: Conference Track Proceedings*, pages 1–15, San Diego, CA.

Barnabé, Fanny (2015). Les détournements de jeux video par les joueurs. *Reset*, 4:1–42.

Bartlet, Michaela, Michel Buch Andersen, Beatrice Compagnon, Mike Dillinger, Declan Groves, and Kirti Vashee (2014). What is the Place of Machine Translation in Today's Gaming Industry? *2014 Game Developers Conference*. San Francisco, CA.

Bernal-Merino, Miguel (2008). Creativity in the Translation of Video Games. *Quaderns de Filologia: Estudis Literaris*, 13:57–70.

Bernal-Merino, Miguel (2015). *Translation and Localisation in Video Games: Making Entertainment Software Global*. Routledge, New York, NY.

Blin, Raoul (2021). Neural machine translation, corpus and frugality. *arXiv* preprint. arXiv:2101.10650.

Bojar, Ondřej, Christian Buck, Christian Federmann, Barry Haddow, Philipp Koehn, Johannes Leveling, Christof Monz, Pavel Pecina, Matt Post, Herve Saint-Amand, Radu Soricut, Lucia Specia, and Aleš Tamchyna (2014). Findings of the 2014 workshop on statistical machine translation. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 12–58, Baltimore, MD. ACL.

Bowker, Lynne (2021). Translation technology and ethics. In Koskinen, Kaisa and Nike K. Pokorn, editors, *The Routledge Handbook of Translation and Ethics*, pages 187–221. Routledge, Abingdon, UK.

Chandler, Heather Maxwell and Stephanie O'Malley Deming (2012). *The Game Localization Handbook*. Jones & Bartlett Learning, Sudbury, MA, 2nd edition.

Chu, Chenhui and Rui Wang (2018). A Survey of Domain Adaptation for Neural Machine Translation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1304–1319, Santa Fe, NM. ACL.

Díaz Montón, Diana (2007). It's a Funny Game. *The Linguist*, 46(3):6–9.

Doherty, Stephen (2016). The Impact of Translation Technologies on the Process and Product of Translation. *International Journal of Communication*, 10:947–969.

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas
Orlando, USA, September 12-16, 2022. Volume 2: Users and Providers Track and Government Track

Page 266

Ellefsen, Ugo and Miguel Bernal-Merino (2018). Harnessing the Roar of the Crowd: A Quantitative Study of Language Preferences in Video Games of French Players of the Northern Hemisphere. *The Journal of Internationalization and Localization*, 5(1):21–48.

Fernández Costales, Alberto (2016). Analyzing Players' Perceptions on the Translation of Video Games: Assessing the Tension between the Local and the Global Concerning Language Use. In Esser, Andrea, Miguel Bernal-Merino, and Robert Smith, editors, *Media Across Borders Localizing TV, Film, and Video Games*, pages 182–201. Routledge, New York, NY.

Gambier, Yves (2016). Rapid and Radical Changes in Translation and Translation Studies. *International Journal of Communication*, 10:887–906.

Geurts, Francine (2015). What do you want to play? the desirability of video game translations from english into dutch according to dutch gamers and non-gamers. Master's thesis, Leiden University.

Hansen, Damien (2021). Les lettres et la machine : un état de l'art en traduction littéraire automatique. In *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles*, volume 2, pages 28–45, Lille, France. ATALA.

Honeywood, Richard and Jon Fung (2012). Best Practices for Game Localization. IGDA Localization SIG.

Houlmont, Pierre-Yves (2022). Traduire le jeu vidéo : un équilibrage des dynamiques intersémiotiques. *Sciences du jeu*, 17.

Hurel, Pierre-Yves (in press). L'amateurisme comme processus au cœur de la culture vidéoludique. In Krichane, Selim, Isaac Pante, and Yannick Rochat, editors, *Penser (avec) la culture vidéoludique : Discours, pratique, pédagogie*, pages 187–221. Presses universitaires de Liège, Liège, Belgium.

IDEA Consult, Goethe-Institut, Sylvia Amann, and Joost Heinsius (2021). Cultural and creative sectors in post-COVID-19 Europe: Crisis effects and policy recommendations. European Parliament.

IGDA (2018). Recap of our Roundtable at GDC. *Game Developer*, April 4.

Jenkins, Henry (2006). *Convergence Culture: Where Old and New Media Collide*. New York University Press, New York, NY.

Klein, Guillaume, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush (2017). OpenNMT: Open-Source Toolkit for Neural Machine Translation. In *Proceedings of ACL 2017, System Demonstrations*, pages 67–72, Vancouver, Canada. ACL.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, pages 177–180, Prague, Czech Republic. ACL.

Kudo, Taku (2018). Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 66–75, Melbourne, Australia. ACL.

Lavault-Olléon, Elisabeth (2011). L'ergonomie, nouveau paradigme pour la traductologie. *ILCEA*, 14:1–17.

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas
Orlando, USA, September 12-16, 2022. Volume 2: Users and Providers Track and Government Track

Page 267

Loponen, Mika (2009). Translating Irrealia: Creating a Semiotic Framework for the Translation of Fictional Cultures. *Chinese semiotic studies*, 2(1):165–175.

Mandelin, Clyde and Tony Kuchar (2017). *This be book bad translation, video games!* Fangamer, Tucson, AZ.

Mandiberg, Stephen (2015). Playing (with) the Trace: Localized Culture in *Phoenix Wright*. *Kinephanos*, 5(1):111–141.

Marie, Benjamin, Atsushi Fujita, and Raphael Rubino (2021). Scientific Credibility of Machine Translation Research: A Meta-Evaluation of 769 Papers. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 7297–7306. ACL.

Muñoz Sánchez, Pablo (2009). Video Game Localisation for Fans by Fans: The Case of Romhacking. *The Journal of Internationalization and Localization*, 1(1):168–185.

O'Hagan, Minako (2013). *Game Localization: Translating for the Global Digital Entertainment Industry*. John Benjamins, Amsterdam, Netherlands.

O'Hagan, Minako and Carmen Mangiron (2013). *Game Localization: Translating for the global digital entertainment industry*. John Benjamins, Amsterdam, Netherlands.

Post, Matt (2018). A Call for Clarity in Reporting BLEU Scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. ACL.

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

Ruete, Borja (2021). El Juego del Calamar usa posedición, ¿hay traducción automática en los videojuegos? *MeriStation*, November 24.

Sakamoto, Akiko (2020). The Value of Translation in the Era of Automation: An Examination of Threats. In Desjardins, Renée, Claire Larsonneur, and Philippe Lacour, editors, *When Translation Goes Digital: Case Studies and Critical Reflections*, pages 231–255. Palgrave Macmillan, Cham, Switzerland.

Theroine, Sarah, María Isabel Rivas Ginel, and Aurélie Perrin (2021). Au coeur de la terminologie du jeu vidéo. L'absence de ressources, frein majeur pour les traducteurs. *Traduire*, 244:27–40.

Tiedemann, Jörg (2012). Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pages 2214–2218, Instanbul, Turkey. ELRA.

Toral, Antonio and Andy Way (2015). Machine-Assisted Translation of Literary Text: A Case Study. *Translation Spaces*, 4(2):240–267.

van Dijck, José (2009). Users like you? Theorizing agency in user-generated content. *Media, Culture & Society*, 31(1):41–58.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin (2017). Attention is All you Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010, Long Beach, CA. Curran Associates.

Vazquez-Calvo, Boris, Leticia Tian Zhang, Mariona Pascual, and Daniel Cassany (2019). Fan translation of games, anime, and fanfiction. *Language, Learning and Technology*, 23(1):49–71.

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas
Orlando, USA, September 12-16, 2022. Volume 2: Users and Providers Track and Government Track

Page 268

Appendix A. Corpus Specifications.

	Game	Developer	Size
1.	*Baldur's Gate*	BioWare, 1998	19 K segments
2.	*Baldur's Gate II: Shadows of Amn*	BioWare, 2000	62 K segments
3.	*Darkest Dungeon*	Red Hook Studios, 2016	8 K segments
4.	*Divinity: Original Sin II*	Larian Studios, 2017	85 K segments
5.	*Fallout*	Black Isle Studios, 1997	26 K segments
6.	*Fallout 2*	Black Isles Studios, 1998	56 K segments
7.	*Fallout 3*	Bethesda Game Studios, 2008	49 K segments
8.	*Fallout: New Vegas*	Obsidian Entertainment, 2010	64 K segments
9.	*Fallout 4*	Bethesda Game Studios, 2015	126 K segments
10.	*Planescape: Torment*	Black Isle Studios, 1999	39 K segments
11.	*Pillars of Eternity*	Obsidian Entertainment, 2015	48 K segments
12.	*Star Wars: Battlefront II*	Pandemic Studios, 2005	4 K segments
13.	*The Elder Scolls III: Morrowind*	Bethesda Softworks, 2002	37 K segments
14.	*The Elder Scrolls IV: Oblivion*	Bethesda Game Studios, 2006	40 K segments
15.	*The Elder Scrolls V: Skyrim*	Bethesda Game Studios, 2011	70 K segments
16.	*The Witcher 2: Assassins of Kings*	CD Projekt, 2011	32 K segments
17.	*The Witcher 3: Wild Hunt*	CD Projekt RED, 2015	78 K segments
18.	*Torment: Tides of Numenéra*	inXile Entertainment, 2017	49 K segments
19.	*Ultima VII: The Black Gate*	Origin Systems, 1992	12 K segments
20.	*Ultima VIII: Pagan*	Origin Systems, 1994	6 K segments
21.	*Ultima IX: Ascension*	Origin Systems, 1999	9 K segments
22.	*Wasteland 2*	inXile Entertainment, 2014	37 K segments
Total			956 K segments

Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas
Orlando, USA, September 12-16, 2022. Volume 2: Users and Providers Track and Government Track
Page 269