

Human-Adapted MT for Literary Texts: Reality or Fantasy?

Damien Hansen ^{1,2} & Emmanuelle Esperança-Rodier ¹

¹ Grenoble Alpes University, France – ² University of Liège, Belgium

Starting Point

- Workshop on Translation Technology for Creative Domains: Individualized MT systems for literary translators?
 - We can adapt a system not just to literary data, but to a specific author or translator, in a specific genre.
 - This might be a way to mitigate the ethical and quality issues that come with the introduction of this technology.
 - But what is produced? How has it actually adapted?

Fine-tuning on literary data

- Trained with OpenNMT v2 (Klein et al. 2017) and base Transformer parameters (Vaswani 2017).
- Vocabulary of 16K encoded with SentencePiece unigram language model (Kudo 2018).

Corpus	Segments	EN tokens	FR tokens
Europarl	2,007,723	49,867,465	54,553,979
Vid. Game	1,370,431	21,041,902	22,804,380
TED	410,443	7,041,745	7,464,033
GVoices	195,387	3,503,600	3,980,602
News	183,251	4,055,180	4,952,704
Books	127,021	2,737,133	2,770,418
Total	4,294,256	88,247,025	96,526,116

Generic corpora

(Tiedeman 2012; Hansen and Houlmont forth.)

Corpus	Segments	EN tokens	FR tokens
Synthetic	338,233	14,339,224	15,130,086
Translator	111,322	3,571,242	3,569,595
Parallel	100,055	4,014,409	4,365,486
Sep. (trn)	37,348	550,536	541,779
Sep. (val)	7,225	109,859	106,621
Sep. (tst)	704	10,181	10,073
Total	594,887	22,595,451	23,723,640

Specialized corpora

Assessing the output of the system

- Automated evaluation:
 - MT quality metrics;
 - lexical richness, to reflect *simplification*;
 - length ratio, to reflect *explicitation*;
 - syntactic equivalence, to reflect *convergence* (Baker 1996).

- Human evaluation:
 - error annotation;
 - general observations.

Evaluation Metrics

- Provided by SacreBLEU (Post 2018) and COMET (Rei et al. 2020).
<https://github.com/mjpost/sacrebleu> / <https://github.com/Unbabel/COMET>
- Paired t-test and bootstrap resampling to confirm statistical significance (Koehn et al. 2004).

System	BLEU ↑	chrF2++ ↑	TER ↓	COMET ↑
Google Tr.	10.79	35.20	91.08	-0.240
DeepL	10.04	34.88	92.81	-0.248
Generic	09.93	33.14	92.24	-0.388
Septimus v1	18.56	40.43	76.06	-0.126
Septimus v2	19.08	41.44	75.98	-0.066

Table 1: Comparison of metrics for in-house and online systems.

Lexical Richness

- Provided by the LexicalRichness python module. Lemmatization with Stanza (Qi et al. 2020).
<https://github.com/LSYS/LexicalRichness> / <https://github.com/stanfordnlp/stanza>
- Window of 500 tokens for MSTTR (Johnson 1944) & MATTR (Covington and McFall 2010).

System	Tokens	Types	TTR	MSTTR	MATTR
Reference	10,348	1,900	0.18	0.45	0.45
Septimus	9,601	1,598	0.17	0.42	0.42
Google Tr.	11,351	1,699	0.15	0.41	0.41
DeepL	11,423	1,688	0.15	0.40	0.40

Table 2: Measures of lexical richness.

Length Ratio

- Average length in words.
- Calculated on segments, which can contain multiple sentences (up to five).

System	Average length	Expanding ratio
Source	14.26	—
Reference	14.70	+3.04
Septimus	13.64	-4.40
Google Tr.	16.12	+13.03
DeepL	16.22	+13.75

Table 3: Average segment length and expanding ratio.

Syntactic Equivalence

- Provided by ASTRÉD (Vanroy et al. 2021). Alignment with AwesomeAlign (Dou and Neubig 2021).
<https://github.com/BramVanroy/astred> / <https://github.com/neulab/awesome-align>
- Calculated on sentences that have a 1-to-1 correspondence in all translations (404 out of 704).

System	SACr	POS changes	ASTrED
Reference	1728	1116	3674
Septimus	891	975	2683
Google Tr.	492	969	2294
DeepL	558	974	2494

Table 4: Measures of syntactic equivalence.

Error Annotation

Adequacy Total: 617

Content				MTese		Vocabulary		Meaning		
Omission	Addition	Over-trans.	Under-trans.	Stuttering	Hallucination	Non-trans.	Mis-trans.	Opposite meaning	Non-sense	Shift in meaning
270	6	8	38	3	12	0	200	11	26	43

Fluency Total: 251

Coherence			Agreement		Conugation		Syntax			
Referential	Relational	Logic	Gender	Number	Tense	Person	Function words	Punctuation		
31	3	60	19	6	51	7	54	20		

Literary aspects Total: 382

Prose				Terminology		Typography				
Style	Register	Unfitting paraphr.	Adap-tation	Cohe-rence	Loss	Dialo-gues	Case			
218	38	8	15	29	10	44	20			

Error Annotation

- Carried out by two annotators in Accolé (Esperança-Rodier and Brunet-Manquat 2021).

◀ Aller au segment précédent
Tableau des segments
Valider le segment courant ✓
Aller au segment suivant ▶

Phrase source

Jenna looked around the entrance hall, which was full of shadows and ghosts. She reached into her pocket for the Candlelight Charm Septimus had given her that afternoon and began lighting **the first of many extinguished candles.**

Phrase référence

Jenna promena son regard autour du hall, peuplé d'ombres et de fantômes, puis elle sortit de sa poche l'allume-bougies magique que Septimus lui avait offert cet après-midi-là et entreprit de rallumer quelques-unes des dizaines de chandelles que la Ténèbre avait soufflées.

Phrase cible

Jenna promena son regard autour du hall d'entrée, qui était rempli d'ombres et de fantômes, et plongea la main dans sa poche pour récupérer le charme que Septimus lui avait offert cet après-midi-là. Elle commença à allumer **les premières bougies éteintes.**

Source	Cible	Erreur	Actions
the first of many extinguished candles	les premières bougies éteintes.	I - Logique	Ajouter l'erreur

Récapitulatif [Supprimer des erreurs](#)

Source	Cible	Erreur	Actions
Candlelight Charm	charme	Adéquation > Contenu > Sous-traduction	

Développé par FBM

Copyright 2014 - GETALP LIG

10

Error examples : (L) Style (17.4%)

Source	DeepL	Custom MT	Reference
<p><i>I believe there may be a reason for this. Over the years the Hermetic Chamber has seen much Darke Magyk, and I suspect the Pick has become corrupted.</i></p> <p>-----</p> <p><i>It was a beautiful frosty morning.</i></p>	<p><i>Je crois qu'il y a peut-être une raison à cela. Au fil des ans, la Chambre Hermétique a vu beaucoup de Darke Magyk, et je soupçonne que le Pick a été corrompu.</i></p> <p>-----</p> <p><i>C'était un beau matin givré.</i></p>	<p><i>Je crois qu'il y a une raison à cela. Au fil des ans, le cabinet hermétique a vu beaucoup de Magyk noire, et je pense que la désignation a été corrompue.</i></p> <p>-----</p> <p><i>C'était une belle matinée glaciale.</i></p>	<p><i>Le cabinet hermétique a maintes fois été exposé à la Ténèbre, et je crains que celle-ci n'ait fini par affecter la désignation.</i></p> <p>-----</p> <p><i>C'était une magnifique matinée d'hiver.</i></p>

Error examples : (L) Register (3%)

Source	DeepL	Custom MT	Reference
Alther, is that <u>you</u> ?"	Alther, c'est toi ?"	Alther, c'est <u>toi</u> ?	C'est <u>vous</u> , Alther ?
-----	-----	-----	-----
"Marcellus, <u>do you have</u> six guineas on <u>you</u> ?" asked Marcia.	"Marcellus, tu as six guinées sur toi ?" demande Marcia.	– Marcellus, <u>tu as</u> six guinées sur <u>toi</u> ?	– Marcellus, <u>auriez-vous</u> six guinées sur <u>vous</u> ? demanda Marcia.
-----	-----	-----	-----
" <u>Let's get out of this</u> <u>place</u> ."	"Sortons de cet endroit."	<u>Tirons-nous d'ici</u> ."	<u>Rentrons vite chez nous</u> ."

Error examples : (L) Dialogues (3.5%)

Source	DeepL	Custom MT	Reference
<p><i>"Does that mean that Beetle is..."</i></p> <p>-----</p> <p><i>"Yeah, Foxo," he said.</i></p> <p><i>"I am. I am very all right."</i></p> <p>-----</p> <p><i>She continued.</i></p> <p><i>"During the last few days in the Wizard Tower, when I was trying to find a way to defeat the Darke Domaine [...]"</i></p>	<p><i>"Cela signifie-t-il que Beetle est..."</i></p> <p>-----</p> <p><i>"Oui, Foxo", a-t-il dit. "Je le suis. Je suis très bien. "</i></p> <p>-----</p> <p><i>Elle a continué.</i></p> <p><i>"Pendant les derniers jours dans la Tour des Sorciers, quand j'essayais de trouver un moyen de vaincre le Domaine de Darke [...]"</i></p>	<p><i>Ça veut dire que Moustique est...</i></p> <p>-----</p> <p><i>– Ouais, Vulpin.</i></p> <p><i>– Je vais très bien.</i></p> <p>-----</p> <p><i>Elle poursuivit :</i></p> <p><i>– Ces derniers jours, à la tour du Magicien, alors que je cherchais un moyen de vaincre le domaine ténébreux [...]"</i></p>	<p><i>– Ça veut dire que Moustique...</i></p> <p>-----</p> <p><i>– Oui, répondit-il. Parfaitement bien.</i></p> <p>-----</p> <p><i>Marcia reprit :</i></p> <p><i>– Ces derniers jours, tandis que je cherchais comment vaincre le domaine ténébreux, enfermée dans la tour [...]"</i></p>

Error examples : (F) Tense (4.1%)

Source	DeepL	Custom MT	Reference
<p><i>Simon <u>pushed his way</u> through the Darke Fog, terrified that at any moment a Thing would recognize him.</i></p>	<p><i>Simon s'est frayé un chemin à travers le brouillard de Darke, terrifié à l'idée qu'à tout moment une Chose puisse le reconnaître.</i></p>	<p><i>Simon <u>se fraya un</u> chemin à travers le brouillard, terrifié à l'idée qu'une créature puisse le reconnaître.</i></p>	<p><i>Simon <u>se frayait un</u> chemin à travers le brouillard, terrifié à l'idée qu'une créature le reconnaisse.</i></p>

Error examples : (F) Logic (4.8%)

Source

The discs seemed to have fused into one and it was clear that the lines that radiated out from the Manuscriptorium Code joined up with some of those on the Wizard Tower Code. Each one of these led to a symbol.

DeepL

Les disques semblaient avoir fusionné en un seul et il était clair que les lignes qui rayonnaient du code du manuscritoire rejoignaient certaines de celles du code de la tour du sorcier. Chacune d'entre elles menait à un symbole.

Custom MT

Les disques semblaient avoir fusionné et il était évident que les lignes rayonnantes du code du Manuscriptorium rejoignaient celles du code de la tour du Magicien, chacune donnant naissance à un symbole.

Reference

Les disques semblaient avoir fusionné, et les lignes qui rayonnaient à partir du centre du plus petit se prolongeaient dans certaines de celles qui étaient gravées à la surface du second pour aboutir à un symbole.

Error examples : (F) Function words (4.3%)

Source	DeepL	Custom MT	Reference
<p><i>I don't want to leave the Manuscriptorium unattended.</i></p> <p>-----</p> <p><i>He thought of Larry. Of Matt, Marcus and Igor at Gothyk Grotto, even the oddly irritating people at Wizard Sandwiches.</i></p>	<p><i>Je ne veux pas laisser le manuscritoire sans surveillance.</i></p> <p>-----</p> <p><i>Il a pensé à Larry. À Matt, Marcus et Igor à la Grotte de Gothyk, et même aux gens bizarrement irritants de Wizard Sandwiches.</i></p>	<p><i>Je ne veux pas laisser le Manuscriptorium <u>sous</u> surveillance.</i></p> <p>-----</p> <p><i>Il songea à Larry, à Matt, Marcus et Igor, <u>à la</u> Grotte-Gothic, même aux gens bizarrement agaçants de Magyk Sandwich.</i></p>	<p><i>Je n'aimerais pas laisser le Manuscriptorium <u>sans</u> surveillance.</i></p> <p>-----</p> <p><i>Puis il pensa à Larry, à Matt, Marcus et Igor, <u>de</u> la Grotte-Gothic, et même à la bande d'hurluberlus parfois exaspérants de Magyk Sandwich.</i></p>

Error examples : (A) Meaning (6.4%)

Source	DeepL	Custom MT	Reference
<p><i>Marcia, Septimus and Jenna emerged from the Great Arch and paused for a moment, looking down <u>the newly liberated Wizard Way</u>.</i></p> <p>-----</p> <p><i>Romilly walked <u>self-consciously</u> into the seven-cornered passage.</i></p>	<p><i>Marcia, Septimus et Jenna sont sortis de la Grande Arche et se sont arrêtés un moment, regardant la Voie des Sorciers nouvellement libérée.</i></p> <p>-----</p> <p><i>Romilly a marché d'un air gêné dans le passage à sept angles.</i></p>	<p><i>Marcia, Septimus et Jenna émergèrent de la Grande Arche et marquèrent une pause, contemplant <u>la nouvelle</u> voie du Magicien.</i></p> <p>-----</p> <p><i>Romilly pénétra <u>inconsciemment</u> dans le couloir qui tournait sept fois sur lui-même</i></p>	<p><i>Marcia, Septimus et Jenna émergèrent de la Grande Arche pour contempler la voie du Magicien <u>nouvellement libérée</u>.</i></p> <p>-----</p> <p><i>Romilly s'avança d'<u>un pas timide</u> vers l'entrée du couloir aux sept détours</i></p>

Error examples : (A) Mistranslation (16%)

Source	DeepL	Custom MT	Reference
<p>"SafeShield's <u>out</u>," Marcia said grimly.</p> <p>-----</p> <p>Foxy rushed over to the tiny sink and <u>grabbed</u> the pan <u>out</u> of Beetle's hands.</p>	<p>"Safeshield est sorti", a déclaré Marcia avec sinistrose.</p> <p>-----</p> <p>Foxy s'est précipité vers le petit évier et a pris la casserole des mains de Beetle.</p>	<p>– Le dôme est <u>sorti</u>, dit Marcia d'un ton sinistre.</p> <p>-----</p> <p>Vulpin se précipita vers l'évier et <u>agrippa</u> la casserole des mains de Moustique.</p>	<p>– Le dôme est <u>mort</u>, annonça Marcia d'un ton sinistre.</p> <p>-----</p> <p>Vulpin se rua vers l'évier et <u>arracha</u> la casserole des mains de Moustique.</p>

Error examples : (A) Undertranslation (3%)

Source	DeepL	Custom MT	Reference
<p><i>The previous afternoon Marcia <u>had successfully gathered</u> all scribes together in the Manuscriptorium.</i></p>	<p><i>L'après-midi précédent, Marcia avait réussi à rassembler tous les scribes dans le manuscritoire.</i></p>	<p><i>La veille, Marcia <u>avait réuni</u> tous les scribes du Manuscriptorium.</i></p>	<p><i>La veille, Marcia <u>était parvenue à réunir</u> tous les scribes au Manuscriptorium.</i></p>

Error examples : (A) Omissions (21.6 %)

Source	DeepL	Custom MT	Reference
<p><i>A simple <u>Reveal</u>—why hadn't she thought of that?</i></p> <p>-----</p> <p><i>Septimus read out the <u>second number of the first pair</u>. "Ninety-eight."</i></p>	<p><i>Une simple révélation - pourquoi n'y avait-elle pas pensé ?</i></p> <p>-----</p> <p><i>Septimus a lu le deuxième chiffre de la première paire. "Quatre-vingt-dix-huit."</i></p>	<p><i>Pourquoi n'y avait-elle pas songé ?</i></p> <p>-----</p> <p><i>— Quatre-vingt-dix-huit.</i></p>	<p><i><u>Un sort révélateur...</u></i></p> <p><i>Comment n'y avait-elle pas pensé elle-même ?</i></p> <p>-----</p> <p><i>— Quatre-vingt-dix-huit.</i></p>

How can we say the system has adapted?

- All the metrics point that way.
- Virtually all the vocabulary specific to the series is maintained.
- A lot of heavy syntactic reorganisations that are in line with our translator (although there are less sentence splits).
- Terms are not systematically translated the same way and are sometimes rendered by synonyms or periphrases also used by the translator:
 - *Wizard Tower courtyard* >> *cour de la tour du Magicien, cour du Magicien, cour de la tour.*

How can we say the system has adapted?

- Translation choices similar to those of the translator:
 - recurrent lexical items or phrases;
 - discourse information (e.g. *said grimly* >> *dit d'un ton sinistre*);
 - modulations (e.g. *as they watched the sun rise* >> *tandis que le soleil se levait*);
 - transpositions (eg. *looked in panic* >> *regardèrent, paniqués*
smiled with relief >> *s'exclama, soulagée*);
 - contractions (e.g. *were beginning to fall; they drifted lazily* >> *tombaient paresseusement*
she had rushed off to the Palace as soon as she could >> *elle s'était précipitée au palais*
I shall go and ask to look for >> *j'irai consulter*);
 - ...
- Many omissions shared by the translator: whole propositions, tag answers, discourse information and speaker indications, adjectives in very long nominal groups, repetitions...

General observations

- With subword segmentation and more literary data, we have no untranslated words, more syntactic reorganisation and less grammatical errors, but much less exact matches with the human reference!
- Despite the paraphrases and the adaptation, a number of translations are still way too literal.
- Even with the use of paraphrases and synonyms by the MT system, the translator plays a lot more with names and mentions of characters:
 - e.g. – *they* [the witches] >> *Pamela et ses disciples*;
 - *the Things* >> *les serviteurs de Merrin*.
- Colloquial and spoken language create frequent disambiguation problems.

General observations

- Much less adaptation issues than expected, because the system omits the problem altogether or uses avoidance strategies, which is also problematic.
- 2 out of the 3 chapter names are problematic.
- Out of 41 sentences merged, 33 are shared by the MT and the translator, 2 are merged by MT only, and 6 in the reference only.
- However, out of 30 split sentences, 4 are shared by the MT and the translator, 8 are split by MT only, and 18 in the reference only.
- Out of 270 omissions, 61 are shared, only 12 are really problematic.
- One spoiler!

Conclusions

- Human translation remains, unsurprisingly, far ahead.
- But there might be a real benefit to using MT as a personalized tool, especially for creative domains.
- Perhaps a way to improve creativity and quality, in addition to allowing of a more human-centric way of using MT and re-empowering translators.
- But we would need tools for that!
- And we still have to find the right balance between translator style and quality.

Future Work

- Investigate other metrics, with more emphasis on stylometry.
- Investigate context-aware NMT.
- Interview the official French translator of the series and gather her views on the output of our custom MT system.
- Experiment with other authors and translators, other genres.

Bibliography

- Baker, Mona. “Corpus-Based Translation Studies: The Challenges That Lie Ahead”. *Terminology, LSP and Translation: Studies in Language Engineering*, edited by Harold Somers, John Benjamins, 1996, pp. 175–186, doi: [10.1075/btl.18.17bak](https://doi.org/10.1075/btl.18.17bak).
- Dou, Zi-Yi, and Graham Neubig. “Word Alignment by Fine-tuning Embeddings on Parallel Corpora”. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, edited by Paola Merlo et al., ACL, 2021, pp. 2112–2128, doi: [10.18653/v1/2021.eacl-main.181](https://doi.org/10.18653/v1/2021.eacl-main.181).
- Esperança-Rodier, Emmanuelle, and Francis Brunet-Manquat. “ACCOLÉ : Annotation Collaborative d’erreurs de traduction pour CORpus aLignÉs, multi-cibles, et Annotation d’Expressions Poly-lexicales”. *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles*, edited by Pascal Denis et al., vol. 3, ATALA, 2021, pp. 1–5, <https://aclanthology.org/2021.jeptalnrecital-demo.1>.
- Hansen, Damien, and Pierre-Yves Houlmont. “A Snapshot into the Possibility of Video Game Machine Translation”. Forthcoming.
- Covington, Michael A., and Joe D. McFall. “Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR)”. *Journal of Quantitative Linguistics*, vol. 17, no. 2, 2010, pp. 94–100, doi: [10.1080/09296171003643098](https://doi.org/10.1080/09296171003643098).
- Johnson, Wendell. “Studies in Language Behavior: A Program of Research”. *Psychological Monographs*, vol. 56, no. 2, 1944, pp. 1–15.
- 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, <https://aclanthology.org/P17-4012>.
- Koehn, Philipp. “Statistical Significance Tests for Machine Translation Evaluation”. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, edited by Dekang Lin and Dekai Wu, ACL, 2004, pp. 388–395, <https://aclanthology.org/W04-3250>.

Bibliography

- 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, vol. 1, ACL, 2018, pp. 66–75, doi: [10.18653/v1/P18-1007](https://doi.org/10.18653/v1/P18-1007).
- Post, Matt. “A Call for Clarity in Reporting BLEU Scores”. *Proceedings of the Third Conference on Machine Translation: Research Papers*, edited by Ondřej Bojar et al., ACL, 2018, pp. 186–191, doi: [10.18653/v1/W18-6319](https://doi.org/10.18653/v1/W18-6319).
- Qi, Peng, et al. “Stanza: A Python Natural Language Processing Toolkit for Many Human Languages”. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, edited by Asli Celikyilmaz and Tsung-Hsien Wen, ACL, 2020, pp. 101–108, doi: [10.18653/v1/2020.acl-demos.14](https://doi.org/10.18653/v1/2020.acl-demos.14).
- Rei, Ricardo, et al. “COMET: A Neural Framework for MT Evaluation”. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, edited by Bonnie Webber et al., ACL, 2020, pp. 2685–2702, doi: [10.18653/v1/2020.emnlp-main.213](https://doi.org/10.18653/v1/2020.emnlp-main.213).
- Tiedemann, Jörg. “Parallel Data, Tools and Interfaces in OPUS”. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, edited by Nicoletta Calzolari et al., ELRA, 2012, pp. 2214–2218, http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf.
- Vanroy, Bram, et al. “Metrics of Syntactic Equivalence to Assess Translation Difficulty”. *Explorations in Empirical Translation Process Research*, edited by Michael Carl, Springer, 2021, pp. 259–294, doi: [10.1007/978-3-030-69777-8_10](https://doi.org/10.1007/978-3-030-69777-8_10).
- Vaswani, Ashish, et al. “Attention is All you Need”. *NIPS'17: 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, <https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>.

Thank you for your attention!

Don't hesitate to get in touch:

@LiteraryLudeme / @E_Esperanca

Slides can be found online on ORBi:

<https://orbi.uliege.be/>