



The International Journal for
Translation & Interpreting
Research
trans-int.org

Translating science fiction in a CAT tool: machine translation and segmentation settings

Lucas Nunes Vieira
University of Bristol, United Kingdom
l.nunesvieira@bristol.ac.uk

Natalie Zelenka
University of Bristol, United Kingdom
natalie.zelenka@bristol.ac.uk

Roy Youdale
University of Bristol, United Kingdom
roy.youdale@bristol.ac.uk

Xiaochun Zhang
University College London, United Kingdom
xiaochun.zhang@ucl.ac.uk

Michael Carl
Kent State University, United States
mcarl6@kent.edu

DOI: 10.12807/ti.115201.2023.a11

Abstract: There is increasing interest in machine assistance for literary translation, but research on how computer-assisted translation (CAT) tools and machine translation (MT) combine in the translation of literature is still incipient, especially for non-European languages. This article presents two exploratory studies where English-to-Chinese translators used neural MT to translate science fiction short stories in Trados Studio. One of the studies compares post-editing with a ‘no MT’ condition. The other examines two ways of presenting the texts on screen for post-editing, namely by segmenting them into paragraphs or into sentences. We collected the data with the Quality plugin for Trados Studio and describe a method for analysing data collected with this plugin through the translation process research database of the Center for Research in Translation and Translation Technology (CRITT). While post-editing required less technical effort, we did not find MT to be appreciably timesaving. Paragraph segmentation was associated with less post-editing effort on average, though with high translator variability. We discuss the results in the light of broader concepts, such as status-quo bias, and call for more research on the different ways in which MT may assist literary translation, including its use for comparison purposes or, as mentioned by a participant, for ‘inspiration’.

Keywords: Literary translation; post-editing; machine translation; neural machine translation; computer-assisted translation; CAT tools; science fiction; Chinese translation

1. Introduction

There is increasing interest in machine translation (MT) as a potential literary tool. Previous research has shown that literary translators are faster when they

post-edit MT (e.g., Toral et al., 2018; Ó Murchú, 2019) even though they prefer translating without it (Moorkens et al., 2018). In professional contexts, MT is often edited in the environment of computer-assisted translation (CAT) tools, but the ways in which MT and CAT tools combine in literary translation remain under-researched. The CAT interfaces in which MT is edited matter for various reasons, not least because of the way in which they segment texts on screen, which can be constraining for literary tasks (Moorkens et al., 2018).

From a methodological perspective, data collection tools used in MT research can also be problematic in how they differ from professional editing environments. Empirical studies on translation technologies are often based on tools that are built for research, such as PET (Post-Editing Tool) (Aziz et al., 2012) or Translog-II (Carl, 2012). While these tools have the benefit of providing detailed logs of the translation process, their user interfaces are a factor to consider in a study's ability to replicate professional working environments. Reports of MT's time-saving potential have recently been criticised because of the often-oversimplified experimental conditions on which research in this area relies (do Carmo, 2020, p. 45). The importance of ecologically valid approaches has also been highlighted for literary tasks (Kenny & Winters, 2020), and although CAT tools may be more popular for non-literary specialisms (Verplaetse & Lambrechts, 2019, p. 16), literary translators use these tools too and find them helpful (e.g., Lombardino, 2015; Zakrajšek, 2020; see also Ruffo (2022), where 25% of a sample of 150 literary translators were CAT tool users).

In this article, we report on two studies where professional translators carried out a series of literary translation tasks in the familiar environment of a commercial CAT tool, namely Trados Studio 2019. We report on a method that allows data collected with the Qualityity plugin for Trados Studio (see Section 3.3) to be analysed through the CRITT Translation Process Research Database (TPR-DB) (Carl et al., 2016). The first study compares a translation (T) or 'no MT' condition with post-editing (P) of NMT (henceforth the 'T-P study'). The second study compares different ways of presenting the texts on screen for post-editing, namely with sentence and paragraph segmentation (henceforth the 'segmentation study'). Science fiction short stories were used in all tasks (see Section 3.1). Unlike most research on literary MT, where the focus is on European languages, translators across our two studies worked from English into Simplified Chinese. English word order and segmentation often need to change in Chinese (e.g., Meifang & Li, 2009), so this language pair lends itself well to our analysis of segmentation settings, whose effects may be particularly felt by English-Chinese translators.

We have a relatively small sample (see Section 4), so our approach is closer to a case study than to predictive hypothesis-testing. We also note that, although we examine the use of a CAT tool, translation memories are outside the scope of the article. We approach CAT from a broad perspective that includes MT as one of its key sources of assistance. To the best of our knowledge, this is the first time that translation process data generated in a commercial CAT tool is used for an empirical investigation of literary post-editing. It is hoped that our results will generate hypotheses, inform professional practice and stimulate future research on methodologies that may better reflect professional working environments.

We organise the remainder of the article as follows: in Section 2, we provide a brief literature review. In Section 3, we present our methodology, including details of the source texts, translators' profiles and the steps for converting and analysing the Qualityity data. We then present results of the two studies in Section 4, discuss the results in Section 5 and provide a summary of our findings, as well as future research directions in Section 6.

2. Research on computer-assisted literary translation

There have been several studies on *computer-assisted literary translation* (CALT), especially as to how literary translators can benefit from corpus tools and quantitative textual analyses (e.g., Youdale, 2020; Zanettin, 2017; Horenberg, 2019; Kolb & Miller, 2022). Research on the use and development of MT for literary texts is also proliferating (e.g., Hadley et al., 2019; Taivalkoski-Shilov, 2019; Toral & Way, 2015; Hansen, 2022). While this area of research is growing rapidly, this section is based predominantly on a small number of studies that looked specifically at the process of post-editing literary texts.

Evaluations of MT as a literary translation tool have so far largely supported MT's timesaving effect. A comparison between post-editing and a 'no MT' (or 'unaided') condition found that MT made the process of translating science fiction 31% faster in tasks from Scottish Gaelic to Irish (Ó Murchú, 2019, p. 24). Similarly, English-to-Catalan tasks showed increases of between 18% and 36% in number of words translated per time when MT was used (Toral et al., 2018). This same study showed reductions in keystrokes and cognitive effort for post-editing (ibid.). In an analysis of attitudes to literary post-editing based on the same data, translators reported preferring to translate without MT and felt that text segmentation had a negative effect on their work due to lack of context and difficulties with coherence and cohesion (Moorkens et al., 2018; 253). More recently, a similar study from English to both Dutch and Catalan presented the texts on screen with paragraph segmentation and found that unaided translation required more time and cognitive effort, although this did not always correspond to translators' perceptions (Guerberof-Arenas & Toral 2022). Şahin and Gürses (2021) asked student and professional translators to post-edit literary machine translations from English into Turkish. They analysed the post-edited texts and gathered further data through a survey and qualitative interviews. While their participants disliked working with MT, they mostly accepted the machine suggestions both lexically and syntactically (p. 197).

Kenny and Winters (2020) were interested in the translator's signature textual style (or 'voice') in literary post-editing. They presented the source text in Microsoft Word in a task where English-to-German translator Hans Christian Oeser post-edited a text that he had translated before. They found that his typical translation style was present to a lesser extent in the post-edited translations compared to his previous work. Macken et al. (2022) also analysed translation products. Based on English-to-Dutch tasks, they compared three versions of a literary text: its unedited machine translation, a corresponding post-edited version and a revision of the post-edited text produced by a second professional. Their analysis showed that most edits were performed in the revision rather than the post-editing phase (p., 109). The results were not conclusive as to whether this three-step workflow is preferable to human translation without MT followed by revision (ibid.).

As for the effect of segmentation, text presentation (sentence and document levels) has recently been examined by Läubli et al. (2022). Their experiments were based on controlled tasks such as copying the source text (p. 322) or revising manipulated human translations (p. 324) and did not involve MT or literary texts. Some literary translators may prefer to segment the text into paragraphs when working in CAT tools (Molines, 2020), but to our knowledge paragraph and sentence segmentations have not to date been compared for literary tasks. This comparison matters not least because sentence segmentation is the default setting in most CAT tools, which may prime translators' choices. For example, previous research has shown how consumers are heavily influenced by 'status quo bias' (Mandl et al., 2011) in adopting default

suggestions when making purchasing decisions. Similarly, a study about Microsoft Word found that more than 95% of the users consulted had not changed the program's default configurations (Spool, 2011).

As regards the post-edited quality of literary translations, earlier research between English and French suggests that texts post-edited by non-professionals were 'acceptable' (Besacier, 2014). Based on a score of narrative engagement, readers of literature have shown higher engagement with texts translated without MT than with post-edited translations, although this difference was not statistically significant (Guerberof-Arenas & Toral, 2020). In an assessment of creativity, operationalised as a mixture of novelty and acceptability, translations produced without MT were found to be more creative (Guerberof-Arenas & Toral, 2022).

Notably, most studies mentioned above were based on European languages. As mentioned in Section 1, text segmentation may be particularly problematic for English-Chinese translation and, by translators' own accounts, it negatively affects literary translation tasks (see above). We therefore explore important issues in literary MT use by presenting studies with Simplified Chinese carried out in a commercial CAT tool involving sentence and paragraph segmentation settings. Below we describe the methodology of both of our studies and describe a process that can be used in future research to analyse translation process data collected in Trados Studio.

3. Methodology

Below we provide details of the source texts (Section 3.1), the translators (Section 3.2) and the procedure for collecting and processing the data (Section 3.3). We make the dataset openly available together with the analysis code and the source texts used in the tasks (see Data Availability Statement).

3.1. Source texts

To share the material afterwards, the source texts needed to be licensed for translation and redistribution. We also needed to ensure that Chinese translations of the texts were not already available online to avoid priming the translators. Additionally, to reflect a realistic literary translation task, it was important for the texts to be challenging and call for translation solutions that deviated significantly from the source-text structure. To meet these criteria, we selected science fiction short stories by Canadian author Peter Watts (Watts, 2014a). The stories are available for non-commercial use under Creative Commons Licence CC BY-NC-SA 2.5¹ (Watts, 2014b).

Of the stories available under this licence, five candidate texts were initially selected. We ran excerpts starting from the beginning of the stories through the Coh-Metrix text analysis tool (McNamara et al., 2014). We compared the five excerpts based on a total of ten Coh-Metrix measures²

¹ See <https://creativecommons.org/licenses/by-nc-sa/2.5/>

² Namely, four lexical diversity measures (Type-token ratio (TTR) for content word lemmas; TTR for all words; Measure of Textual Lexical Diversity (MTLD); VOCD-D, a measure of vocabulary density), three word frequency measures based on the CELEX database (CELEX average frequency of content words, CELEX average frequency of all words, CELEX average minimum word frequency in sentences), average sentence length, Flesch Reading Ease and Coh-Metrix L2 Readability, an index designed for learners of English as a second language. We aggregate the lexical diversity and word frequency measures in Table 1 by taking the means. The formula for Flesch Reading Ease is "206.835 - (1.015 * sentence length) - (84.6 * word length)" (McNamara et al., 2014, p. 78). The one for Coh-Metrix L2 Readability is "45:032 + (52:230 x CRFCWO1

involving sentence length, lexical diversity, readability and word frequency, factors that have been shown to correlate with translation difficulty (e.g., Hvelplund, 2011; Sun & Shreve, 2014). In addition, two members of the team manually analysed the texts. One is an experienced literary translator and native speaker of English and the other a translation expert with native knowledge of Chinese. They identified idiomatic expressions as well as words and phrases which, in their assessment, stood out as likely to cause difficulty in English-to-Chinese translation. The difficult words or phrases were terms including slang or non-standard spelling (e.g., ‘the SanFran wireframe’, Text 1) or those which might not be found in dictionaries and/or had little surrounding context to suggest a particular interpretation (e.g., ‘paired lifting surfaces’, Text 3; ‘unleashed atmospheres’, Text 2). This manual analysis served to complement the objective measures. Counting idiomatic expressions (e.g., ‘out of whack’, Text 1; ‘plunked her down’, Text 3) also served to quantify ambiguity, a well-known negative factor in MT quality (e.g., Futral et al., 2022). Based on this procedure, we selected four texts that could be grouped by difficulty into two pairs. The texts are summarised in Table 1.

Table 1: Source texts selected for the investigation. Idiomatic expressions, difficult words and difficult phrases were identified in the manual expert analysis. The other metrics were sourced from Coh-Metrix.

	<i>INCORRUPTIBLE</i> (Text 1)	<i>ZeroS</i> (Text 2)	<i>The Colonel</i> (Text 3)	<i>Repeating the past</i> (Text 5)
Word count*	229	236	215	224
Sentence length (mean)	15.3	13.9	19.5	12.4
Lexical diversity (mean)	66.1	67.7	51	57
Word frequency (mean)	1.9	2.1	2.2	2.4
Flesch Reading Ease	59.1	75.9	61.9	76.4
Coh-Metrix L2 Readability	6.3	4.6	0.2	19.4
Idiomatic expressions	3	2	1	1
Difficult words	3	6	5	4
Difficult phrases	2	4	10	3

As shown in Table 1, some variables were similar for all texts while others varied slightly more. Coh-Metrix L2 Readability, for example, was lower for Text 3, which suggests that this text’s cohesion may pose a bigger challenge for second-language readers (McNamara et al., 2014, p. 81). Despite some of these differences, the texts could be grouped into two largely homogeneous pairs. Texts 1 and 3 (longer sentences, lower readability and more translation difficulties) were used in the T-P study. Texts 2 and 5 (shorter sentences, higher readability and fewer translation difficulties) were used in the segmentation study. The assignment of text pairs to each study was arbitrary. Our main concern was to ensure that texts used in the same study were comparable.

+ (61:306 x SYNSTRUT) + (22.205 x WRDFRQmc)” (p. 80), where CRFCWO1, SYNSTRUT and WRDFRQmc are measures of content word overlap, sentence-to-sentence syntax similarity and word frequency, respectively (McNamara et al., 2014).

3.2. Translators

Familiarity with Trados Studio and prior experience with literary translation were the main criteria for recruiting translators. Experience with MT post-editing was desirable. We posted an announcement on ProZ.com and used this forum directory to approach translators who advertised relevant experience and qualifications – for example, those who mentioned use of Trados Studio or ‘Art/Literary’ as a specialism. Our initial selection criteria were relatively narrow since we needed experienced literary translators who could all use a specific version of Trados Studio for the tasks (see Section 3.3). To recruit a larger sample, we accepted translators who did not have literature as their primary specialism as long as they had some literary experience. While this is a limitation of our method, relaxing this requirement was necessary to allow us to recruit translators who fit all criteria even if with a lower overall level of literary experience than initially desired. Translators in the authors’ networks were also considered provided they had a similar profile. Fifteen translators took part in the T-P study. Of these, eleven also took part in the segmentation study. Two translators were recruited through the authors’ contacts and the remainder were recruited through ProZ.com. All translators therefore either self-identified as professional online or were known by the authors to have paid translation experience. They all worked from English to Simplified Chinese. The project was managed from the United Kingdom (UK) and received ethical approval from the University of Bristol. Translators took part remotely (see Section 3.3.2). They were based in different countries including China, the UK and Australia.

Technical errors or other issues that could spuriously affect the results (see Section 3.3.2) forced us to exclude nine translators from the T-P analysis and one from the segmentation analysis. This means that the samples retained for analysis consisted of six translators in the T-P study and ten translators in the segmentation study. Table 2 provides a summary of their profile.

Table 2: Self-reported profile of translators retained for analysis in each study

	T-P Study (n = 6)	Segmentation Study (n = 10)
Mean professional experience in years	8.1 (range 4-11)	10.3 (range 2-26)
Previous literary experience		
1-10 projects	67%	60%
11-20 projects	0%	20%
Over 20 projects	33%	20%
Previous post-editing experience		
No experience	16.5%	20%
1-10 projects	16.5%	40%
Over 10 projects	67%	40%

3.3. Data collection and processing

3.3.1. *Qualitativity and the CRITT TPR database*

As mentioned, the data for the study was collected with the Qualitativity plugin for Trados Studio (henceforth, ‘Studio’). Like Studio itself, the plugin needs to be locally installed on the computer. Translators retained full control of any data records, which were not automatically shared with the researchers.³ Qualitativity

³ While from a research perspective automatic sharing may avoid technical errors, the plugin may also be used commercially, so the need for local installation protects

Translation & Interpreting Vol. 15 No. 1 (2023) 221

runs unobtrusively in the background while it logs text edits and translation time. After a translation session, the plugin generates reports that can be visualised in Studio or exported for further analysis.

Of interest for the studies presented here are the activity logs, referred to as ‘activities’, which can be exported from Qualityity. These records provide a time-stamped log of the translation process for each segment. They resemble the output of research-oriented tools such as Translog-II (Carl 2012) or PET (Aziz et al. 2012). Qualityity’s keyboard logging method is based on the `LowLevelKeyboardProc` callback function (Microsoft, 2018), which approximates keystrokes based on the modifications that take place in the editor. Keyboard inputs corresponding to applications outside of Studio (e.g., web browsers) are not recorded. The use of this function also means that only keystrokes corresponding to textual changes are logged, which is different from how research tools work. The plugin cannot record isolated ‘Ctrl + C’ operations, for instance, since in this case a string is copied to the clipboard without any changes in the text. For studies into Chinese or other languages with indirect or phonetic input methods, such as pinyin, Microsoft’s default keyboards need to be used for a more complete capture of the typing process. Some alternative methods (e.g., Sogou) require the input to be entered into an external IME (input method editor) window to which Qualityity does not have access. Qualityity’s keylogging method is therefore relatively non-invasive, which in commercial settings may be necessary to avoid the tool being mistakenly flagged as malware.

To generate CRITT TPR-DB features (see Carl et al., 2016) based on the Qualityity data, we developed a script that converts Qualityity’s XML output into an XML structure that is compatible with Translog-II. A ‘Trados’ flag can be selected when uploading the Qualityity data to the CRITT TPR-DB via the management interface. The ‘Trados’ setting then takes care of the conversion and generation of the TPR-DB feature tables. Eye tracking is now also supported and synchronised with the Qualityity data (Yamada 2022). After the Qualityity-to-TPR-DB conversion, all other TPR-DB features based on keystrokes and source-target textual alignments can be generated through the same procedure implemented for Translog-II or CASMACAT (Alabau et al., 2013). This allows variables that are not automatically available in the Qualityity output to be generated through the CRITT TPR-DB.

As in other studies analysed in the CRITT TPR-DB framework, the translations produced in our tasks were aligned with the source text at a word level in the YAWAT tool (Germann, 2008). This is a necessary step for generating the CRITT TPR-DB standard variables. Four MA students in Translation with native knowledge of Chinese aligned the content. The source-target correspondences were as granular as possible, with phrases kept as a single unit only when aligning individual components within a phrase was not possible or logical.

3.3.2. Study design and procedure

The Qualityity data was obtained remotely. Translators received the translation jobs as Studio packages and worked either in their own copy of Studio or in the trial version of the software, which did not impose restrictions on the tasks required. We provided task instructions in online forms that guided translators in a step-by-step fashion akin to a wizard. They used the wizard to download the Studio packages and to upload return packages. The data collection process

translators’ autonomy over their data. Commercial uses of the plugin are outside the scope of this article, but we argue that in any such case translators should be the judges of whether and how their activity is recorded and actively consent to any sharing of it (see Vieira et al., 2021).

for each study consisted of three stages. Each stage corresponded to a separate wizard.

The first stage was for setting up the plugin, testing the set-up and becoming familiarised with the data collection process. Details of translators' reported professional background and experience were also collected at this stage. Translators had to keep the option to log keystrokes selected as well as select an option to record the time spent reading segments that remained unedited.⁴ Figure 1 shows the part of the setting up wizard that specified the task requirements.

→ Please note that the study requires using **Trados Studio 2019 - Service Release 2** with **Microsoft Pinyin** as the keyboard input method (Sogou is *not* allowed). We will also ask you to complete each phase in a **single sitting**. This means that **once you start working on a document, pausing and coming back later is *not* allowed**. Select all options below to confirm you understand this.

NOTE: if you have recently installed the trial version of Trados Studio 2019 to take part in the study, the version you installed was most likely Service Release (SR) 1. If that's the case, please update your trial copy to SR2 by clicking on 'Help' and then 'Check for Updates'. *

Choose as many as you like

A I have Trados Studio 2019 SR2 installed ✓

B I'll use Microsoft Pinyin ✓

C I understand any documents need to be translated/edited in one sitting ✓

OK ✓ press Enter ↵

Figure 1. Section of the task wizard where translators had to confirm they had understood the task requirements

In the second and third wizards, translators received Studio packages that corresponded to the study conditions. The third wizard also included questions about translators' perceptions of the tasks and any comments they wished to make.

Translators were instructed to produce high-quality texts that would be suitable for publication. They also received background information about the source author and the stories. Figure 2 shows instructions provided for tasks where translators carried out post-editing.

⁴ Translators also had to set the timer to pause automatically after five inactive minutes, but the plugin code had an error (corrected in a subsequent version) that prevented the timer from being paused in this way. To avoid counting breaks as part of the translation process, we use the FDur variable from the CRITT output, which excludes pauses longer than 200 seconds (see Section 4). A similar variable that excludes interruptions is not available for pause duration metrics in the CRITT tables, so we do not provide an analysis of pause durations in this article.

The text you will be editing has been translated with machine translation software. Your job is to edit the translation so that it is suitable for publication and can be read by Chinese readers of Peter Watts's work.



Figure 2. Instructions on expected level of target quality provided in the post-editing task wizard

Version 2019 (Service Release 2) of Trados Studio was used throughout the investigation. We pre-translated the source texts using the RWS (formerly SDL) Language Cloud NMT system.⁵ In the segmentation study, we selected paragraph and sentence options by changing the translation memory settings.

To ensure consistency within the sample and higher levels of data quality, we asked translators to carry out each task in a single sitting and to use Microsoft Pinyin as their keyboard input method (see Figure 1). Translators could consult reference sources and search the internet as normal, but we asked them not to use any additional CAT resources such as their own translation memories. The packages they received contained just the source text and the MT output when it was part of the task condition.

The studies were realistic professional tasks, so in some respects they were inevitably difficult to control. By inspecting the data records, it became apparent that five translators pasted entire sentences into the Studio interface in the T-P study at the beginning of the translation process for each segment in the unaided condition. This suggests that these translators post-edited MT even when they were supposed to be translating without it. We therefore excluded these translators from the T-P analysis.⁶ A further four translators were excluded from the T-P study: two did not use Microsoft Pinyin and therefore had incomplete keylogging records (see Section 3.3.1); the other two failed to generate accurate data reports, in one case because of a technical error and in the other because the participant did a lot of the drafting outside of Studio and then pasted the content into the interface. In the segmentation study, we excluded one translator who manually segmented paragraphs into shorter segments. This was only changed on this occasion, so we excluded the data for consistency across the sample such that the paragraph condition corresponded to the same configuration for all translators throughout the task.

In both studies, translators worked in one condition first and then received the package for the second condition. We counterbalanced the assignment of text to condition and the texts' and conditions' order of presentation. All translators saw all conditions, though not all text-condition combinations since this would have required translating the same text twice. We examined the conditions in two separate studies, rather than in a single study with four conditions, because this simplified the data collection and therefore avoided errors. However, since we had to exclude participants from the T-P analysis, the resulting order in which translators saw the conditions became unbalanced in that study, with five of six translators post-editing last and one translator post-editing first. Given the exploratory nature of the analysis, and because of our efforts to ensure the use of comparable texts in each study, we do not consider

⁵ The texts were pre-translated in November 2019.

⁶ We also noted that a sixth participant pasted a longer passage into the interface in the unaided condition for one sentence. Since this was an isolated instance, it could also reflect the use of online reference materials, so we retained the data in this case.

this a cause for concern. In the segmentation study, translators saw the two conditions (paragraph vs. sentence) the same number of times as their first and last tasks.

3.3.3. *Variables of interest*

After converting the Quality data into the CRITT TPR-DB format, we generated and concatenated segment-level (SG) tables for analysis. By ‘segment’, we refer to strings of text as segmented in Studio. In the T-P study, these consisted of sentences or fragments as per default Studio segmentation rules. In the segmentation study, by contrast, segments consisted of sentences or fragments, when the default sentence segmentation setting was selected, and paragraphs, when the paragraph-level segmentation setting was selected. We used measures of task time, pauses and keystrokes to check how the task conditions influenced translators’ working processes. These three types of metrics can be regarded as proxies for temporal, cognitive and technical effort, respectively (Krings, 2001). Below we outline how we pre-processed the variables.

We used the FDur variable for the analysis of task time. This variable excludes typing pauses longer than 200 seconds (see Section 3.3.2). The segment-level temporal variables in the SG data are based on the interval between the first and last keystroke for a given segment. When a segment is not edited, FDur takes a value of zero. MT segments were left unedited in 3 instances in the T-P study (1.9% of the data) and 19 instances (9% of the data) in the segmentation study. Although reading time for unedited segments was recorded by the plugin, we excluded these data points from the temporal analysis since cases where FDur is zero do not correspond to the time translators spent on these segments but rather to the fact that the segments did not involve any keystrokes.

Typing pauses are a well-established proxy for the levels of cognitive effort required by translation and post-editing where the higher the number of pauses the higher the cognitive effort (Lacruz et al., 2014). We obtained the number of pauses from the SG tables by adding a constant of two to all values of the TB300 variable, which represents the number of typing bursts interspersed by 300-millisecond intervals. The number of typing bursts is equivalent to the number of pauses in addition to one initial and one final pause for each segment – i.e., pauses which, within a single segment, are not preceded or followed by any typing. While these pauses may in principle not occur if translators take less than 300 milliseconds to start editing the segment or if they move to a new segment within 300 milliseconds, previous research based on the 300-millisecond threshold suggests the presence of these pauses (Vieira, 2016). Adding the constant was therefore our preferred approach.⁷

Finally, to calculate the number of keystrokes, we added up the ‘Ins’ and ‘Del’ variables from the SG tables. These variables count the number of characters involved in insertions and deletions, respectively. We also consider the ‘Nedit’ variable to explore specific aspects of self-editing and how translators moved through the documents. This variable corresponds to the number of editing visits to each segment.

For simplicity, we refer to the variables described above as ‘translation time’ (FDur), which we express in seconds, ‘pauses’ (TB300+2), ‘keystrokes’ (Ins+Del) and ‘editing visits’ (Nedit).

Due to exclusions (see Section 3.3.2) and the fact that some of the data was collected at a paragraph rather than sentence level, the data available for

⁷ A constant of two overshoots the number of pauses per segment in one pause for cases where the MT output is left unedited. These cases were few and the difference in any case is a single pause, so we do not deem this to be problematic.

analysis has a small number of observations: 151 for the T-P study and 210 for the segmentation study. These size restrictions would inherently limit an inference about the population. We therefore take a more conservative approach and provide a descriptive exploratory analysis. We present results for the two studies separately below.

4. Results

4.1. Translation vs. post-editing

Before exploring differences between the post-editing and unaided conditions, we checked for potential correlations between our variables for translation time, pauses and keystrokes. There were very high correlations between pauses and keystrokes ($r = 0.93$) and between translation time and pauses ($r = 0.89$). The correlation between translation time and keystrokes was slightly lower ($r = 0.81$). Although we use all these variables in the analysis provided together with our dataset, to reduce redundancy we focus on keystrokes and translation time below given their lower correlation. Correlations between editing visits and the other variables were lower (between $r = 0.48$ and $r = 0.59$), so we also use editing visits to illustrate specific aspects of translators' working processes.

Across sentences and participants, there were more keystrokes per source character for the unaided condition (mean = 3.71, SD = 1.75) than for post-editing (mean = 2.25, SD = 1.35), which represents a 39.3% reduction on average. Based on the medians, the reduction in keystrokes for post-editing was 35.8%. In absolute terms, these keystroke counts are quite high. We checked the legacy English-to-Chinese news translation data of the CRITT TPR-DB (Carl et al., 2016) for comparison. We found that the RUC17 study had 1.13 keystrokes ('Ins+Del') per source character for unaided translation and 0.47 for post-editing. We hypothesise that the high counts we observe here are linked to the literary domain (and the difficulty of the texts) and the natural environment of the task, where translators were largely free to behave as they normally would in any professional commission.

Figure 3 illustrates the difference between unaided and post-editing conditions with boxplots for keystrokes and pauses per source character per translator. The boxplots show a clear difference between the two conditions with lower medians (middle horizontal lines in each box) for post-editing. This was the case for all participants in terms of keystrokes. In terms of pauses (right pane), P08 was the only translator who paused more often when using MT.

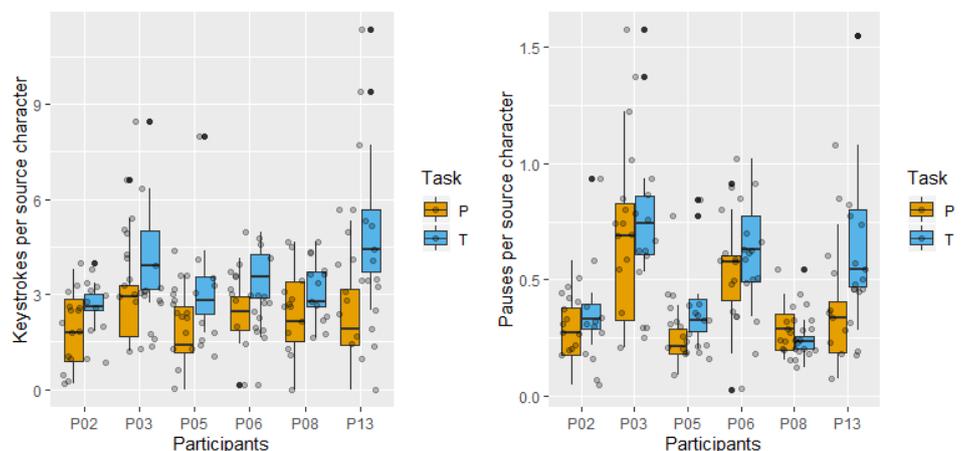


Figure 3. Boxplots per translator showing the keystroke (left) and pause (right) data distribution for post-editing (P, amber) and unaided translation (T, blue).

Figure 4 shows how the number of keystrokes changed in relation to sentence length, with non-parametric local weighted (loess) regression lines for post-editing (P, amber) and unaided translation (T, blue). There were consistently fewer keystrokes for post-editing irrespective of sentence length. This difference was slightly narrower for shorter sentences as can be seen by the widening gap between the lines along the x-axis (source segment length) in Figure 4.

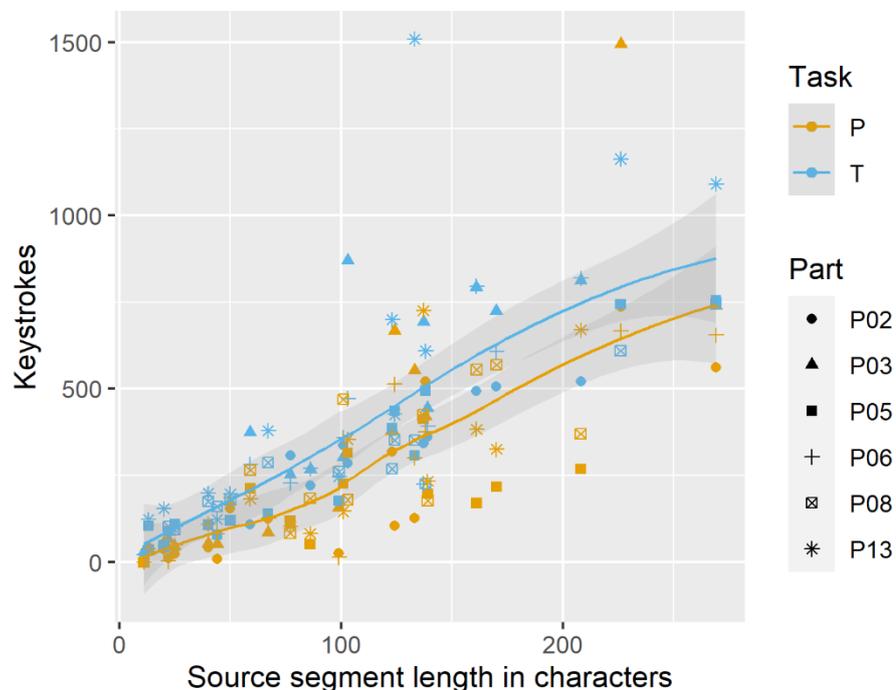


Figure 4. Keystrokes (y-axis) and source segment length in characters (x-axis) with loess lines for post-editing (P, amber) and unaided translation (T, blue). Shades represent 95% confidence intervals.

In terms of translation time, on average there were more seconds per character in translation (mean = 3.18, SD = 3.90) than in post-editing (mean = 2.66, SD = 2.14). This is a 16.6% reduction for the post-editing condition. Based on the medians, we observed a reduction of 9.3% for post-editing. We plot results for translation time in Figure 5, which shows boxplots per participant on a log scale⁸ (left) and the overall effect of sentence length (right). Based on per-participant median values, post-editing required more time for four out of six translators. This suggests that the overall average difference observed for post-editing was driven by just two of the participants. Figure 5 (right) also shows that post-editing required more time than translation for particularly long segments.

We checked for a potential effect of temporal outliers, but this was not objectively evident. As mentioned, the FDur variable by default excludes interruptions longer than 200 seconds. Furthermore, we did not identify artificial behaviour for segments where the translation process was particularly long. For example, the longest time spent on a segment corresponds to P13, who took 25.6 minutes to translate unaided a segment with 25 tokens and 133 characters. P13 went into this segment 13 times to edit it (i.e., Nedit = 13). Some of these visits lasted over three minutes and involved constant self-editing. The

⁸ Throughout the article, we occasionally use the natural log or square root scale in graphs to aid visualisation. Numeric tick labels are displayed in the original (i.e., non-transformed) scale in these cases for interpretability.

time spent on the segment therefore corresponded to actual editing behaviour. We saw no basis to exclude this data point as an outlier. There were similar data points in the post-editing condition. For instance, P03 took 22.7 minutes to post-edit a source segment with 41 tokens and 226 characters. P03 paid two editing visits to this segment, which were long and involved substantial editing. We saw no basis to exclude this data point either.

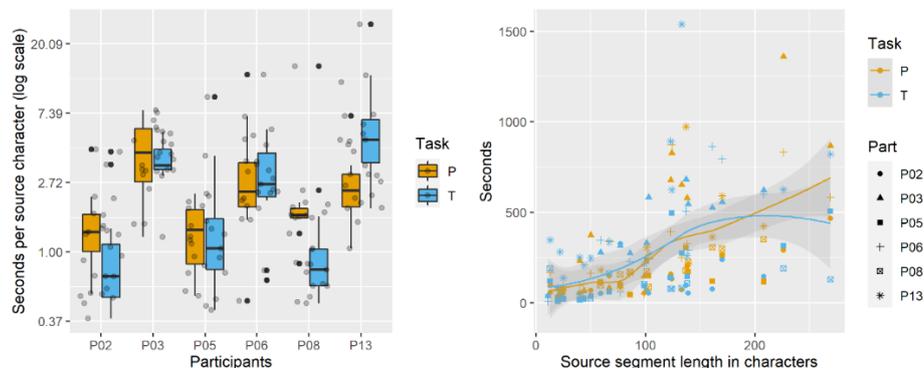


Figure 5. Per-participant boxplots showing distribution of seconds per source character (left, log scale) and seconds as a function of source-segment length (right).

The results above paint a mixed picture for the effect of MT on literary translation. While translators in our study generally typed and paused less when they used MT, this did not lead to appreciable reductions in translation time since only two translators were faster when using MT. Some translators repeatedly revisited segments for additional editing, which may have made the tasks longer for both conditions, even if with fewer pauses and keystrokes for post-editing. We discuss these results further in Section 5.

4.2. Paragraph vs. sentence segmentation

We used the same variables described in Section 3.3.3 to analyse the segmentation study. There were again high correlations between pauses and keystrokes ($r = 0.96$) and between translation time and pauses ($r = 0.91$). The correlation between keystrokes and translation time was slightly lower ($r = 0.87$), so we concentrate on these two variables below. Like in the T-P study, we also use editing visits, which had lower correlations with the other variables (between $r = 0.39$ and $r = 0.45$).

On average there were more seconds per source character when the texts were segmented into sentences (mean = 2.72, SD = 2.65) than when they were segmented into paragraphs (mean = 2.30, SD = 1.77). This represents a 15.7% reduction for the paragraph condition. Based on the medians, the reduction for the paragraph condition changes to 5.3%.

We present boxplots per translator in Figure 6 showing the distribution of keystrokes (left pane) and seconds (right pane) per source character for sentence and paragraph conditions.

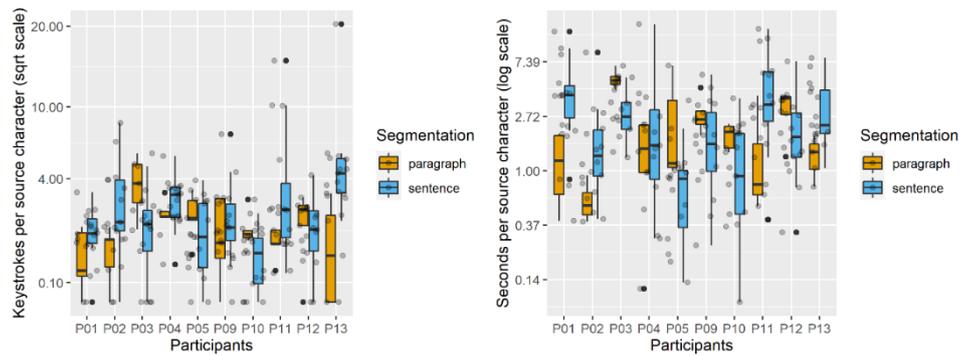


Figure 6. Per-participant boxplots showing the distributions of keystrokes per source character (left, square root scale) and seconds per source character (right, log scale)

In terms of keystrokes, there were reductions for six out of ten translators for the paragraph condition. In terms of translation time, five translators were faster for paragraphs and five of them were faster for sentences. As can be seen in Figure 6 (right pane), there were very clear time reductions for the paragraph condition for some participants (e.g., P01, P02, P11 and P13) though an equally clear opposite effect can be observed for others (e.g., P03 and P05).

Paragraph segmentation was associated with a more substantial effort reduction in terms of keystrokes. On average, there were fewer keystrokes per character for paragraphs (mean = 1.62, SD = 1.28) than for sentences (mean = 2.25, SD = 2.55), which is a 28% difference. Based on the medians, the reduction for paragraphs changes to 13.8%.

In Figure 7, we plot the number of visits (y-axis) and seconds (x-axis), all per source character. We use the natural log scale for both variables and plot linear regression lines to facilitate visualisation. Notably, Figure 7 shows that translators paid more editing visits to sentences than to paragraphs. It also shows that editing time was more closely associated with multiple visits for sentences than for paragraphs. This is possibly linked to the fact that paragraphs are more self-contained, which may allow translators to solve any editing issues in fewer visits. Breaking the text into sentences, by contrast, may spread attention, which is likely to be an underlying factor in more subsequent visits.

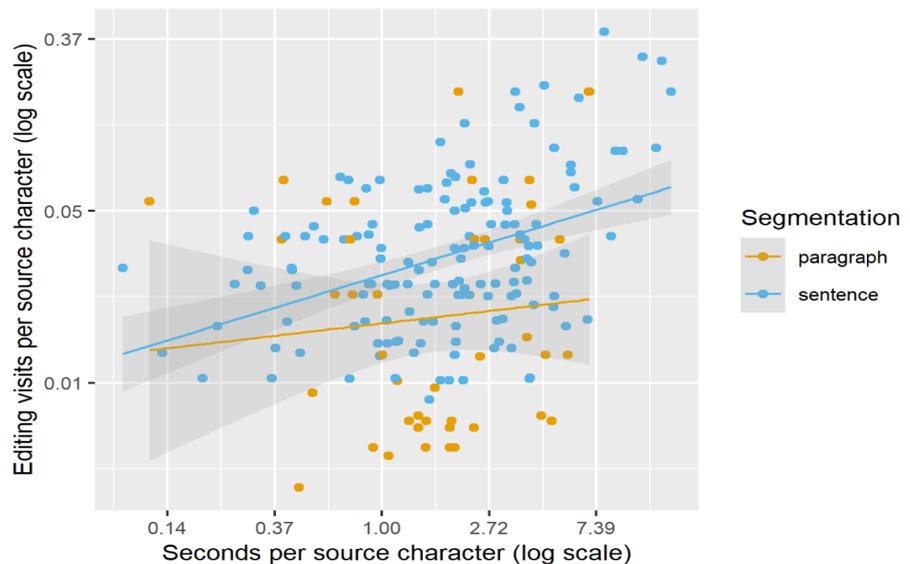


Figure 7. Number of editing visits per source character (y-axis, log scale) as a function of seconds per source character (x-axis, log scale) with linear regression lines for paragraph and sentence segmentation settings.

In short, the results for different segmentation settings suggest that on average segmenting the text into paragraphs saved post-editing effort, especially in terms of keystrokes. In terms of translation time, per-translator medians showed no majority pattern in the sample.

4.3. Translators' perceptions of CAT and MT

While we do not have enough qualitative data to provide an in-depth analysis of translators' perceptions of the tasks, in this section we present brief details of their comments on the use of MT and CAT tools for translating literature. In the T-P study, three of six translators thought that MT had been useful. We checked translators' perceptions of MT's usefulness against their median task times per character for the T-P study. Their perceptions largely matched the measurements except for a translator who thought MT was useful but had been faster without it. In the segmentation study, only one of ten translators did not find MT to be useful. While translators did not have the unaided reference for the segmentation study, they could assess the usefulness of MT by how much of it they decided to retain or by their perception of the MT output quality.

Translators' comments on MT highlight how, even for the same text type, the usefulness of MT may vary depending on a text's level of complexity and translation difficulty. One translator who took part in both studies (and was later excluded from the T-P analysis due to the issues reported in Section 3.2) found the texts in the segmentation study to be easier, which they mentioned as a potential reason for finding MT to be more useful in that study. Irrespective of text difficulty, translators' assessment of usefulness was broad and did not necessarily concern translation time. For example, one translator mentioned MT as a version of the text they could use for comparison and as a source of inspiration: "When the language is simple and unadorned [...], [MT] gives you some clue of what's going on. In other cases, though most of the time it makes little sense, it does provide some contrast and inspiration occasionally".

Half of the translators retained in each study reported having used CAT tools for literary translation before. In terms of how they thought CAT tools could be useful for literary tasks, one translator mentioned that the bilingual layout of CAT interfaces can be inherently beneficial ergonomically by facilitating access to both the source and target text on screen. This was caveated with a comment on how this applied to most text types and not just literary texts, but the comment illustrates how some of the benefits of CAT can apply across domains. Also noteworthy is that on two occasions translators mentioned that CAT environments might make it easier to keep names of people and places consistent, for example in how these are transliterated. While term bases were not used in our tasks, some of these comments served to highlight how terminological resources may present opportunities for literary translation if used to ensure consistency with proper names or other elements that recur in the story (see Horenberg, 2019, p. 70).

5. Discussion

5.1. Using MT in literary translation tasks

The results presented above suggest that, overall, MT was a useful tool for translating science fiction prose from English to Simplified Chinese, especially in terms of reductions in technical effort. The effort-saving potential of MT was not as clear in terms of translation time, however, since four of six translators had longer median task times when they used MT. This differs somewhat from previous research, where significant temporal differences are reported between post-editing and unaided translation of literary texts (see Section 2). While

sample size may be a factor in our mixed results for translation time, the fact remains that for those four translators MT did not have a timesaving effect. This should give pause for thought in discussions of MT as a literary translation tool. Even where results strongly support MT use, there will invariably be exceptions linked to the content, language pair or the MT system, which are all potential factors in our results.

Unlike in previous research on literary post-editing, which used in-domain MT systems trained on literary data (e.g., Toral et al. 2018; Guerberof and Toral 2022), here we used commercial ‘off-the-shelf’ machine translations. This is a useful set-up to examine since it replicates how literary translators with no access to in-domain systems – or the background knowledge to train one – would be likely to use MT if they chose to, client and copyright permitting. As for the language pair, while there have been improvements in English-to-Chinese MT in the past years (Wu et al., 2016), it is well known that this is still a challenging language combination for MT. With respect to the content, one translator felt that texts in the T-P study were harder to translate, which is consistent with our source-text analysis (see Section 3.1). We do not consider the fact that the texts were more difficult in that task to be problematic since our text selection procedure ensured the use of largely comparable texts in each study. What this does show, however, is that text difficulty should be considered in task-based assessments of MT. While it has been known for some time that source-text characteristics can affect the MT output (see Section 2), we argue that this is even more important to note for literary texts, which by their very nature can encompass a wide range of formats, linguistic conventions, and styles.

Despite the lack of a clear timesaving effect for MT, translators’ opinions on MT were largely positive. This is slightly different from the main findings presented by Moorkens et al. (2018), where translators preferred to translate unaided. This may be because, by comparison, translators in the present sample had mixed levels of experience with literary translation and substantial experience with MT post-editing. In any case we do not focus on translators’ preferences, but rather on whether they thought MT had been useful, which across the two studies on most occasions they did. Indeed, these results show how MT’s usefulness can be manifested in different ways. Market-driven concepts such as speed and efficiency are often at the forefront of MT evaluations, though more fluid parameters such as using MT as a source of ideas or ‘inspiration’ or simply as an example of a contrasting translation (see above) should arguably receive more attention in future work.

5.2. CAT tools and literary texts

As mentioned, English word order often needs to be inverted in Chinese (see Meifang & Li, 2009). If in this process sentences need to be merged, this requires merging segments in the CAT interface or breaking the source-target correspondence in the segmentation. Coupled with the more target-oriented solutions likely to be required by literary translation, the constraining effect of sentence-level segmentation possibly explains why segmenting the texts into paragraphs can save post-editing effort. This effort reduction is supported by summary statistics from the segmentation study, but we note that the difference between the sentence and paragraph conditions was subject to high translator variability. Depending on sentence, paragraph and task length, the additional context and freedom provided by paragraph segmentation might also come at a price. This setting may, for instance, make it easier for translators to lose their place in the text when working with longer paragraphs. This is an issue that merits future research, ideally based on larger samples.

Importantly, as we allude to in Section 2, the power of defaults should not be underestimated. For literary translators who are new to CAT, we emphasise the importance of experimenting with different segmentation settings as this may save time and effort. CAT tools were not designed with literary translators in mind, so deviating from the norm in how these tools are configured may be even more important for this text type.

There are other ways in which CAT tools may be beneficial for literary texts, as mentioned in Section 4.3 (see also Rothwell & Youdale, 2022). On the other hand, flagship CAT functionalities like translation memory would be expected to be of limited use for literary texts, where repetition is more likely to occur at the word level rather than at the level of entire phrases or sentences. MT presents new possibilities in this respect, since unlike translation memories the potential benefits of MT do not, at the point of use, depend on textual repetition. This means that not only is MT becoming a more central feature of CAT tools, but also that compared to other CAT features it may be of use for a wider range of tasks.

6. Conclusion

While our sample is small, we found that on average post-editing of literary machine translations required fewer keystrokes compared to unaided translation of literary texts. In terms of temporal differences, post-editing had a much smaller average effect. When considering per-participant results, four of six translators were faster when they translated without MT. Based on this study, therefore, although post-editing clearly required less typing, it was not appreciably timesaving.

In relation to text segmentation, we found that using Trados Studio's paragraph segmentation setting, as opposed to sentence segmentation, was associated with average reductions in keystrokes and post-editing time. For six of ten translators, paragraph segmentation required fewer keystrokes. In terms of translation time, per-participant results on segmentation showed no majority pattern.

Since we do not provide an inferential analysis in this article, these results generate rather than test hypotheses. We call for future research on the merits of using different CAT-tool settings for literary translation and mention how status quo bias can be detrimental for literary translators who are new to CAT. We also note that future research on this subject should pay further attention to target-text evaluations, which we are unable to include in this article. Empirical research on target-text assessments is arguably particularly required to examine the extent to which methods adopted in previous MT research might suit literary texts. Literature is an area where the use of MT can be explored in relation to factors that transcend efficiency or effort savings, including concepts such as inspiration and creativity, which we hope will feature more prominently in research on MT as a human translation tool.

Data availability statement

The dataset analysed in this article is available at the University of Bristol data repository at <https://doi.org/10.5523/bris.8jrz12338nd2169npievph4w>

References

- Alabau, V., Bonk, R., Buck, C., Carl, M., Casacuberta, F., García-Martínez, M., González, J., Koehn, P., Leiva, L., Mesa-Lao, B., Ortiz, D., Saint-Amand, H., Sanchis, G., & Tsoukala, C. (2013). CASMACAT: An open source workbench for advanced computer aided translation. *The Prague Bulletin of Mathematical Linguistics*, 100(1), 101-112. <https://ufal.mff.cuni.cz/pbml/100/art-alabau-et-al.pdf>
- Aziz, W., Castilho, S., & Specia, L. (2012). PET: A tool for post-editing and assessing machine translation. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, J. Odiijk, & S. Piperidis (Eds.), *Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation, 21-27 May 2012, Istanbul, Turkey* (pp. 3982-3987). European Language Resources Association.
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). *Neural machine translation by jointly learning to align and translate* [Paper presentation]. The International Conference on Learning Representations 2015 (ICLR 2015). <http://arxiv.org/abs/1409.0473>
- Besacier, L. (2014). *Traduction automatisée d'une oeuvre littéraire: Une étude pilote* [Paper presentation]. The 21ème Traitement Automatique du Langage Naturel (TALN 2014), Marseille. <http://www.aclweb.org/anthology/F14-2001>
- Carl, M. (2012). Translog-II: A program for recording user activity data for empirical reading and writing research. In N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, J. Odiijk, & S. Piperidis (Eds.), *Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation, 21-27 May 2012* (pp. 4108-4112). European Language Resources Association.
- Carl, M., Schaeffer, M., & Bangalore, S. (2016). The CRITT translation process research database. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New directions in empirical translation process research: Exploring the CRITT TPR-DB* (pp. 13-54). Cham: Springer International Publishing.
- do Carmo, F. (2020). 'Time is money' and the value of translation. *Translation Spaces*, 9(1), 35-57. <https://doi.org/10.1075/ts.00020.car>
- Futeral, M., Schmid, C., Laptev, I., Sagot, B., & Bawden, R. (2022). *Tackling ambiguity with images: Improved multimodal machine translation and contrastive evaluation*. <https://arxiv.org/abs/2212.10140>
- Germann, U. (2008). Yawat: Yet another word alignment tool. In *Proceedings of the ACL-08: HLT Demo Session (Companion Volume)* (pp. 20-23). Association for Computational Linguistics.
- Guerberof-Arenas, A., & Toral, A. (2022). Creativity in translation: Machine translation as a constraint for literary texts. *Translation Spaces*, 11(2), 184-212. <https://doi.org/10.1075/ts.21025.gue>
- Guerberof-Arenas, A., & Toral, A. (2020). The impact of post-editing and machine translation on creativity and reading experience. *Translation Spaces* 9(2), 255-282. [doi:https://doi.org/10.1075/ts.20035.gue](https://doi.org/10.1075/ts.20035.gue)
- Hadley, J., Popović, M., Afli, H., & Way, A. (Eds.) (2019). *The qualities of literary machine translation* [Paper presentation]. Machine Translation Summit XVII, 19 August 2019, Dublin, Ireland. <https://www.aclweb.org/anthology/W19-73.pdf>
- Hansen, D. (2022). *La traduction littéraire automatique : Adapter la machine à la traduction humaine individualisée*. <https://hal.archives-ouvertes.fr/hal-03583562/document>
- Horenberg, L. W. (2019). *Using CAT in literary translation - How tools may support translators in source-text analysis, translation and retranslation: A case study of "Mr Loveday's Little Outing"* (Master's thesis). Utrecht University, Utrecht.
- Hvelplund, K. T. (2011). *Allocation of cognitive resources in translation: An eye-tracking and key-logging study*. (PhD Thesis). Copenhagen Business School, Copenhagen. <http://hdl.handle.net/10398/8314>
- Kenny, D., & Winters, M. (2020). Machine translation, ethics and the literary translator's voice. *Translation Spaces* 9(1), 123-149. <https://doi.org/10.1075/ts.00024.ken>
- Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes* (G. S. Koby Ed.). Kent State University Press.

- Kolb, W. & Miller, T. (2022). Human–computer interaction in pun translation. In Hadley, J. L., Kristiina Taivalkoski-Shilov, K., Teixeira, C. S. C., & Toral, A. (Eds.), *Using technologies for creative-text translation* (pp. 66-88). Routledge.
- Lacruz, I., Denkowski, M., & Lavie, A. (2014). Cognitive demand and cognitive effort in post-editing. In S. O'Brien, M. Simard, & L. Specia (Eds.), *Proceedings of the Third Workshop on Post-Editing Technology and Practice, 11th Conference of the Association for Machine Translation in the Americas, 22-26 October 2014, Vancouver, BC Canada* (pp. 73-84). <https://aclanthology.org/2014.amta-wptp.6.pdf>
- Läubli, S., Simianer, P., Wuebker, J., Kovacs, G., Sennrich, R., & Green, S. (2022). The impact of text presentation on translator performance. *Target* 34(2), 309-342. <https://doi-org.bris.idm.oclc.org/10.1075/target.20006.lau>
- Lombardino, R. (2015, June 15). *CAT tools in literary translation: What are they good for?* <https://rafalombardino.com/news/2018/1/4/cat-tools-in-literary-translation-what-are-they-good-for>
- Macken, L., Vanroy, B., Desmet, L., & Tezcan, A. (2022). Literary translation as a three-stage process: Machine translation, post-editing and revision. In Lieve Macken, L., Rufener, A., Van den Bogaert, J., Daems, J., Tezcan, A., Vanroy, B., Fonteyne, M., Barrault, L., Costa-Jussà, M. R., Kemp, E., Pilos, S., Declercq, C., Koponen, M., Forcada, M. L., Scarton, C., & Moniz, H. (Eds.) *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation, Ghent, Belgium* (pp. 101–110). European Association for Machine Translation. <https://aclanthology.org/2022.eamt-1.pdf>
- Mandl, M., Felfernig, A., Tiihonen, J., & Isak, K. (2011). Status quo bias in configuration systems. In: Mehrotra, K.G., Mohan, C.K., Oh, J.C., Varshney, P.K., & Ali, M. (Eds.), *Modern approaches in applied intelligence* (pp. 105-114). Springer.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- Meifang, Z., & Li, P. (2009). Introducing a Chinese perspective on translation shifts. *The Translator*, 15(2), 351-374. doi:10.1080/13556509.2009.10799285
- Microsoft. (2018). *LowLevelKeyboardProc callback function*. [https://docs.microsoft.com/en-gb/previous-versions/windows/desktop/legacy/ms644985\(v=vs.85\)](https://docs.microsoft.com/en-gb/previous-versions/windows/desktop/legacy/ms644985(v=vs.85))
- Molines, N. (2020). *Creative translation and technologies expert meeting* [Paper presentation]. Creative Translation and Technologies Expert Meeting, Friday 29th May, University of Surrey, UK. https://surrey-ac.zoom.us/rec/play/uJcodr2u-zM3HtSRsgSDCv5xW420eqOs1CQb_KFfxEm8B3IDNVaIMLYTM7FO1h7mhKk4iDA5fD0Ac0Ah
- Moorkens, J., Toral, A., Castilho, S., & Way, A. (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, 7(2), 240–262. <https://doi.org/10.1075/ts.18> (pp. 105–114). 014.moo
- Ó Murchú, E. P. (2019). Using Intergaelic to pre-translate and subsequently post-edit a sci-fi novel from Scottish Gaelic to Irish. In J. Hadley, M. Popović, H. Afli, & A. Way (Eds.), *Proceedings of The Qualities of Literary Machine Translation, Machine Translation Summit XVII, 19 August 2019, Dublin, Ireland*. (pp. 20-24). <https://aclanthology.org/W19-7303.pdf>
- Rothwell, A. & Youdale, R. (2022). Computer-assisted translation (CAT) tools, translation memory and literary translation. In Deane-Cox, S., & Spiessens, A. (Eds.), *The Routledge handbook of translation and memory* (pp. 381–402). Routledge.
- Ruffo, P. (2022). Collecting literary translators' narratives: Towards a new paradigm for technological innovation in literary translation. In Hadley, J. L., Kristiina Taivalkoski-Shilov, K., Teixeira, C. S. C., & Toral, A. (Eds.), *Using technologies for creative-text translation* (pp. 18-38). Routledge.
- Şahin, M., & Gürses, S. (2021). English-Turkish literary translation through human-machine interaction. *Revista Tradumàtica* 19, 171-203. <https://doi.org/10.5565/rev/tradumatica.284>
- Spool, J. (2011). *Do users change their settings?* <https://archive.uie.com/brainsparks/2011/09/14/do-users-change-their-settings>
- Sun, S., & Shreve, G. (2014). Measuring translation difficulty: An empirical study. *Target*, 26(1), 98-127. <https://doi.org/10.1075/target.26.1.04sun>

- Taivalkoski-Shilov, K. (2019). Ethical issues regarding machine(-assisted) translation of literary texts. *Perspectives*, 27(5), 689-703. doi:10.1080/0907676X.2018.1520907
- Toral, A., & Way, A. (2015). Machine-assisted translation of literary text: A case study. *Translation Spaces*, 4(2), 240-267. <https://doi.org/10.1075/ts.4.2.04tor>
- Toral, A., Wieling, M., & Way, A. (2018). Post-editing effort of a novel with statistical and neural machine translation. *Frontiers in Digital Humanities* 5(9). <https://doi.org/10.3389/fdigh.2018.00009>.
- Verplaetse, H., & Lambrechts, A. (2019). Surveying the use of CAT tools, terminology management systems and corpora among professional translators: General state of the art and adoption of corpus support by translator profile. *Parallèles* 31(2), 3-31. <https://doi.org/10.17462/para.2019.02.01>
- Vieira, L. N. (2016). How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data. *Machine Translation*, 30(1-2), 41-62. <https://doi.org/10.1007/s10590-016-9188-5>
- Vieira, L. N., Ragni, V., & Alonso, E. (2021). Translator autonomy in the age of behavioural data. *Translation, Cognition & Behavior*, 4(1), 126-149. <https://doi.org/10.1075/tcb.00052.nun>
- Watts, P. (2014a). *About the author*. <https://rifiers.com/real/author.htm>
- Watts, P. (2014b). *The blacklist*. <https://rifiers.com/real/shorts.htm>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., & Dean, J. (2016). *Google's neural machine translation system: Bridging the gap between human and machine translation*. <http://arxiv.org/abs/1609.08144>
- Yamada, M., Mizowaki, T., Zou, L., & Carl, M. (2022). Trados-to-Translog-II: Adding gaze and Quality data to the CRITT TPR-DB. In Lieve Macken, L., Rufener, A., Van den Bogaert, J., Daems, J., Tezcan, A., Vanroy, B., Fonteyne, M., Barrault, L., Costa-Jussà, M. R., Kemp, E., Pilos, S., Declercq, C., Koponen, M., Forcada, M. L., Scarton, C., & Moniz, H. (Eds.) *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation, Ghent, Belgium* (pp. 295–296). European Association for Machine Translation.. <https://aclanthology.org/2022.eamt-1.pdf>
- Youdale, R. (2020). *Using computers in the translation of literary style: Challenges and opportunities*. London: Routledge.
- Zanettin, F. (2017). Issues in computer-assisted literary translation studies. *inTRAlinea*. <http://www.intralinea.org/specials/article/2254>
- Zakrajšek, K. (2020). CAT tools – The literary translator's new assistant? *Counterpoint* 4, 24-25. https://www.ceatl.eu/wp-content/uploads/2020/12/Counterpoint_2020_04_article_06.pdf