

# A Snapshot into the Limits and Risks of Video Game Machine Translation

Pierre-Yves Houlmont & Damien Hansen

Premises

# Growing number of voices in creative sectors

3



العربية DEUTSCH FRANÇAIS ΕΛΛΗΝΙΚΑ ENGLISH  
ESPAÑOL EUSKERA עברית ITALIANO POLSKI  
PORTUGUÊS ROMÂNĂ РУССКИЙ ЯЗЫК SLOVENŠČINA  
SRPSKI SVENSKA TÜRKÇE

## Littérature, cinéma, presse, jeux vidéo : non à des traductions sans âme

Plusieurs modèles génératifs de langage et d'image ont fait récemment leur apparition dans l'espace public et privé ; ils se développent à une vitesse fulgurante, rendus accessibles à quiconque pour la moindre tâche rédactionnelle et créative. Ces modèles façonnent un monde où la création se passe peu à peu de l'humain, et précipitent l'automatisation de nombreux métiers intellectuels, autrefois réputés inaccessibles à toute mécanisation.

Comme tout le secteur de la culture et de la création, la traduction littéraire et audiovisuelle est touchée de plein fouet. Plusieurs organisations professionnelles ont publié des tribunes très claires pour dénoncer l'impact de ces technologies sur notre travail, sur nos œuvres et sur nos vies : l'AVTE (Audiovisual Translators Europe) en 2021, le STAA (Syndicat des Travailleur.euses Artistes-Auteur.ices) en 2022, ATLAS



العربية DEUTSCH FRANÇAIS ΕΛΛΗΝΙΚΑ ENGLISH  
ESPAÑOL EUSKERA עברית ITALIANO POLSKI  
PORTUGUÊS ROMÂNĂ РУССКИЙ ЯЗЫК SLOVENŠČINA  
SRPSKI SVENSKA TÜRKÇE

## Literature, Film, Press, Video Games: Say No to Soulless Translations

Several generative models of language and images have recently appeared in the public and private domains; they are developing at breakneck speed, accessible to all for any task which involves writing and creating. These models are shaping a world where, little by little, creation can do without human beings, thereby hastening the automation of many creative and intellectual professions formerly deemed inaccessible to mechanization.

As with the entire cultural and creative sector, literary and audio-visual translation is being hit hard by this technological development. Several professional organizations have published very clear manifestos to denounce the impact of these technologies on our work, our creations, and our lives: AVTE (Audiovisual Translators Europe) in 2021, STAA (Union of Artist-Author Workers) in 2022, ATLAS (Association



العربية DEUTSCH FRANÇAIS ΕΛΛΗΝΙΚΑ ENGLISH  
ESPAÑOL EUSKERA עברית ITALIANO POLSKI  
PORTUGUÊS ROMÂNĂ РУССКИЙ ЯЗЫК SLOVENŠČINA  
SRPSKI SVENSKA TÜRKÇE

## Literatura, cine, prensa, videojuegos : NO a traducciones sin alma

Recientemente, hicieron su aparición en el espacio público y privado numerosos modelos generativos de lenguaje e imagen; se desarrollan a una velocidad meteórica y a los que cualquier persona accede libremente gracias a una mínima tarea de redacción y creación. Estos modelos construyen un mundo donde la creación depende cada vez menos del ser humano y aceleran la automatización, antes considerada imposible, de numerosos oficios intelectuales.

Como todo el sector de la cultura y de la creación, la traducción literaria y audiovisual se ve afectada directamente. Numerosas organizaciones profesionales publicaron artículos de opinión que denuncian claramente el impacto de estas tecnologías en nuestro trabajo, en nuestras obras y en nuestras vidas: la AVTE (Audiovisual Translators Europe) en 2021, el STAA (Syndicat des Travailleurs.euses Artistes-Autrices) en 2022, ATLAS (Association pour la promotion de la traduction

En chair et en os. “Littérature, cinéma, presse, jeux vidéo : non à des traductions sans âme.” *En chair et en os*, Sept. 2023.



# Growing interest within the industry

4

**The future is here**  
**Neural machine translation  
for games**



Sim City: ea.com/games



**Cristina Anselmi**  
Cristina Anselmi is a machine translation specialist at Electronic Arts Inc. based in Cologne, Germany. She's been in the video game localization industry since 2012, managing multilingual projects from small mobile games to AAA titles.



**Inés Rubio**  
Inés Rubio is a localization veteran with 15 years of gaming industry background, with further experience in eCommerce and marketing. Her interest in process improvement drove her to research potential MT applications to game localization.

## ◇ EA as an “early adopter” (2019):

- High expectations of gamers
- Challenges (terminology, tags and creativity)
- But willing to implement MT (speed, cost and speed & cost)

## ◇ Other custom services:

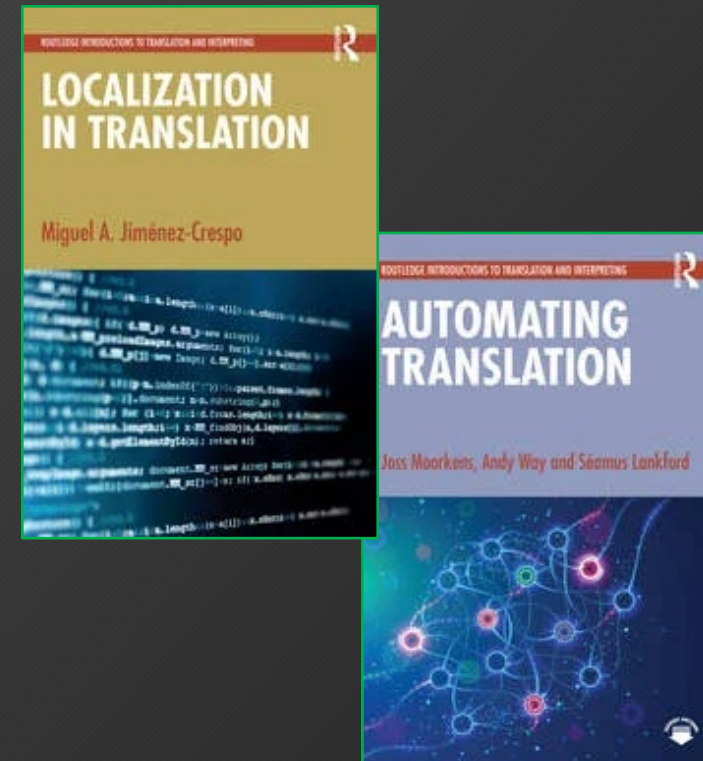
- Language AI at Keywords
- Alpha Intelligence at Alpha Games

# What do translators have to say about it?

- ◇ More than 35% use MT, but...
- ◇ Top “inconvenient” feature:
  - ↳ growing “imposed usage”
  - ↳ including proprietary software (“but they still suck”)
  - ↳ “EA’s piece of shit software”
- ◇ All the more reason to study this!

# Growing interest in the academic world

- ❖ Chapter on “**Perspectives in localization: From MT to AI**” in *Localization in Translation* (2024)
- ❖ Chapter on “**Machine Translation in Multimedia Translation and Localisation**” in *Automating Translation* (2024)
- ❖ Two ongoing thesis (presented here!)\*



\* Brenner, Judith. “[The MTxGames Project: Creative Video Games and Machine Translation - Different Post-Editing Methods in the Translation Process.](#)” *Proceedings of the 25th Annual Conference of the European Association for Machine Translation*, edited by Carolina Scarton et al., vol. 2, EAMT, 2024, pp. 47-48.

\* Copet, Simon. Preparation of a doctoral project on MT adapted to video games.



# Aims and previous work

7

## A Snapshot into the Possibility of Video Game Machine Translation

Damien Hansen<sup>1,2,3</sup> danien.hansen@uliege.be  
Pierre-Yves Houlmont<sup>1,2</sup> py.houlmont@uliege.be  
<sup>1</sup> University of Liège, CIRTP, 4020 Liège, Belgium  
<sup>2</sup> University of Liège, Liège Game Lab, 4000 Liège, Belgium  
<sup>3</sup> Univ. Grenoble Alpes, CNRS, Grenoble INP<sup>\*\*\*</sup>, 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 Wong, 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 growing 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 (IEA Consult et al., 2021): “Few winners can be found in the CCS during the COVID-19 global crisis. One of these, 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.”

<sup>\*\*\*</sup>Centre Interdisciplinaire de Recherche en Traduction et en Interprétation.  
<sup>\*\*\*</sup>Institute of Engineering Univ. Grenoble Alpes.

## ◆ Previous contribution (Hansen & Houlmont 2022):

- assessing the technical feasibility of the process
- automatic evaluation

## ◆ Aims of this presentation:

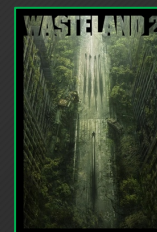
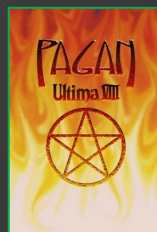
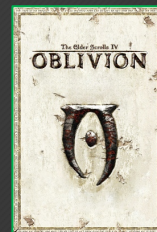
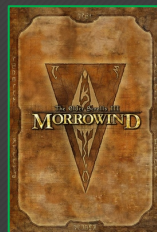
- more detailed evaluation
- emphasis on current context and risks

System



# Training corpus

9



I heard you recovered the staff. I thank the Nine for their gifts of sight, and I thank you for...  
As I told Kaye, I believe the Silver Staff of Shaming, and the remains of Linus Iulus, will be found in  
the shadow of death which has fallen over Linus Iulus. I hope you retrieve the staff. And I hope you find  
I'm glad you recovered the staff, and hope you can make good use of it.

# Pre-processing

10

You are in the Balmora Guild of Fighters now. I am the Guild Steward.  
Show some respect! You are in the most stalwart guild of warriors.  
%PCRank, what can be said. We are the few, the proud, the...  
Yes, I might have one I could sell you. I have been trying...  
You already brought me a copy of Chronicles of Nchuleft.  
Good. I will not forget this, %PCName. As you have brought me all three books. P  
Good. I will not forget this, %PCName.  
I do not want to play games with you, %PCName. Bring me a copy of Chronicles  
I know it may be difficult to find this book. You could try the Dwemer ruin  
Have you found a copy of Chronicles of Nchuleft?  
"Chronicles of Nchuleft" is very rare. There are probably still copies  
Are you still a member of the guild? Once Trebonius gives the aut  
Were you expelled? Or did you clear that up already? You ought  
Chronicles of Nchuleft has been invaluable in my research  
I'm impressed, %PCRank. I hope you are always this eff  
I thought we had an agreement, %PCName. Apparently  
What? You already have a copy of Chronicles  
You actually found a copy? Excellent, now  
I simply \*must\* have this book for m  
Excellent. Here's 250 septims to cover your expenses  
I guess I'll have to get Bribanne Erien to do it.  
It is a rare Dwarven tome. None of the Mages Guild branches have a c  
Yes, Jobasha knows the Chronicles of Nchuleft. Perhaps Jobasha has  
%Name owns no copies of Chronicles of Nchuleft. Speak with Joba  
Try Jobasha's Rare Books in the Market Canton.  
Chronicles of Nchuleft? It is a very rare book, but  
Sixth House references  
Here's a list of cheap books; all of them will tell  
Dorisa Darvel over in the Commercial District. She'll have at least some c  
Orcs at a Daedric ruin  
You have prevented the Barbarian Burub gra-Bamog from becoming a thre  
I have no reason to hate Orcs. In fact, I admire King Gortwog f  
theft, and banditry. If Burub is gathering forces in Ash  
I'm glad you got rid of those Orcs, %PCName. You ha

◇ Re-alignment: *LogiTerm* (Terminotix 2000)

◇ Tokenization: *Moses* (Koehn *et al.* 2007)

E.g. You are Dovahkiin-Dragonborn!

→ You are Dovahkiin - Dragonborn !

◇ Segmentation: *SentencePiece* (Kudo 2018)

E.g. You are Dovahkiin - Dragonborn !

→ \_You \_are \_Dovah kiin \_- \_Dragon born \_ !

```
# Training parameters:
batch_type: "tokens"
batch_size: 4096
valid_batch_size: 16
```

# Training

11

```
queue_size: 10000
bucket_size: 32768

# Optimization
model_dtype: "fp32"
optim: "adam"
learning_rate: 2
warmup_steps: 8000
decay_method: "noam"
average_decay: 0.0005
adam_beta2: 0.998
max_grad_norm: 0
label_smoothing: 0.1
param_init: 0
param_init_glorot: true
normalization: "tokens"
```

```
# Model
encoder_type: transformer
decoder_type: transformer
enc_layers: 6
dec_layers: 6
heads: 8
rnn_size: 512
word_vec_size: 512
transformer_ff: 2048
dropout_steps: [0]
dropout: [0.1]
attention_dropout: [0.1]
position_encoding: true
```

- ◇ *OpenNMT-py* toolkit (Klein *et al.* 2017)
- ◇ *Transformer* model (Vaswani *et al.* 2017)
- ◇ Training corpus (Hansen & Houlmont 2022)
  - 1.4 M segments (+ 500 K segments)
  - Adding more generic data proved detrimental



Evaluation



# Test case

13



- ◆ Project originally inspired by practices in the modding community (corpus building, fan translations, MT...)
- ◆ Makes sense given the training corpus
- ◆ Ideally, training and testing systems on different genres

# Error annotation

14

	Content				Meaning			Vocabulary		Inference		TOTAL
Adequacy	Omiss.	Addit.	Un.-tr.	Ov.-tr.	Oppos.	Nons.	Shift	Lex. c.	Non-tr.	Stutt.	Hallu.	
#	10	2	11	0	6	10	14	40	3	0	4	100

	Coherence			Agreement		Conjugation		Syntax		Structure		TOTAL
Fluency	Refer.	Relat.	Logic.	Gender	Numb.	Tense	Pers.	Funct.	Punct.	Guid.	Format.	
#	5	1	10	1	2	8	3	13	4	1	2	50

	Style			Terminology			Cultural ref.		Adapt			TOTAL
Localization	Coher.	Akw.	Regist.	Coher.	Mech.	Capit.	Realia	Irreal.	Creati.	Gender	Idiol.	
#	1	47	6	4	0	3	0	5	5	12	2	85

Hansen, Damien and Emmanuelle Esperança-Rodier. “[Human-Adapted MT for Literary Texts: Reality or Fantasy?](#)” *Proceedings of the New Trends in Translation and Technology Conference*, edited by Sheila Castilho et al., Incoma Ltd, 2023, pp. 178-190.

- ◆ Selection of a random range of segments make it hard to evaluate all aspects (in this case, only dialogues: no UI, no quests, no books, no journal entries...)
- ◆ Depending on the game, some criterions may be less or not at all represented (e.g. realia)
- ◆ Some aspects are very hard (if not impossible) to evaluate out-of-game, with text only (e.g. relational cohesion, gender of characters, guiding mechanisms...)

- ◇ In the ‘less problematic’ cases:
  - ↳ negative effect on player enjoyment, the overall appeal of the game and brand image
  
- ◇ In most cases:
  - ↳ negative effect on guiding mechanisms and contingency\*

---

\* Houlmont, Pierre-Yves. *La localisation vidéoludique au prisme de la contingence : pour un changement de paradigme*. Doctoral thesis, University of Liège, in press.



◇ Gender avoidance strategies:

↳ Thank you, my friend. ↳ Merci.

◇ Found gems and arguably very good translations (deamer poet)

↳ Thou wilt come to me there, my love, our senses tingling, through the inky ocean delight, our voices mingling. ↳ Tu viendras me voir là-bas, mon amour, nos sens sont en éveil, et nos voix s'entrechoquent dans l'océan noir des délices.

◇ 'Errors' that highlight a difficulty or suggest alternative translations:

↳ These things can hinder, but [...] ↳ Ces choses peuvent gêner, mais [...]

# Discussion

# Foreseeable risks

19

## ◇ Possible advantages:

- Errors ↗ (Ó Murchú 2019)
- Cognitive load ↗ (Larsonneur 2023)
- Creativity ↗ (Hansen & Pickford 2022)
- Appropriating the text (Hansen & Pickford 2022)
- More enjoyable work (Taivalkoski-Shilov 2019)

## ◇ Possible disadvantages:

- Errors ↗ (Ó Murchú 2019)
- Cognitive load ↗ (Taivalkoski-Shilov 2019)
- Creativity ↘ (Guerberof-Arenas & Toral 2022)
- Less apparent voice (Kenny & Winters 2020)
- Less enjoyable work (Moorkens *et al.* 2018)

# Foreseeable risks

20

## Traduction automatique et post-édition

Enquête ATLF menée du 20 novembre au 13 décembre 2022



- ◆ Entirely dependant on how and why MT is implemented
- ◆ Tendency to replace and cut costs rather than aid and improve, to the detriment of quality and creativity, client satisfaction...
- ◆ Interesting survey in a similar (literary) field, especially reg.:
  - the visibility of translators
  - the lowering of remuneration
  - the devaluation of the work

ATLF. “[Traduction automatique et post-édition](#).” Survey from the ATLF, March 2023.



# Limitations and future work

21

- ◆ Assessment of the quality during actual play sessions:
  - integration of subtitles and text in-game
  - better evaluation of guiding strategies and semantic splits
  - effect on overall enjoyment...
- ◆ Assessment of the (usefulness of the) MT system during an actual PE task
- ◆ More research and the socio-ergonomic issues of MT and video games
- ◆ More research on the views of translators / interviews with translators

- ◆ ATLF. “Traduction automatique et post-édition.” Survey from the ATLF, March 2023.
- ◆ Brenner, Judith. “The MTxGames Project: Creative Video Games and Machine Translation - Different Post-Editing Methods in the Translation Process.” *Proceedings of the 25th Annual Conference of the European Association for Machine Translation*, edited by Carolina Scarton *et al.*, vol. 2, EAMT, 2024, pp. 47-48.
- ◆ Guerberof-Arenas, Ana and Antonio Toral. “Creativity in Translation: Machine Translation as a Constraint for Literary Texts.” *Translation Spaces*, vol. 11, no. 2, 2022, pp. 184-212.
- ◆ Hansen, Damien and Emmanuelle Esperança-Rodier. “Human-Adapted MT for Literary Texts: Reality or Fantasy?” *Proceedings of the New Trends in Translation and Technology Conference*, edited by Sheila Castilho *et al.*, Incoma Ltd, 2023, pp. 178-190.
- ◆ Hansen, Damien and Pierre-Yves Houlmont. “A Snapshot into the Possibility of Video Game Machine Translation.” *Proceedings of the 15th Biennial Conference of the AMTA*, edited by Janice Campbell *et al.*, vol. 2, AMTA, 2022, pp. 257-269.

# Bibliography

23

- ◆ Hansen, Damien and Susan Pickford. “Traduire la littérature populaire avec l'aide de la machine : quels avantages et préoccupations pour les professionnel·le·s ?” Presentation at the seminar *TRACT 2021-2022*, 14 Apr. 2022, Paris, France.
- ◆ Houlmont, Pierre-Yves. “La localisation vidéoludique au prisme de la contingence.” Presentation at the conference *Interculturalité et sexualité dans les jeux vidéo*, 17 Nov. 2022, Mons, Belgium.
- ◆ Jiménez-Crespo, Miguel A. “Perspectives in localization: From MT to AI.” *Localization in Translation*. Routledge, 2024.
- ◆ Kenny, Dorothy and Marion Winters. “Machine Translation, Ethics and the Literary Translator’s Voice.” *Translation Spaces*, vol. 9, no 1, John Benjamins Publishing Company, 2020.
- ◆ Klein, Guillaume *et al.* “OpenNMT: Open-Source Toolkit for Neural Machine Translation.” *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, edited by Mohit Bansal and Heng Ji, ACL, 2017, pp. 67-72.



# Bibliography

24

- ◆ Koehn, Philipp *et al.* “Moses: Open Source Toolkit for Statistical Machine Translation.” *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, edited by Sophia Ananiadou, ACL, 2007, pp. 177-180.
- ◆ Kudo, Taku. “Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates.” *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, edited by Iryna Gurevych and Yusuke Miyao, ACL, 2018, pp. 66-75.
- ◆ Larssonneur, Claire. “L’algorithme sert-il les traducteurs ? Conditions et contexte de travail avec les outils de traduction neuronale.” *Parallèles*, vol. 35, no. 2, 2023, pp. 90-103.
- ◆ Moorkens, Joss, Andy Way and Séamus Lankford. “Machine Translation in Multimedia Translation and Localisation.” *Automating Translation*. Routledge, 2024.



- ◆ Moorkens, Joss *et al.* “Translators’ perceptions of literary post-editing using statistical and neural machine translation.” *Translation Spaces*, vol. 7, no 2, 2018, pp. 240-262.
- ◆ Ó Murchú, Eoin P. “Using Intergaelic to Pre-Translate and Subsequently Post-Edit a Sci-Fi Novel from Scottish Gaelic to Irish.” *Proceedings of the Qualities of Literary Machine Translation*, edited by James Hadley *et al.*, EAMT, 2019, pp. 20-25.
- ◆ Rivas-Ginel, María Isabel. *The ergonomics of CAT tools for video game localisation*. Doctoral thesis, Université Bourgogne Franche-Comté, 2023.
- ◆ Taivalkoski-Shilov, Kristiina. “Ethical Issues Regarding Machine(-Assisted) Translation of Literary Texts.” *Perspectives*, vol. 27, no. 5, 2019, pp. 689-703.
- ◆ Vaswani, Ashish *et al.* “Attention Is All You Need.” *Proceedings of the 31st International Conference on Neural Information Processing Systems*, edited by Ulrike von Luxburg *et al.*, Curran Associates Inc., 2017, pp. 6000-6010.

Don't hesitate to reach out:

[pierre-yves.houlmont@heaj.be](mailto:pierre-yves.houlmont@heaj.be)

[damien.hansen2@ulb.be](mailto:damien.hansen2@ulb.be)  
(@dhansen.bsky.social)

These slides and bibliography  
will be made available  
on our repository  
(ORBi)

Thank you for your attention

