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

Damien Hansen
Univ. Grenoble Alpes, CNRS, Grenoble INP*,
LIG, 38000 Grenoble, France
University of Liège, CIRTI**, 4020 Liège, Belgium
damien.hansen
@univ-grenoble-alpes.fr

Emmanuelle Esperança-Rodier
Univ. Grenoble Alpes, CNRS,
Grenoble INP*, LIG,
38000 Grenoble, France
emmanuelle.esperanca-rodier
@univ-grenoble-alpes.fr

Abstract

This article presents the evaluation of a bespoke machine translation (MT) system adapted to the work of a literary translator who specializes in fiction. Actualizing a research avenue for the English–French pair and complementing a growing number of studies for other language combinations, our work on literary machine translation shows that it is possible to fine-tune a base model trained with publicly available tools and datasets and obtain interesting results with a much smaller subset of custom training data. Although the raw unedited output remains nowhere near human production, metrics show that the adapted system produces translations that are much closer to the desired human reference, while further automatic evaluations reflect improvements regarding important textual features for literary texts. Finally, a manual evaluation details the type of errors produced by our system, indicating that these primarily involve adequacy errors that are already well documented for MT and typically harder to spot, rather than purely literary aspects. This annotation process also revealed unsuspected observations and suggests new perspectives for the use of MT. Namely, that of an individually personalized and interactive tool better suited to the translation of creative texts, although its application to the literary domain opens many ethical and practical questions.

1 Introduction

Given the rising interest in the quality, application and ethical considerations of literary machine translation (LMT), we examine the feasibility of adapting a machine translation system not only on literary data, but more specifically on the work of an individual translator.

For this purpose, we trained a Transformer model on a subset of six fantasy novels from a single series, as described in Hansen et al. (2022). The aim of this article is to go beyond simple metrics of performance, to investigate commonly addressed features of MT and human translation (HT), and to provide more insight on if, how much, and in what respect our human-adapted system has adapted—or not—to the individual style displayed in these novels.

Thus, the next section presents some of the related works that have also attempted to train or fine-tune MT models on literature, as well as other studies interested in the production and context of production of literary (machine) translations. Following these, we briefly describe the experiment setting and methodology before highlighting the added value of the adaptation process through standard evaluation metrics, as well as measures of lexical richness, sentence length, and syntactic equivalence. We then end with the results of an error annotation evaluation and overall discussion on concerns and potential directions for LMT.

Ultimately, we find that our system is able to learn and achieve much better performance, even with scarce data, when compared to in-house and online generic tools. The human reference remains well above MT of course, and the comparison between HT and MT in this domain illustrates aspects with which the latter still struggles. Nevertheless, this individualized adaptation hints at a possible scenario where MT might become in the long run a potentially useful interactive aid in complement to other translation and corpus exploration tools. If they were to be trained and used by translators themselves, we likewise wonder if individualized MT systems might help regarding some questions of ergonomics and working conditions, quality and creativity, as well as ethics and language normalization, that surround LMT.

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

New Trends in Translation and Technology, pages 178–190
4–6 July, 2022. ©NeTTT 2022

2 Related Works

Although there has been a long-standing and consistent curiosity for the constrained or automatic production of literary pieces, and more recently for the specific topic of LMT, it is with the advances of neural machine translation (NMT) that the interest really took off (Hansen, 2021). Hence, we can now rely on a growing number of studies investigating the application of LMT with different paradigms or language pairs, and the adaptation of MT systems to literary data. Of particular importance for our work are also studies evaluating the output of either adapted or generic systems on literary texts, as well as research investigating the societal issues surrounding the implementation of such a tool.

To our knowledge, no other study has looked into the topic of LMT for the English–French pair since the pioneering work of Besacier (2014) with statistical machine translation (SMT), who already considered adapting a system on the work of the same author being evaluated, as did Toral and Way (2015), even though the size of the literary corpus was still limited in both cases. Since then, other researchers have used NMT in literature, such as Matusov (2019) with the English–Russian and German–English pairs, Kuzman et al. (2019), who similarly applied it to the English–Slovene combination and found that having a training corpus from a specific author was more useful than a larger corpus of varied literature, and finally Toral and Way (2018) on the English–Catalan pair, who did not fine-tune a generic model on literary texts, but instead trained it exclusively on in-domain data.

In connection with this, Guerberof-Arenas and Toral (2020) have also delved into the evaluation of the translations produced by their adapted system for the same language combination, while other researchers similarly explored this avenue with generic online tools for English–Dutch (Tezcan et al., 2019; Webster et al., 2020; Fonteyne et al., 2020; Macken et al., 2022).

Lastly, another area of research has also focused on the ethical considerations that should become more and more pressing as MT continues to improve and integrate previously unsuspected domains. Such questions might include for example notions of translation quality and work conditions (Taivalkoski-Shilov, 2019), retranslation and plagiarism (Şahin and Gürses, 2019), translator's voice (Kenny and Winters, 2020), and creativity constraints (Guerberof-Arenas and Toral, 2022).

3 Experiment Setting

This study is based on the *Septimus Heap* (Sage, 2015–2013) / *Magyk* (Serval, 2015–2013) saga: a series depicting a world set so far in the future that technology became indistinguishable from magic. Aside from the poetic image that comes with studying this genre in relation to a technology that is often depicted as uncanny and futuristic in mainstream media, our test case was also chosen as it is very representative of the heroic fantasy genre, which itself represents one of the main segments of the literary translation industry. These novels also pose several unique challenges to be found, for instance, in the high variation between very formal or informal register depending on the character speaking, or in the use of archaisms, regionalisms, neologisms, wordplay, etc. Indeed, the same observations encouraged us to choose these books in a previous computer-assisted literary translation experiment which further motivated this selection, as we plan to compare it with this LMT scenario. What is more, a comparison with a small sample of French translations from canonical works seemed to indicate that they are particularly challenging novels for MT, as evidenced by perplexity measures suggested in Toral and Way (2015), which we attributed to the high degree of liberty taken in the translations, and to the lexical field of the fictional universe created in the saga (Hansen et al., 2022).

The saga comprises 7 volumes, originally written in English by Angie Sage, but only 6 of them were published in and translated by Nathalie Serval, as the last tome was never released by the French publisher—which makes us wonder if the use of MT could somehow help French-speaking fans to finally have access to the *dénouement* of the story. This corpus (of about 45K segments) nevertheless serves as a good starting point for a domain adaptation scenario, inasmuch as it constitutes a relatively homogeneous and coherent corpus.

To train our generic system, we used publicly available and commonly used corpora for our language pair (cf. Table 1): Europarl v8, GlobalVoices 2018Q4, News-Commentary v16 and TED2020 v1 (Tiedemann, 2012). We also used the Books corpus from the OPUS repository, after correcting a third of the dataset for which the language direction was inverted, as well as a corpus of video game translation described in Hansen and Houlmont (2022), due to both is size and quality, but also its closeness to the fantasy genre.

	Segments	Tokens EN	Tokens FR
Europarl	2,007,723	49,867,465	54,553,979
Video Game	1,370 431	21,041,902	22,804,380
TED	410,443	7,041,745	7,464,033
GlobalVoices	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

Table 1: Corpora used for generic training.

	Segments	Tokens EN	Tokens FR
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. (train)	37,348	550,536	541,779
Sep. (val.)	7,225	109,859	106,621
Sep. (test)	704	10,181	10,073
Total	594,887	22,595,451	23,723,640

Table 2: Corpora used for fine-tuning.

The fine-tuning process was carried out using our custom corpus of 6 novels, divided into a training set that includes the first 5 volumes and a validation set made up of the sixth tome, with the exception of the last three chapters which form our test set (all labelled "Sep." in Table 2). We trained a first model using only this data for comparison purposes, then trained a second model with a larger literary corpus, as in Toral and Way (2018). We therefore compiled a synthetic corpus ("Synthetic") in the manner of Caswell et al. (2019), made up of 150 books originally written in French and specifically selected from different genres, time periods and French-speaking countries. All were translated with DeepL to create the English source text. We also built a parallel corpus of 35 ebooks ("Parallel") chosen from fantasy classics and aligned with *LogiTerm* (Terminotix Inc., 2018). Given that our aim is to adapt a system to a specific translator, we did not include more than two or three works per author or translator. Finally, we built a second parallel corpus ("Translator") containing 40 translations from Nathalie Serval, who has worked extensively in the *fantastique*, fantasy and science-fiction domains.

Both of our systems were trained (200,000 steps) and fine-tuned (50,000 steps) with OpenNMT v2 (Klein et al., 2017), using the base parameters of the Transformer model (Vaswani et al., 2017) and a vocabulary size of 16K, set with SentencePiece's unigram encoding (Kudo and Richardson, 2018).

4 Automatic Evaluations

In an attempt to avoid common pitfalls of MT evaluation (Marie et al., 2021) and in order to provide detailed insight into the output of our custom system, we show in this section a combination of traditional and more recent metrics, along with their signature and a statistical significance test. These are followed by other automatic measures reflecting problems of MT which are of particular interest to the literary domain, namely lexical richness (Vanmassenhove et al., 2019), length ratio (Castilho and Resende, 2022), and structure being too close to the source text (Ahrenberg, 2017), that reflect the simplification, explicitation and convergence universals respectively (Baker, 1996). We then add to this analysis with a more qualitative evaluation of the resulting text in the next section.

4.1 Evaluation Metrics

The easiest way to confirm the improvement that results from the adaptation process is, of course, to lean on automatic evaluation metrics for MT. To that end, we report three metrics provided by SacreBLEU (Post, 2018), along with their signature,[1] as well as the more recent COMET metric, that is said to better correlate with human judgement (Rei et al., 2020).[2] Although it is now widely accepted that BLEU (Papineni et al., 2002) is not representative of quality, it is of interest to us here in that we are hoping to see if our translation is closer to the one produced by the official translator, and in that BLEU actually measures the closeness to this one reference (by comparing n-grams). ChrF2++ (Popović, 2017) relies in the same principle but also takes strings of characters into account on top of word n-grams and has shown to better correlate with human evaluation, while TER (Snover et al., 2006) rather gives an idea of the time and effort needed to edit the MT output until it matches the reference exactly. Lastly, COMET (Rei et al., 2020) is supposed to measure semantic similarity by comparing the vector representations of a translation in relation to the representations of both the reference and source texts.

[1]Metric signatures for sacreBLEU:
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.

[2]We ran the evaluation with the wmt20-comet-da model, and apply a bootstrapped t-test (Koehn, 2004) to confirm that the difference for our tuned model is statistically significant with a 95% confidence interval.

	BLEU ⇑	chrF2++ ⇑	TER ⇓	COMET ⇑
Google	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
Sep-only	18.56	40.43	76.06	-0.126
Sep-large	19.07	41.43	75.98	-0.066

Table 3: MT metrics for two public systems, in-house generic baseline, and our two literary-adapted systems.

We use multiple metrics to see if all point in the same direction (an arrow indicates for each if better performance is marked by a higher or lower score). To have some sort of a basis for comparison, we first show the scores of two generic online systems, Google Translate and DeepL, in Table 3.[3] Reporting scores for these can be of interest, seeing as they are the MT systems that the general public might be most familiar with for this language combination, but also because we assume they are much more robust than our own generic system (labelled as such in the table) as it is trained with only the 4M sentences given in Table 1.[4] The results for all three are quite low if we look at the absolute BLEU points, which we take as another confirmation that MT appears to have some difficulties with this particular literary genre—or is at least very distant from the official reference.

Still looking at BLEU, our first custom system ("Septimus-only") reaches almost twice the performance of our generic model with only the 45K segments of our initial corpus. The added 550K segments of the second model ("Septimus-large"), which was tuned on all of the corpora listed in Table 2, only improved BLEU by about 0.5 points. The translation output of these systems is very different, however, as somewhat reflected in the COMET score. We believe this is because the use of a language model favours systems that score higher on fluency, whereas our first model shows a higher degree of adaptation to translator style, as we note in our observations (cf. *infra*).

Yet, all metrics attest to the interest of the adaptation procedure, which almost doubles the BLEU score. This is much more than in other successful literary adaptation experiments (Matusov, 2019), and in keeping with the conclusion that translator-specific corpora are more useful than larger and general literary corpora (Kuzman et al., 2019).

4.2 Lexical Richness

Various studies have looked at other, more specific phenomenons similar to lexical richness to better understand how MT outputs differ from HT. As a result, it has been found that texts produced by MT are lexically less diverse than human translations, and the same observation held evidently true with literary works, adding to the importance of such features in addition to traditional quality metrics (Webster et al., 2020).

To verify whether or not the adaptation of our system also showed improvements with regard to textual features such as lexical richness, we used the LexicalRichness python module[5] to look at the Type-Token Ratio (TTR) of our reference ("REF") and compared it to our "Septimus-large" model ("SEP") as well as the translations produced by Google Translate ("GGL") and by DeepL ("DPL"). Results are reported in Table 4. Since the TTR is skewed by sentence length, we also computed the Mean Segment TTR (MSTTR) (Johnson, 1944) and the Moving Average TTR (MATTR) (Covington and McFall, 2010), as they seem to be better indicators of TTR (Tezcan et al., 2019). All texts were lemmatized with Stanza (Qi et al., 2020) and we used a suggested window of 500 tokens for the MSTTR and MATTR measures.

	Tokens	Types	TTR	MSTTR	MATTR
REF	10,348	1,900	0.18	0.45	0.45
SEP	9,601	1,598	0.17	0.42	0.42
GGL	11,351	1,699	0.15	0.41	0.41
DPL	11,423	1,688	0.15	0.40	0.40

Table 4: Measures of lexical richness for the human reference, our bespoke MT system, Google and DeepL.

Unsurprisingly, the human reference exhibits the highest degree of lexical richness across the board, including the simple TTR that sometimes attributes a higher score to MT in some studies (cf. Tezcan et al., 2019) due to the abundance of mistranslations. All three methods also confirm an increase for our custom model in that respect, even though one might notice the considerable gap in the overall number of tokens between the same model and the reference or, even more so, the two generic systems. This is due to the number of omissions generated by our system and will be touched in the manual evaluation section (cf. *infra*).

[3] Publicly available systems tested on 25/11/2020.

[4] As a note, we tried running experiments with the dozens of millions of sentences from the WMT 2014 translation task (Bojar et al., 2014), but the quality as indicated by automatic and human evaluations dropped systematically, due to—in our opinion—the noise and dissimilarity of these datasets.

[5] `https://github.com/LSYS/LexicalRichness`.

4.3 Segment length

Although the sentence length of human translations tends to be longer than the original,[6] Toral (2019) has observed that MT produced shorter sentences, thus hinting at a possible interference of the source text in that regard. In order to see if the same held true with our literary text, we indicate in Table 5 the average length of all translations in addition to the source ("SRC"), obtained by dividing the number of tokens with the total number of segments (704), as well as the expanding ratio, simply calculated by the difference in length of the source and target texts, normalized by the length of the source.

	Average length	Expanding ratio
SRC	14.26	—
REF	14.70	+3.04
SEP	13.64	-4.40
GGL	16.12	+13.03
DPL	16.22	+13.75

Table 5: Average segment length and expanding ratio for the reference, bespoke system, Google and DeepL.

While our test set contains a lot of very small segments (e.g. "Marcia smiled.", "She continued."), these appear amongst segments that very often contain two or more sentences (up to five) and that can be subject to major syntactic reorganizations. Despite the manual alignment of this whole custom corpus to ensure that each translation unit is as small as possible, while having the exact same content, this scenario happens rather frequently, which is why we prefer to speak of segments instead of sentences. Still, it appears overall that the average segment length is quite similar between the source and the reference, which we can explain by the fact that the translator makes ample use of omissions, contractions and shortcuts to produce a text that has approximately the same number of pages as the original novel. However, both public MT tools display a much higher ratio, while the translation of our custom system generates segments that are even shorter than the source, echoing our last observation regarding lexical richness.

4.4 Syntactic Equivalence

After looking at segments in terms of length and richness, we investigate what might be perhaps the biggest challenge for MT in the literary field, which would be its tendency to follow the syntactic structure of the source text too closely. To do so, we utilize another readily available python module, ASTrED,[7] in its fully automated mode. Based on word-aligned and POS-tagged sentences,[8] ASTrED offers various metrics aimed at quantifying syntactic equivalence and assessing translation difficulty (Vanroy et al., 2021). Given that sentences have to be compared individually and that our test set contains many sentences that do not have a one-to-one correspondence, we have removed all segments containing more than one sentence in any of the four compared translations and were eventually able to evaluate 404 sentences out of the 704 total.

	SACr	POS changes	ASTrED
REF	1728	1116	3674
SEP	891	975	2683
GGL	492	969	2294
DPL	558	974	2494

Table 6: Measures of syntactic equivalence for the reference, bespoke system, Google and DeepL.

The first measure suggested by ASTrED is the number of Syntactically Aware Crosses (SACr), meaning that words are first grouped together in linguistically motivated syntagmas, of which the alignment between source and target sentence is then compared to determine how much reorganizing (alignment crosses) has taken place during translation. The number of changes in Part-of-Speech (POS) tags for each of these aligned pairs can also be used as an indicator of syntactic equivalence, while the third Aligned Syntactic Tree Edit Distance (ASTrED) aims at identifying deeper structural differences by contrasting and evaluating the distance between dependency parse trees of the source and target sentence while still keeping the alignment into account.[9]

The number of SACr once again ranks HT first, with the custom system showing heavier reorganization when compared to generic systems. And even though this is less visible in terms of POS changes, the ASTrED metric seems to confirm this trend, both supporting the assumption that relying on specific and quality data can really improve MT, including when it comes to literary texts, and showing that HT remains well ahead when other textual features are taken into account.

[6] Usually 20–25% longer for translations from English into French according to De Clercq et al. (2021).

[7] https://github.com/BramVanroy/astred.

[8] The alignment is carried out by AwesomeAlign (Dou and Neubig, 2021). POS-tagging is done once again with Stanza.

[9] See Vanroy et al. (2021) for deeper concrete examples and further details on each measure.

5 Manual Evaluation

All of the automated evaluations presented above thus seem to corroborate the positive influence of our adaptation experiment. To further substantiate and enlarge on these results, we conducted an error annotation and general evaluation of the output.

The annotation was performed by two annotators (the authors, who are also translators) in *Accolé* (Esperança-Rodier and Brunet-Manquat, 2021). The typology used for this study inspired by Vilar et al. (2006), Tezcan et al. (2019), and Schumacher (2020), but was set up with three aims in mind: it adopts the translator rather than the MT engineer's point of view, it specifically takes literary aspects into consideration, and it aims at minimizing the number of categories while maximizing ease of use.

A preliminary evaluation was carried out on one chapter (30%) of our test set and then discussed between the two annotators to simplify the annotation scheme and remove possible ambiguities. Then, another 10% were re-annotated with the final scheme to assess the inter-annotator agreement and therefore the relevance of our typology. We present here the subsequent findings of one annotator over the complete test set. Readers can refer to Appendix A for an overview and description of the whole typology, and we provide a few examples in Appendix B to illustrate the following section.

5.1 Error Annotation

The observed agreement for the first tenth of the test set was 80.85% for adequacy, 84% for fluency and 70.49% for the literary aspects, revealing more variance for this last point despite the otherwise high agreement.

The annotation process was quite aggressive, in the sense that repeating errors were systematically annotated and that one issue could receive multiple labels, as with "during a Do-or-Die exercise in the Young Army" > "*lors d'un exercice de désœuvrement de la Jeune Garde*", where a specific term was lost (LOS) due to a mistranslation (MTR) and resulted in a nonsensical (NON) segment. Overall, there are 145 (20%) "error-free segments", which is much less than the 44% reported in Fonteyne et al. (2020) with Dutch, but closer to the 20% and 28% indicated by Matusov (2019) for Russian and German.

In total, we count 617 adequacy errors, 251 for fluency, and 382 regarding literary aspects. This confirms the common observation that, while being very fluent, NMT produces more errors related to content and meaning (Loock, 2019), which are typically harder to spot when correcting the output. Surprisingly, there were a lot fewer errors than we expected when it comes to literary aspects.

The biggest issues, by far, are omissions (OMI), mistranslations (MTR) and style (STY), as well as logic problems (LOG). (See examples 1, 2 and 3 in Appendix B.) This is consistent with other findings in the literary domain (Matusov, 2019; Fonteyne et al., 2020), but the number of omissions remains notable, as highlighted in previous sections. Many of these are not errors strictly speaking, and are equally made by the translator: out of 270, 61 omissions are shared, and only 12 are really problematic (see examples 4 and 5 for both cases). In addition to the necessary "stylistic omissions", this strategy can result from conscious and desirable translation choices (Dimitriu, 2004) that our system seems to have learned from, almost to an exaggerated point.

Adequacy										
Content				NMT Glitches		Vocabulary		Meaning		
OMI	ADD	OTR	UTR	STU	HAL	NTR	MTR	OME	NON	SME
270	6	8	38	3	12	0	200	11	26	43

Fluency									
Coherence			Agreement		Conjugation		Syntax		
REF	REL	LOG	GEN	NUM	TEN	PER	FUN	PUN	
31	3	60	19	6	51	7	54	20	

Literary Aspects								**Total**
Prose				Terminology		Typography		704 segments
STY	REG	PAR	ADA	COH	LOS	DIA	CAS	799 sentences
218	38	8	15	29	10	44	20	1,250 errors

Table 7: Number of errors annotated for adequacy, fluency and literary aspects.

Close to omissions, other adequacy issues include undertranslation (UTR), as well as a number of meaning errors (cf. examples 6 and 7). The latter mainly relate to shifts in meaning (SME), nonsense (NON) or—this is more problematic—opposite meanings (OME).

For fluency, the biggest issues besides logic are tenses (TEN) and function words (FUN), mainly (cf. examples 8 and 9). We expected to see a higher number of errors in most of the categories for fluency, especially when it came to coherence, agreement and conjugation. We assume that the adaptation process has helped a lot in that regard, in particular with tenses (in fact, our system seems to like using the French *passé simple* so much that it forgot using the *imparfait* when it should have). The same goes for referential cohesion, which is of notable importance for literary texts (Voigt and Jurafsky, 2012). It would therefore be interesting to compare this with the output of a generic MT system on the one hand, as well as with context-aware NMT (Voita et al., 2018) on the other.

Regarding literary aspects, style really is the principal difficulty, as we have already suggested. This means that some translations remain too literal and not very natural in French at times, even after the adaptation procedure. As with fluency, we also anticipated more errors for these categories, notably for register (REG) and dialogues (DIA) for instance. There remains ambiguities regarding the formal and informal translation of “you” in French (cf. example 10), or the use of a tone that does not fit a given character, but these are not too frequent. Likewise, dialogues must meet specific typographic conventions in French (colon in the preceding segment, quotes only in precise cases and em dash to designate a new speaker), but even if a few dashes are missing or wrongly introduced here and there, these are generally well observed, even when there are no indications of a dialogue in the English segment (cf. examples 11, 12 and 13).

5.2 Observations

While we used our best model (“Septimus-large”) for all of these evaluations, since it showed the best performance on all metrics, one important observation is that in comparison with our preliminary evaluation on the output of an older system, there seemed to be less obvious signs of adaptation. This system, like our “Septimus-only” model, was only tuned on the six novels translated by Nathalie Serval and did not use SentencePiece. With subword segmentation and more literary data, there were no longer any untranslated words (NTR), more syntactic reorganization and fewer grammar errors, but much also fewer exact matches with the reference. This is something that is not revealed by quality metrics, and also something that we will have to remedy in the long run.

But how can we say that the system has adapted nonetheless? In addition to all the metrics pointing that way, the biggest indicator is that virtually all of the vocabulary specific to the series is maintained, but also that there is a lot of heavy syntactic reorganizations that are in line with the translator. Terms are not systematically translated the same way and are sometimes rendered by synonyms or periphrases also used in the reference (e.g. “Wizard Tower courtyard” > “*cour de la tour du Magicien*”, “*cour du Magicien*”, “*cour de la tour*”). Likewise, translation choices are closer overall to those of the translator, whether it be 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 (e.g. “looked in panic” > “*regardèrent, paniqués*”), contractions (e.g. “were beginning to fall; they drifted lazily” > “*tombaient paresseusement*”)... Many omissions, interestingly, are made by the translator as well (entire clauses, tag answers, discourse information and speaker indications, adjectives in long nominal groups, repetitions...), and, equally surprising, the MT system usually merges segments when the translator similarly does so (33 in both, 6 in HT only, 2 in MT only), contradicting the common idea that MT can only work with one-to-one translation of sentences.

However, we can generally see that translations remain quite literal despite the adaptation. 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. “the Things” > “*les serviteurs de Merrin*” instead of “*les créatures*”), and although our custom MT learned to merge sentences, splits are more frequent in the reference (18 in HT only, 8 in MT only, 4 in both). Finally, it appears that colloquial and spoken language create frequent disambiguation mistakes, while chapter names (2 out of the 3) are also problematic, as was the spoiler that we found!

The code and scripts used for this project will eventually be made available on https://gitlab.uliege.be/dhansen/literary-machine-translation.

6 Discussion

In this study, we have tackled the challenge of adapting an MT system not simply on literary data, but on a much more specific corpus of individual translations and a subset of fantasy fiction novels. Even though we still face the issue of finding the ideal balance between translator style and overall quality, current results from automatic metrics have proven the output to be both lexically richer, while remaining very coherent with the series's universe, and syntactically more varied than generic systems.

What is more, the annotation and comparison with the reference reveals that our custom system has also adapted to the translator's individual style. Indeed, the main and most significant finding of our analysis, which supports the result of the previous automated measures, is that having a system trained on stylistically rich and coherent literary data yields substantial progress in the adaptation process, but also on a range of criteria that are all cited as limitations of MT and major differences with HT: reordering, sentence splits (to which we add merging of sentences), syntactic shifts, explicitation, modulation and paraphrasing (Ahrenberg, 2017).

Of course, the raw output is still very far from being on par with human translation. This appears clearly in the error annotation process, but the results also indicate that most issues fall within typical and well-known errors of MT across domains, rather than purely literary considerations. This points to the idea that MT might eventually become a useful tool, capable of providing adequate suggestions on-the-fly to literary translators. This scenario, however, could be dependent on further improvements to machine translation (regarding context aware MT for example) as well as careful implementation of this tool into the workflow (consider, for instance, having MT suggestions appearing outside the editing area in a CAT tool, against having the text pre-translated in the same interface, or having isolated sentences appearing in a bare and rudimentary PE tool, or even having no tool at all and post-editing the raw MT directly). Although further removed from our study, Läubli et al. (2022) show that text and user interface can have a direct effect on performance and affects.

6.1 A New Paradigm for LMT

In light of all this, we thus suggest a new paradigm for MT that could prove beneficial even outside the literary field, where tools are adapted on a deeper, more human level, and are seen more as an interactive aid rather than a pre-processing step. Our approach thus actively rejects technological determinism, as defined by Ruffo (2018), "whereby technology acts as a subject in shaping society and culture. On the contrary, humans regain their active role of agents in determining, accepting, rejecting and interpreting technological artefacts." Having a system that is tailored to individual human productions could furthermore play an important role in the emotional response and therefore acceptance of such a tool, if we keep in mind that translators are not against translation technologies *per se*, but rather against the tools that do not account for the specific challenges of their work and the "human aspects" of it (*Ibid.*; Koskinen and Ruokonen, 2017; Daems, 2021).

Personal MT systems might additionally represent the only way to mitigate the ethical challenges that arise with translation technologies. The most evident issue would be the loss of voice and stylistic normalization otherwise induced by MT (Kenny and Winters, 2020), but also the creativity constraints (Guerberof-Arenas and Toral, 2022), as well as work conditions, remuneration or property rights (Taivalkoski-Shilov, 2019). And while we could be tempted to think that LMT would not become a reality before a number of years, companies have already started offering literary post-editing services (Macken et al., 2022), and unrevised automated translations of literary works are not uncommon on the Web.

Having tools and data that remain entirely in the hands of translators could help to address some of those legal concerns, especially now that initiatives like OPUS-CAT aim at facilitating the integration and fine-tuning of MT. But as with interactive and adaptive MT, there still is some way to go. With better systems, dedicated translation environments, proper PE experience and MT literacy in addition to strong literary translation skills, human-adapted LMT might provide a means to support rather than lessen creativity and quality. What is certain is that it would allow for a more human-centred use of MT, and perhaps a way to re-empower translators.

6.2 Limitation and Future Work

One limitation and challenge of our work remains, as discussed, in preserving the delicate balance between robustness, or having a system that produces grammatical and fluent sentences, and having a system that adapts more closely to translator style. As a continuation to this study, we plan to gather the views and opinions of the most interested parties and confront the output of systems adapted on various domains to their respective translators.

References

Lars Ahrenberg. 2017. Comparing Machine Translation and Human Translation: A Case Study. In *Proceedings of the First Workshop on Human-Informed Translation and Interpreting Technology*, pages 21–28, Varna, Bulgaria. ACL.

Mona Baker. 1996. Corpus-Based Translation Studies: The Challenges That Lie Ahead. In Harold Somers, editor, *Terminology, LSP and Translation: Studies in Language Engineering*, pages 175–186. John Benjamins, Amsterdam, Netherlands.

Laurent Besacier. 2014. Traduction automatisée d'une oeuvre littéraire : une étude pilote. In *Actes de la 21e conférence sur le Traitement Automatique des Langues Naturelles*, pages 389–394, Marseille, France. ATALA.

Ondřej Bojar, 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.

Sheila Castilho and Natália Resende. 2022. Post-Editese in Literary Translations. *Information*, 13(2):1–22.

Isaac Caswell, Ciprian Chelba, and David Grangier. 2019. Tagged Back-Translation. In *Proceedings of the Fourth Conference on Machine Translation*, volume 1, pages 53–63, Florence, Italy. ACL.

Michael A. Covington and Joe D. McFall. 2010. Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2):94–100.

Joke Daems. 2021. Wat denken literaire vertalers echt over technologie? *WEBFILTER*, 24 January.

Orphée De Clercq, Gert De Sutter, Rudy Loock, Bart Cappelle, and Koen Plevoets. 2021. Uncovering Machine Translationese Using Corpus Analysis Techniques to Distinguish between Original and Machine-Translated French. *Translation Quarterly*, 101:21–45.

Rodica Dimitriu. 2004. Omission in translation. *Perspectives*, 12(3):163–175.

Zi-Yi Dou and Graham Neubig. 2021. Word Alignment by Fine-tuning Embeddings on Parallel Corpora. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2112–2128. ACL.

Emmanuelle Esperança-Rodier and Francis Brunet-Manquat. 2021. ACCOLÉ : Annotation Collaborative d'erreurs de traduction pour COrpus aLignÉs, multi-cibles, et Annotation d'Expressions Polylexicales. In *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles*, volume 3, pages 1–5. ATALA.

Margot Fonteyne, Arda Tezcan, and Lieven Macken. 2020. Literary Machine Translation under the Magnifying Glass: Assessing the Quality of an NMT-Translated Detective Novel on Document Level. In *Proceedings of the 12th Conference on Language Resources and Evaluation*, pages 3790–3798, Marseille, France. ELRA.

Ana Guerberof-Arenas and Antonio Toral. 2020. The Impact of Post-Editing and Machine Translation on Creativity and Reading Experience. *Translation Spaces*, 9(2):255–282.

Ana Guerberof-Arenas and Antonio Toral. 2022. Creativity in Translation: Machine Translation as a Constraint for Literary Texts. *Translation Spaces*.

Damien Hansen. 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.

Damien Hansen, Emmanuelle Esperança-Rodier, Hervé Blanchon, and Valérie Bada. 2022. La traduction littéraire automatique : Adapter la machine à la traduction humaine individualisée. *Journal of Data Mining & Digital Humanities*, pages 1–19.

Damien Hansen and Pierre-Yves Houlmont. 2022. A Snapshot into the Possibility of Video Game Machine Translation. In *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas*, volume 2, pages 257–269, Orlando, FL. AMTA.

Wendell Johnson. 1944. Studies in Language Behavior: A Program of Research. *Psychological Monographs*, 56(2):1–15.

Dorothy Kenny and Marion Winters. 2020. Machine translation, ethics and the literary translator's voice. *Translation Spaces*, 9(1):123–149.

Guillaume Klein, 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.

Philipp Koehn. 2004. Statistical Significance Tests for Machine Translation Evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 388–395, Barcelona, Spain. ACL.

Kaisa Koskinen and Minna Ruokonen. 2017. Corpus-Based Translation Studies: The Challenges That Lie Ahead. In Dorothy Kenny, editor, *Human Issues in Translation Technology*, pages 8–24. Routledge, Abingdon, United Kingdom.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Bruxelles. ACL.

Taja Kuzman, Špela Vintar, and Mihael Arčan. 2019. Neural Machine Translation of Literary Texts from English to Slovene. In *Proceedings of the Qualities of Literary Machine Translation*, pages 1–9, Dublin, Ireland. EAMT.

Rudy Loock. 2019. La plus-value de la biotraduction face à la machine. *Traduire*, 241:54–65.

Samuel Läubli, Patrick Simianer, Joern Wuebker, Geza Kovacs, Rico Sennrich, and Spence Green. 2022. The impact of text presentation on translator performance. *Target : international journal of translation studies*, 34(2):309–342.

Lieve Macken, Bram Vanroy, Luca Desmet, and Arda Tezcan. 2022. Literary Translation as a Three-Stage Process: Machine Translation, Post-Editing and Revision. In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pages 101–110, Ghent, Belgium. EAMT.

Benjamin Marie, 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*, volume 1, pages 7297–7306. ACL.

Evgeny Matusov. 2019. The Challenges of Using Neural Machine Translation for Literature. In *Proceedings of the Qualities of Literary Machine Translation*, pages 10–19, Dublin, Ireland. EAMT.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, PA. ACL.

Maja Popović. 2017. ChrF++: Words Helping Character N-grams. In *Proceedings of the Second Conference on Machine Translation*, pages 612–618, Copenhagen, Denmark. ACL.

Matt Post. 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.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108. ACL.

Ricardo Rei, Craig Stewart, Ana Farinha, and Alon Lavie. 2020. COMET: A Neural Framework for MT Evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 2685–2702. ACL.

Paola Ruffo. 2018. Human-Computer Interaction in Translation: Literary Translators on Technology and Their Roles. In *Proceedings of the 40th Conference Translating and the Computer*, pages 127–131. Éditions Tradulex.

Angie Sage. 2015–2013. *Septimus Heap*. 7 vols. HarperCollins, New York, NY.

Mehmet Şahin and Sabri Gürses. 2019. Would MT Kill Creativity in Literary Retranslation? In *Proceedings of the Qualities of Literary Machine Translation*, pages 26–34, Dublin, Ireland. EAMT.

Perrine Schumacher. 2020. Post-édition et traduction humaine en contexte académique : une étude empirique. *Transletters*, 3:239–274.

Nathalie Serval. 2015–2013. *Magyk*. 6 vols. Albin Michel, Paris, France.

Mathhew Snover, Bonnie Dorr, Rich Schwartz, Linnea Micciulla, and John Makhoul. 2006. A Study of Translation Edit Rate with Targeted Human Annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, pages 223–231, Cambridge, MA. AMTA.

Kristiina Taivalkoski-Shilov. 2019. Ethical Issues Regarding Machine(-Assisted) Translation of Literary Texts. *Perspectives: Studies in Translation Theory and Practice*, 27(5):689–703.

Terminotix Inc. 2018. *LogiTerm*. V. 5.8.0.18042.

Arda Tezcan, Joke Daems, and Macken Lieve. 2019. When a 'Sport' Is a Person and Other Issues for NMT of Novels. In *Proceedings of the Qualities of Literary Machine Translation*, pages 40–49, Dublin, Ireland. EAMT.

Jörg Tiedemann. 2012. Parallel Data, Tools and Interfaces in OPUS. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2214–2218, Istanbul, Turkey. ELRA.

Antonio Toral. 2019. Post-editese: an Exacerbated Translationese. In *Proceedings of Machine Translation Summit XVII*, volume 1, pages 273–281, Dublin, Ireland. EAMT.

Antonio Toral and Andy Way. 2015. Translating Literary Text between Related Languages using SMT. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 123–132, Denver, CO. ACL.

Antonio Toral and Andy Way. 2018. What Level of Quality can Neural Machine Translation Attain on Literary Text? In Joss Moorkens, Sheila Castilho, Federico Gaspari, and Stephen Doherty, editors, *Translation Quality Assessment: From Principles to Practice*, pages 263–287. Springer, New York, NY.

Eva Vanmassenhove, Dimitar Shterionov, and Andy Way. 2019. Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation. In *Proceedings of Machine Translation Summit XVII*, volume 1, pages 222–232, Dublin, Ireland. EAMT.

Bram Vanroy, Orphée De Clercq, Arda Tezcan, Joke Daems, and Lieve Macken. 2021. Metrics of Syntactic Equivalence to Assess Translation Difficulty. In Michael Carl, editor, *Explorations in Empirical Translation Process Research*, Machine Translation: Technologies and Applications, pages 259–294. Springer, Cham, Switzerland.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6000–6010, Long Beach, CA. Curran Associates Inc.

David Vilar, Jia Xu, Luis Fernando D'Haro, and Hermann Ney. 2006. Error Analysis of Statistical Machine Translation Output. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, pages 697–702, Genoa, Italy. ELRA.

Rob Voigt and Dan Jurafsky. 2012. Towards a Literary Machine Translation: The Role of Referential Cohesion. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, pages 18–25, Montreal, Canada. ACL.

Elena Voita, Pavel Serdyukov, Rico Sennrich, and Ivan Titov. 2018. Context-aware neural machine translation learns anaphora resolution. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistic*, volume 1, pages 1264–1274, Melbourne, Australia. ACL.

Rebecca Webster, Margot Fonteyne, Arda Tezcan, Lieve Macken, and Joke Daems. 2020. Gutenberg Goes Neural: Comparing Features of Dutch Human Translations with Raw Neural Machine Translation Outputs in a Corpus of English Literary Classics. *Informatics*, 7(3):1–21.

A Error annotation scheme

Omission (OMI) 21.6% of all errors
Word or idea that is present in the source but not in the target (even if also omitted in the reference).

Addition (ADD) 0.5% of all errors
Word or idea that is present in the target but not in the source.

Overtranslation (OTR) 0.6% of all errors
Translation that is correct but that leads to a redundancy or adds a nuance in the target that was not present in the source.

Undertranslation (UTR) 3.0% of all errors
Translation that is correct but that leads to the loss of a nuance in the target.

Stuttering (STU) 0.2% of all errors
Words repeated for no apparent reason by the MT system.

Hallucination (HAL) 1.0% of all errors
Completely illogical fragments of sentence added or replaced in the target; invented terms due to subword segmentation.

Non-translation (NTR) 0.0% of all errors
Source term left untranslated in the target.

Mistranslation (MTR) 16.0% of all errors
Terms or syntagmas wrongly translated or translated with the wrong sense.

Opposite meaning (OME) 0.9% of all errors
Translation leading to a meaning that is contradictory to the source.

Nonsense (NON) 2.0% of all errors
Translation that does not make any sense.

Shift in meaning (SME) 3.4% of all errors
Translation leading to a different meaning than the one expressed in the source.

Referential cohesion (REF) 2.5% of all errors
Break in the logical relation of co-referring items (e.g. anaphora): pronoun resolution, lack of antecedent, lexical choice...

Relational cohesion (REL) 0.2% of all errors
Break in the logical articulation and flow of a text's sentences or clauses.

Logic (LOG) 4.8% of all errors
Sentence that is grammatically correct but syntactically or semantically inaccurate based on the source sentence or story.

Gender (GEN) 1.5% of all errors
Any issue related to a grammatical or character gender (excluding pronoun resolution).

Number (NUM) 0.5% of all errors
Any issue related to agreement based on grammatical number.

Tense (TEN) 4.1% of all errors
Wrong tense; problem with the sequence of tenses.

Person (PER) 0.6% of all errors
Subject-verb agreement.

Function words (FUN) 4.3% of all errors
Mistranslation of a determiner, preposition, etc. (anything but content words).

Punctuation (PUN) 1.6% of all errors
Any punctuation issue, with the exception literary-specific typographic conventions (e.g. related to dialogues).

Style (STY) 17.4% of all errors
Literal translation, repetition, unnatural wording or collocation, translation that makes the text more difficult to understand...

Register (REG) 3.0% of all errors
Confusion between the French "*tu*" and "*vous*"; character speaking with an inappropriate tone.

Unfitting paraphrase (PAR) 0.6% of all errors
Term rendered by an equivalent syntagma but leading to a ponderous or poor translation in the given context.

Adaptation (ADA) 1.2% of all errors
Text fragment that requires a particular translation solution due to language differences, cultural context, formal constraints...

Coherence with previous volumes (COH) 2.3% of all errors
Translation of a common or series-specific term that is not in line with previous volumes.

Loss (LOS) 0.8% of all errors
Term that is specific to the series's universe and translated with a flat and neutral instead of an original or obligatory solution.

Dialogues (DIA) 3.5% of all errors
Problem tied to the typographic conventions of the series.

Case (CAS) 1.6% of all errors
Translation going against previous choices regarding the capitalization of series-specific terms (always capitalized in English).

B Translation and error examples

Ex.	Source	DeepL (01/07/2022)	Custom MT	Reference
1	Foxy rushed over to the tiny sink and grabbed the pan out of Beetle's hands.	Foxy s'est précipité vers le petit évier et a pris la casserole des mains de Beetle.	Vulpin se précipita vers l'évier et agrippa la casserole des mains de Moustique.	Vulpin se rua vers l'évier et arracha la casserole des mains de Moustique.
2	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.	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.	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.	Le cabinet hermétique a maintes fois été exposé à la Ténèbre, et je crains que celle-ci n'ait fini par affecter la désignation.
3	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.	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.	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.	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.
4	Septimus read out the second number of the first pair. "Ninety-eight."	Septimus a lu le deuxième chiffre de la première paire. "Quatre-vingt-dix-huit."	— Quatre-vingt-dix-huit.	— Quatre-vingt-dix-huit.
5	A simple Reveal—why hadn't she thought of that?	Une simple révélation - pourquoi n'y avait-elle pas pensé ?	Pourquoi n'y avait-elle pas songé ?	Un sort révélateur... Comment n'y avait-elle pas pensé elle-même ?
6	The previous afternoon Marcia had successfully gathered all scribes together in the Manuscriptorium.	L'après-midi précédent, Marcia avait réussi à rassembler tous les scribes dans le manuscritoire.	La veille, Marcia avait réuni tous les scribes du Manuscriptorium.	La veille, Marcia était parvenue à réunir tous les scribes au Manuscriptorium.
7	Romilly walked selfconsciously into the seven-cornered passage.	Romilly a marché d'un air gêné dans le passage à sept angles.	Romilly pénétra inconsciemment dans le couloir qui tournait sept fois sur lui-même.	Romilly s'avança d'un pas timide vers l'entrée du couloir aux sept détours.
8	Simon pushed his way through the Darke Fog, terrified that at any moment a Thing would recognize him.	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.	Simon se fraya un chemin à travers le brouillard, terrifié à l'idée qu'une créature puisse le reconnaître.	Simon se frayait un chemin à travers le brouillard, terrifié à l'idée qu'une créature le reconnaisse.
9	He thought of Larry. Of Matt, Marcus and Igor at Gothyk Grotto, even the oddly irritating people at Wizard Sandwiches.	Il a pensé à Larry. À Matt, Marcus et Igor à la Grotte de Gothyk, et même aux gens bizarrement irritants de Wizard Sandwiches.	Il songea à Larry, à Matt, Marcus et Igor, à la Grotte-Gothic, même aux gens bizarrement agaçants de Magyk Sandwich.	Puis il pensa à Larry, à Matt, Marcus et Igor, de la Grotte-Gothic, et même à la bande d'hurluberlus parfois exaspérants de Magyk Sandwich.
10	"Marcellus, do you have six guineas on you?" asked Marcia.	"Marcellus, tu as six guinées sur toi ?" demande Marcia.	— Marcellus, tu as six guinées sur toi ?	— Marcellus, auriez-vous six guinées sur vous ? demanda Marcia.
11	"Does that mean that Beetle is..."	"Cela signifie-t-il que Beetle est...«	Ça veut dire que Moustique est...	— Ça veut dire que Moustique...
12	She continued. [cont.]	Elle a continué. [cont.]	Elle poursuivit : [cont.]	Marcia reprit : [cont.]
13	"During the last few days in the Wizard Tower, when I was trying to find a way to defeat the Darke Domaine [...]	"Pendant les derniers jours dans la Tour des Sorciers, quand j'essayais de trouver un moyen de vaincre le Domaine de Darke [...]	— Ces derniers jours, à la tour du Magicien, alors que je cherchais un moyen de vaincre le domaine ténébreux [...]	— Ces derniers jours, tandis que je cherchais comment vaincre le domaine ténébreux, enfermée dans la tour [...]