

ÉTUDE COMPARATIVE ENTRE LA TRADUCTION ET DE LA POST-ÉDITION DE BALDUR'S GATE III

Nouvelles didactiques de la traduction et de
l'interprétation, Paris, 14 février

Simon Copet

Contexte

- Ces dernières années : amélioration de la qualité de la traduction automatique neuronale (TAN) → PE est devenue une pratique professionnelle (Cui *et al.*, 2023 : 1)
- Avènement des grands modèles de langue (LLM) et amélioration de leurs TA brutes (Peng *et al.*, 2023 : 1)
- Le secteur de la localisation de jeu vidéo (ici : traduction) ne fait pas exception (Mangiron & O'Hagan, 2013 : 96 ; Hansen & Houlmont, 2022)
- → Les développeurs et les studios utilisent la traduction automatique (TA) pour limiter les coûts dans un marché saturé (Rivas Ginel & Theroine, 2022 : 2)

La localisation de JV

- La localisation est devenue un incontournable pour les studios, car elle leur permet de maximiser leur retour sur investissement (Mangiron, 2018 : 123)

- Localisation vers de plus en plus de langues (Bernal-Merino, 2015 : 80)

- Contraintes particulières :
 - 1) Privation du contexte multimodal (*ibid.* : 81) → Seuls 29,68% des localisateurs y ont accès (Theroine *et al.* 2021 : 28)

 - 2) Non-linéarité textuelle (Díaz Montón, 2007 : 3 ; Bernal-Merino, 2015 : 80 ; Pettini, 2022 : 47)

 - 3) Manque de ressources terminologiques (Theroine *et al.* 2021 : 28)

La localisation de JV & TA/PE

Recherches sur la TA/PE de jeux vidéo sont peu nombreuses (processus et produit), même si la TA est bien présente et que la localisation de JV présente certaines contraintes.

Rivas Ginel et Theroine (2022) : Étude de la traduction du genre en JV par des moteurs génériques.

Hansen et Houlmont (2022) : Spécialisation d'un moteur de TAN adapté au JV

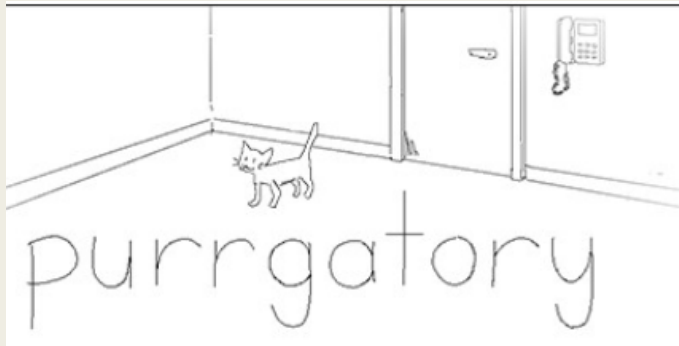
- Résultats prometteurs malgré le peu de données (1 million de segments)
- Soulignent la nécessité d'une bonne intégration dans la chaîne de développement

État de l'art



Copet & De Faria Pires (2023) :

- Analyse d'un Visual Novel (*Killer Trait*, Prikarin, 2023), avec évaluations humaines (fidélité et fluidité) et métriques d'évaluation automatique (Copet & De Faria Pires, à paraître)
- Analyse du *post-edite* : « the unique features that set machine translated post-edited texts apart from human-translated texts » (Castilho & Resende, 2022 : 1)
- Jeu contenant peu d'irrealia (= éléments propres à un univers fictif, Loponen, 2009) : 7
- Résultats de ChatGPT (PE) ~ TH et DeepL < ChatGPT (TA brutes et PE)



De Faria Pires (2025) :

- Jeu riche en jeux de mots, mais très pauvre en irrealia (Purrigatory, 2021) : 2
- Métriques d'évaluation automatique
- Résultats différents : DeepL < Google < ChatGPT

Objectifs

- Perspective contrastive par rapport à Copet & De Faria (2023) en analysant un jeu riche en irrealia
- Analyse de la qualité de la PE de jeux vidéo (évaluation humaine)
 - *Comparaison entre moteurs*
 - *Comparaison avec la TH*
- Analyse du *post-edite*



Baldur's Gate 3

- Game Award du jeu de l'année 2023
- Univers fantastique basé sur Donjons et Dragons
- RPG (= riche sur le plan narratif)
- 228558 segments (extraction le 2 février 2024, avec LSLIB v.1.18.7)
- 20 langues (SteamDB, 2025).

Cadre expérimental : étudiants et évaluation

- Cohorte de 21 étudiants de Master 2 (finalité traduction multidisciplinaire) ayant suivi une introduction à la localisation en Master 1
- 3 modalités (distribution aléatoire) : 7 TH, 7 PE (DeepL) et 7 PE (ChatGPT 4o-mini)
- Choix des moteurs basé sur leur popularité (cf. Loock & Holt, 2024 : 17)
- Accès exclusivement aux données textuelles = conditions authentiques
- Contexte de non-linéarité textuelle = conditions authentiques
- 2 heures (max) pour 296 mots, pour 26 irrealia
- Collecte des données par PosEdiOn (Alvarez-Vidal *et al*, 2023)

PosEdiOn

- Métriques d'évaluation automatique (BLEU et HTER)
- Diverses informations : nombre de frappes, substitutions, ajouts etc.
- Temps (par segment et total)

The screenshot shows a window titled "PosEdiOn" with a text area containing two lines of text. The first line is in English: "We disarmed the Monastery's powerful trap, saving Crèche Y'llek and ourselves - and escaped with the Blood of Lathander." The second line is in French: "Nous avons désamorcé le puissant piège du Monastère, sauvant Crèche Y'llek et nous-mêmes, et nous avons échappé avec le Sang de Lathandre." The French text is highlighted in green. At the bottom of the window, there is a control bar with a timer showing "1:43:10", a progress indicator "8/24", and buttons for "RESTART", "<<", "ACCEPT", ">>", "GO TO", and "Clear".

```
PosEdiOn
We disarmed the Monastery's powerful trap, saving Crèche Y'llek and ourselves -
and escaped with the Blood of Lathander.

Nous avons désamorcé le puissant piège du Monastère, sauvant Crèche Y'llek et
nous-mêmes, et nous avons échappé avec le Sang de Lathandre.

1:43:10  8/24  RESTART  <<  ACCEPT  >>  GO TO  Clear
```

Évaluation : fidélité et fluidité

Évaluation de la qualité en traduction (TQA)

- Question délicate en raison de la subjectivité humaine inhérente à l'opération (Freitag *et al.*, 2021 : 1460)
- Fidélité et fluidité : approche régulièrement utilisée (ex. Daems *et al.*, 2013)
 - Quelques peu réductrice mais aisée à mettre en œuvre pour les évaluateurs
 - Données chiffrées fournies par évaluateurs humains : résultats exploitables
 - Objectif : idée globale de la qualité selon ces deux critères

2 évaluateurs externes professionnels et familiers à la localisation de JV :

- Évaluations TH/PE/localisation officielle à l'aveugle
- Score /5 (échelle de Likert) pour la fidélité et la fluidité

Évaluation avec métriques à venir...

METEOR (Rei *et al.*, 2019) et celles proposées par SacreBLEU, à savoir BLEU, HTER, chr++ (Post, 2018)

Résultats : données chiffrées

La fidélité

	Évaluateur 1	Évaluateur 2	Temps moyen
Localisation officielle	4,82	4,88	N.A.
TH étudiantes	3,92	3,9	1h32
PE DeepL (TA brute)	4,16 (4,1)	4,11 (4,08)	1h03
PE ChatGPT 4o-mini (TA brute)	4,28 (3,91)	4,2 (4,05)	1h05

Localisation officielle = scores élevés chez les 2 évaluateurs

TA brutes plus élevées que les TH

Plus de gains avec ChatGPT qu'avec DeepL en PE

Différence de temps significative entre les TH et les PE

Résultats plus bas que dans Copet et De Faria Pires (2023) :

	Évaluateur 1 (TA brute)	Évaluateur 2 (TA brute)
PE DeepL (n = 9)	4,3089 (4,0244)	4,4282 (4,4878)
PE ChatGPT (n = 10)	4,4439 (4,3171)	4,7049 (4,6098)
TH (n = 9)	4,4580	4,4905

Résultats : données chiffrées

La fluidité

	Évaluateur 1	Évaluateur 2	Temps moyen
Localisation officielle	4,92	4,87	N.A.
TH étudiantes	3,91	3,95	1h32
PE DeepL (TA brute)	3,77 (3,69)	3,78 (3,82)	1h03
PE ChatGPT 4o-mini (TA brute)	4,09 (3,52)	4,04 (3,63)	1h05

Scores élevés pour la localisation officielle

TH toujours au-dessus des TA brutes

Amélioration systématique des TA, sauf pour DeepL chez éval. 2

Comme pour la fidélité, la PE ChatGPT dépasse la TH et la PE DeepL

Résultats plus bas que dans Copet et De Faria Pires (2023) :

	Évaluateur 1 (TA brute)	Évaluateur 2 (TA brute)
PE DeepL (n = 9)	4,7507 (4,8049)	4,8591 (4,7561)
PE ChatGPT (n = 10)	4,8098 (4,8537)	4,7951 (4,7805)
TH (n = 9)	4,8158	4,7751

Quelques tendances notables

Davantage de fautes de distraction dans les TH → influence sur le fond et la forme

Fond :

TS : *You see Dror Ragzlin writhe. A tadpole clings to the mind flayer's fingers.*

LO : *Vous voyez Dror Ragzlin être pris de spasmes. Une larve est accrochée aux doigts du flagelleur mental.*

TH : *Tu aperçois Dror Ragzlin **en train d'écrire**. Un têtard est accroché aux doigts du Flagelleur mental.* (2 cas sur 7)

Forme :

Nous avons désarmé le puissant piège du Monastère, **saucé** la Crèche Y'llek et nous-mêmes, et nous nous sommes échappés avec le Sang de Lathandre.

Tu rends le garçon mal à l'aise - C'est toi qui **a** éliminé **Ketheruc** dans les ombres non ? (à la place Ketheric)





Post-editeuse

Certaines des erreurs induites par les sorties de TA ne sont pas corrigées, et ce sur plusieurs plans :

Typographique

TS : *You see Dror Ragzlin writhe. A tadpole clings to the mind flayer's fingers.*

TA DeepL (et ChatGPT) : *Vous voyez Dror Ragzlin se tordre. Un têtard s'accroche aux doigts du tueur d'esprit.

⇒ Il manque un astérisque

⇒ Seuls 3 étudiants sur 14 ont corrigé l'erreur (TH : 6/7)

L'humour

TS : By Graz'zt's cock, it's about time you came.

Loca officielle : Par le mandrin de Graz'zt, j'ai failli attendre !

TA DeepL : Par la bite de Graz'zt, il est temps que tu jouisses.

Erreur de sens systématiquement corrigée (7/7)

Mais perte de l'humour (atténuation du registre) : « Par tous les dieux », « par tous les princes », etc.

TA ChatGPT : Par la queue de Graz

- Graz à la place de Graz'zt (3/7)
- Traductions littérales (5/7)

TH : créativité

Par les tétons de Graz'zt, il était temps que tu te pointes !

Par la bite de Graz'zt, je jouis de ta venue tardive...

Mais problèmes de registre :

Pour l'amour de Graz'zt, il était temps.

Par la barbe de Graz'zt, il t'en aura fallu du temps.



L'idiomaticité

Segments non-idiomatiques résiduels :

TS : Didst thou mother sex a plague-rat

Loca officielle : Votre mère vous a eu avec un rat porteur de la peste bubonique, c'est ça ?

TA ChatGPT (et DeepL) : Ta mère a-t-elle couché avec un **rat de peste** ?

PE étudiantes :

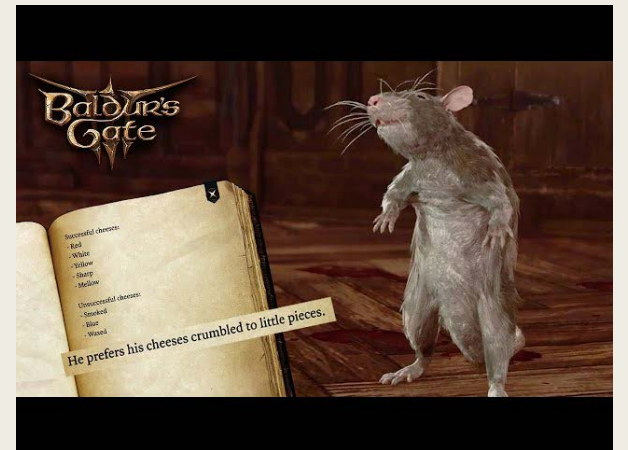
Ta mère a-t-elle couché avec un **rat de (la) peste** ? (9 PE sur 14)

TH :

Vostre mère a-t-elle copulé avec un rat pestiféré ?

Ta matriarche ne serait-elle pas envoyée en l'air avec un un rat ayant la peste ?

Votre mère se serait-elle dévergondée avec un rat pestiféré ?



Les pronoms

Choix de pronom

Source : You are too modest, great one. This name perfectly captures your essence - your fury

Loca officielle : Vous êtes trop modeste, Votre Sainteté. Ce nom saisit parfaitement votre essence, votre fureur.

TA DeepL : **Tu** es trop modeste, **mon grand**. Ce nom reflète parfaitement ton essence, **ta** fureur.

PE étudiantes :

Tu es trop modeste, **mon grand**. Ce nom reflète parfaitement ton essence, **ta** fureur. (10 PE sur 14).

Davantage de créativité dans la TH :

« ô tout puissant », « grand chef », « ô vénérable », etc.

Mais tout de même un problème de pronom (2 cas sur 7)

La traduction des irréalias

TS: Spoken like someone who's never heard a dragonborn throatsong.

Loca officielle : Voilà qui est parlé comme quelqu'un qui n'a jamais entendu un chant de gorge **drakéide**.

TA brute ChatGPT : On dirait quelqu'un qui n'a jamais entendu un chant guttural de **drakéide**. = Bonne terminologie → bonne PE

TA brute DeepL : Vous parlez comme quelqu'un qui n'a jamais entendu le chant d'un **dragonborn**.

TH : résultats très mitigés.

Innovations de la part des étudiants (= créativité), mais glissements de sens, manque d'idiomatisme :
« dragonborn », « dragon-né », « enfant de dragon » (= référence à *Skyrim*), « bébé dragon ».



La traduction des irrealia

TS : I'll find my way back, as I always do. I admit defeat. Baldur's Gate is my home.

Loca officielle : Mais je finirai bien par revenir, comme toujours. Je reconnais ma défaite. La Porte de Baldur, c'est chez moi, désormais.

TA brutes : Je retrouverai mon chemin, comme toujours. J'admets ma défaite. **Baldur's Gate** est ma maison.

PE : « Baldur's Gate » = 5/14

TH « « Baldur's Gate » = 3/7.

Pourtant, « Porte de Baldur » était bien présent dans l'extrait du jeu **qu'ils ont lu en amont.**

Discussion

Résultats à nuancer :

- Les étudiants ne sont pas des professionnels, et pas forcément adeptes des JV
- Leur créativité apporte toutefois une réelle plus-value
- Les étudiants éprouvent davantage de difficultés lorsqu'ils sont confrontés à des irrealia

Cependant, la PE de ChatGPT a montré de meilleurs résultats (fidélité/fluidité) que la TH

- IA doit rester un outil
- *Post-editese*
- Qualité largement en-deçà de la localisation officielle (dans tous les cas)

Cette rejoint la conclusion de Copet et De Faria (2023), mais en exacerbant les tendances observées (>< De Faria Pires, 2025)

- Soulève la question du choix du moteur en fonction du genre

Perspective de recherches

- Davantage d'évaluateurs
- Prise en compte de la composante multimodale (avec et sans linéarité textuelle)
- Comparaison avec des professionnels
- Comparaison avec des moteurs spécialisés
- Comparaison avec différents prompts

Références

Alvarez-Vidal, S., & Oliver, A. (2023). Assessing MT with measures of PE effort. *Ampersand*, 11. <https://www.sciencedirect.com/science/article/pii/S2215039023000176>

Bernal-Merino, M. Á. (2015). *Translation and Localisation in Video Games*. New-York : Routledge.

Castilho, S. & Resende, N. (2022). *Post-Editese in Literary Translations*. *Information*, 13(66), 1-22. <https://www.mdpi.com/2078-2489/13/2/66>

Copet, S., & De Faria Pires, L. (2023). Post-édition de TAN et localisation de jeux vidéo : quelles conséquences sur la qualité et l'immersion des joueurs ? . [Paper presentation]. *Journée d'étude Traduire le jeu vidéo: entre immersion, interactivité et interaction*, Orléans, France.

Copet, S. (2025). Alice in localizationland: an English-French translation and transmedia study of Disney characters' speech in films and video games. [Paper presentation]. *Fun for All VII: International Conference on Video Game Translation and Accessibility - "Shaping the Future of Game Accessibility and Localization"*, Barcelone, Espagne.

Daems, J., Macken, L., & Vandepitte, S. (2013). Quality as the sum of its parts: a two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE. In S. O'Brien, M. Simard, & L. Specia (Eds.), *MT Summit XIV Workshop on Post-editing Technology and Practice, Proceedings*, 63–71.

De Faria Pires, L. (2025). NMT and LLM Post-Editing Process in Videogame Localization: How Can Students' Productions Help Us Become Better Localisation Lecturers? [Paper presentation]. *Fun for All VII: International Conference on Video Game Translation and Accessibility - "Shaping the Future of Game Accessibility and Localization"*, Barcelone, Espagne.

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

Références

Cui, Y., Liu, X., & Cheng, Y. (2023). A Comparative Study on the Effort of Human Translation and Post-Editing in Relation to Text Types: An Eye-Tracking and Key-Logging Experiment. *SAGE Open*, 13(1), 1-15.

https://www.researchgate.net/publication/370056187_A_Comparative_Study_on_the_Effort_of_Human_Translation_and_Post-Editing_in_Relation_to_Text_Types_An_Eye-Tracking_and_Key-Logging_Experiment

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, Errors and Context: A Large-Scale Study of Human Evaluation for Machine Translation. *Transactions of the Association for Computational Linguistics*, 9, 1460-1474.

<https://arxiv.org/pdf/2104.14478.pdf>

Hansen, D. & Houlmont, P-Y. (2022). A Snapshot into the Possibility of Video Game Machine Translation. *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas*, 2, 257-269.

<https://orbi.uliege.be/bitstream/2268/294581/1/2022.amta-upg.18.pdf>

Loock, R., & Holt, B. (2024). Augmented linguistic analysis skills: Machine translation and generative AI as pedagogical aids for analyzing complex English compounds. *Technology in Language Teaching & Learning*, 6(3), 1489. <https://doi.org/10.29140/tltl.v6n3.1489>

Loponen, M. (2009). Translating Irrealia – Creating a Semiotic Framework for the Translation of Fictional Cultures. *Chinese Semiotic Studies*, 2(1), 165-175. <https://doi.org/10.1515/css-2009-0117>

Références

Mangiron, C. (2018). Game on! Burning issues in game localisation. *Journal of Audiovisual Translation*, 1(1), 122-138.

Pettini, S. (2022). *The Translation of Realia and Irrealia in Game Localization: Culture-Specificity between Realism and Fictionality*. New-York : Routledge.

Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Ouyang, Y., & Tao, D. (2023). Towards making the most of ChatGPT for machine translation. *EMNLP 2023*, 1-12. <https://doi.org/10.48550/arXiv.2303.13780>

Rivas Ginel, M. I., & Theroine, S. (2022). Machine Translation and Gender Biases in Video Game Localisation: A Corpus-Based Analysis. *Journal of Data Mining and Digital Humanities (Vers une robotique du traduire)*. <https://jdmdh.episciences.org/10245>

Theroine, Sarah, Rivas Ginel, M. I., & Perrin, A. (2021). Au cœur de la terminologie du jeu vidéo. L'absence de ressources, frein majeur pour les traducteurs. *Traduire*, 244, 27-40.



Merci pour votre
attention

Contact : Simon.Copet@ulb.be

