# Measured and perceived effort: assessing three literary translation workflows

Paola Ruffo
Joke Daems
Lieve Macken

**Paola Ruffo**
Ghent University / University of St Andrews;
paola.ruffo@ugent.be;
ORCID: 0000-0002-1688-4758

**Joke Daems**
Ghent University;
joke.daems@ugent.be;
ORCID: 0000-0003-3734-5160

**Lieve Macken**
Ghent University;
lieve.macken@ugent.be;
ORCID: 0000-0001-7516-7487

## Abstract

Professional literary translators were asked to translate three short stories using MS Word, Trados Studio 2022, and a proprietary machine translation postediting platform. This article compares measured and perceived temporal, technical, and cognitive effort across the three workflows. Data was collected via questionnaires, keylogging, and screen capturing.

Keywords: computer-aided literary translation, machine translation post-editing, literary translation, computer-aided translation, literary machine translation, user testing, keylogging.

## Resumen

Se pidió a traductores literarios profesionales que tradujeran tres cuentos cortos utilizando MS Word, Trados Studio 2022 y una plataforma propia de posedición para traducción automática. Este artículo compara el esfuerzo de tiempo, técnico y cognitivo, tanto medido como percibido, en los tres flujos de trabajo. Los datos se recopilaron mediante cuestionarios, registro de teclas (keylogging) y capturas de pantalla.

Palabras clave: traducción literaria asistida por ordenador, posedición de traducción automática, traducción literaria, traducción asistida por ordenador, traducción literaria automática, pruebas con usuarios, registro de teclado.

## Resum

S'ha demanat a traductors literaris professionals que tradueixin tres contes curts utilitzant MS Word, Trados Studio 2022 i una plataforma pròpia de postedició en traducció automàtica. Aquest article compara l'esforç de temps, tècnic i cognitiu, tant mesurat com percebut, en tots tres fluxes de treball. Les dades s'han recopilat per mitjà de qüestionaris, registre de tecles (keylogging) i captures de pantalla.

Paraules clau: traducció literària assistida per ordinador, postedició de traducció automàtica, traducció literària, traducció assistida per ordinador, traducció literària automàtica, proves amb usuaris, registre de teclat.

*Rebuda: 31 de octubre de 2023 / Acceptació: 4 de novembre de 2024 / Publicació avançada: 20 de desembre de 2024*

Revista Tradumàtica. Número 22 (Desembre 2024)
**Citació recomanada:** RUFFO, PAOLA; DAEMS, JOKE; MACKEN, LIEVE (2024). «Measured and perceived effort: assessing three literary translation workflows». Revista Tradumàtica. Tecnologies de la Traducció, 22, 238-257. <https://doi.org/10.5565/rev/tradumatica.378>
http://revistes.uab.cat/tradumatica
ISSN 1578-7559

## 1. Introduction

The once widespread assumption that translation technology has no place in literary translation has recently started to falter, leaving space for an exploration of the possibilities and implications of translation technology use in creative-text translation. In this respect, the last few years have seen an increase in the number of panels, events and journal issues dedicated to the topic. Recent examples of this in academia are the 2024 special issue of *Translation in Society* on literary translatorship in digital contexts (2024), the Creative-text Translation and Technology (CTT) workshop co-located with the 25th Annual Conference of The European Association for Machine Translation (EAMT 2024), the 2023 *Tradumàtica* special issue on computer-aided literary translation (2023), and the edited volume *Computer-Assisted Literary Translation* (Rothwell et al., 2024). The impact of automation on literary translation has also been the subject of recent articles in mainstream media such as *The Atlantic* (Klemin, 2024) and *The Guardian* (Aslanyan, 2024), and literary translation associations have issued statements and guidelines regarding the use of artificial intelligence (AI) in literary translation (e.g. CEATL, 2023; The Society of Authors, 2023).

As technology and discourse in public and academic spheres both advance, they bring to the fore numerous questions related to technology adoption in literary translation. For example, much still needs to be uncovered on the effects of postediting on the literary translation process and product, translation quality, translator productivity, literary translators' use of (translation) technology, their workflow preferences, and the exact ways in which tools such as machine translation (MT), generative AI, and computer-aided translation (CAT) tools can be beneficial to literary translation and/or affect translators' working conditions and socio-economic status. In recognising this socio-technological shift, Declercq and van Egdom (2023) advocate for a nuanced and critical approach to the study of literary translation and technology, emphasising the importance of balancing "the promises of efficiency, depth of interpretation and accessibility offered by language automation" with the "need to remain attuned to the intricate fabric of human communication and the craft of literariness across languages" (Declercq and van Egdom, 2023: 59). In a similar vein, Way et al. (2023) urge their readers to reevaluate the assumption that CAT tools and MT are neither useful nor appropriate for the translation of literature. However, they also warn against conceding to narratives revolving around the idea of replacing translators, rather stressing the importance for practitioners to "[maintain] control over their preferred translation workflow" while exploring how these tools can be used "to help not only achieve improvements in productivity, but also as an essential aid in the ideation process itself" (Way et al., 2023: 97).

If a nuanced approach to the role of translation technology in literary translation is to be hoped for, then it becomes increasingly important to explore the issues at hand in as much depth as possible and from a wealth of different angles and perspectives, in order to reflect the social and economic complexities that technological innovation — and automation in particular — engender. It is in this context that the EU-funded Developing User-centred Approaches to Technological Innovation in Literary Translation (DUAL-T) project operates. In particular, DUAL-T aimed at addressing the lack of involvement of literary translators in translation technology research, as well as testing and comparing the use of word processors and CAT tools in addition to machine translation postediting (MTPE). During the project, professional literary translators were asked to translate three short stories from English into Dutch using, respectively, Microsoft Word, Trados Studio 2022, and a proprietary web-based MTPE platform. The project looked at both literary translators' attitudes and effort, employing questionnaires, interviews, keylogging and screen capturing as data production methods. This article will focus on temporal, technical, and cognitive effort (Krings, 2001), reporting results on translation time, time spent outside the tool, number of keystrokes and pauses, and participants' perceived effort for each of the three workflows. The research questions that will be addressed are the following:

1. How do overall translation time and time spent outside of the tool differ between the three translation workflows?
2. How does technical effort differ between the three translation workflows?
3. How does cognitive effort differ between the three translation workflows?
4. Does perceived effort match measured effort?

The following sections will provide an overview of state-of-the-art literature on translation technology research in literary translation, outline the study's methodology, and present and discuss results on temporal, technical, and cognitive effort, as well as participants' perceived effort.

## 2. Literature review

While the use of automated translation processes and/or postediting in the publishing industry is not currently regarded as an established practice, this is starting to change. For example, what has been described as "the world's first fully AI powered publisher" (Reedz, 2023) was launched in 2023. The rise in availability of generative AI tools has further complicated matters, resulting in the increasing convergence of conversational AI and MT.

So far, research on postediting for literary texts using neural machine translation (NMT) has shown its use tends to increase productivity. A 36% reduction in translation time was found by Toral et al. (2018), who also observed a reduction in technical and cognitive effort (by 23% and 42%, respectively) for translation from English into Catalan. Postediting was also faster in a study of literary MT from English to Slovene (Kuzman et al., 2019). However, claims that postediting results in temporal gains have recently started to be questioned, both in literary (Guerberof-Arenas and Toral, 2022; Noriega-Santiáñez and Pastor, 2023) and non-literary translation (Terribile, 2023). Postediting of literary texts has also been linked to lower technical and cognitive effort, having been shown to reduce the number of keystrokes and pauses (Toral et al., 2018; Guerberof-Arenas and Toral, 2022; Kolb, 2024).

The effects of MT and postediting on literary translation quality have also been investigated in recent years, often via employing automated evaluation metrics to compare different MT outputs and human translations. In a study evaluating the translation of literary prose sentences by a tailored MT system, Matusov (2019) found that 28-30% of them were of acceptable quality for the language pair German-English. Overall, higher quality tends to be achieved when using MT systems adapted to literature rather than commercial MT. However, human translations present higher levels of literariness and lexical variety (Toral et al., 2024). This echoes results from Webster et al. (2020), who, in addition to numerous errors in the MT output, found literary NMT to present a lower degree of lexical richness and local cohesion

Despite promising results, MT output is not currently deemed of sufficient quality to produce a publishable literary translation without human intervention (European Commission, Directorate-General for Education, Youth, Sport and Culture, 2022; Guerberof-Arenas and Toral, 2022; Oliver, 2024). As Van Egdom et al. (2023: 142) put it, "custom-built MT solutions do seem to hold potential, even in the literary domain, but this potential should never be overstated". Hongtao (2023) is of a similar opinion, stating that "the particularity and complexity of literary translation have been underestimated". This is particularly salient in light of the fact that automated evaluation metrics might provide a misleading understanding of literary MT quality (Van Egdom et al., 2023). Furthermore, the priming effect associated with using MT output at the outset of the literary translation process has so far been confirmed in studies of creativity, as well as translator voice and style. Higher levels of creativity were found in human literary translation in Guerberof-Arenas and Toral (2022) and Noriega-Santiáñez and Pastor (2023). According to Kolb (2024: 65), "MT priming seems to be an intrinsic element of PE processes". The question of priming is also relevant when discussing translator voice, which has been shown to be affected by the use of MT (Kenny and Winters, 2020). The study of postediting literary translation is, thus, fundamental to reducing the potential risks associated with losing linguistic richness (Toral, 2019; Vanmassenhove et al., 2019), as well as to exploring ethical issues that result from the inclusion of automation tools in creative-text translation workflows.

As a response to the focus on MT and postediting, a strand of research has emerged that explores the use of CAT and other tools for the translation of literary texts (see, for example, Horenberg, 2019; Youdale and Rothwell, 2022; Vieira et al., 2023), either in isolation or in combination with interactive MT (Rothwell, 2024). In Ruffo (2022), 25% of the 150 literary translators who were surveyed specifically mentioned using CAT tools for their literary translation work, while 18% of Daems's (2022) 153 respondents used translation technology in their practice. The discussion of CAT tool use in literary translation is often accompanied by another on augmented translation and the development of one or more human-centred tools tailored to literary translation which expand on existing CAT tool functionalities. CAT tools are also viewed as having the potential to enhance the creative process while enabling literary translators to retain control over their workflow. A creativity-enhancing approach has been tested, for example, by Kolb and Miller (2022), who found the software PunCAT to be effective in inspiring translators to find more creative solutions when working with puns. According to Rothwell (2024: 124), "the translation tool of the future [...] would act as an outward-facing creative portal augmenting the literary translator with flexible access, within a single interface, to potentially numerous linguistic and stylometric resources". Flexibility appears to be the common denominator of an ideal future computer-aided literary translation (CALT) tool, as Rudan et al. (2024) identify it as one of the attributes of augmentation-centred technology. This being said, Hadley (2024) highlights how currently available CAT tools were not developed with literary translation in mind, and thus, in the future, we might see the development of several different pieces of software that each address distinctive aspects of text analysis, processing, and translation, as well as a shift towards features allowing human translators to agilely manage both the type and quantity of machine input they receive while working on a translation.

Virtually absent from translation technology research for several years, studies centring on professional literary translators' attitudes towards and use of technology have increased in the past decade. This research has often uncovered a complex portrait of practitioners' relationship to technology. In terms of use, MT and CAT tools are still relatively rare in the literary translator toolkit. In particular, literary translators exhibit low levels of confidence and familiarity with translation technology (Daems, 2022; Ruffo, 2022), which directly affects their willingness to adopt such tools, as well as their attitudes towards them. In this respect, literary translators are more positive towards general technology tools (e.g. word processors, online dictionaries, etc.), and their view of translation-specific tools is also affected by the way these tools are presented as a potential substitute for professional translators (Ruffo, 2022). Additionally, the ways literary translators employ translation technology in their work tend to diverge from the original (e.g. using CAT tools for a first draft, using MT to check whether any parts of the source text were skipped) (Slessor, 2020; Daems, 2022; Ruffo, 2022). Professional literary translators are sceptical of MT's contribution to their practice (Şahin and Gürses, 2021) and, even after trying postediting, they prefer translating from scratch, regardless of any reductions in effort (Moorkens et al., 2018). Ultimately, the personal and professional spheres of literary translators are deeply intertwined, and the way they see their profession affects how they use and relate to technology (Kolb, 2017; Ruffo, 2024). Thus, it is important to guarantee that literary translators' input and participation in translation technology research are actively sought.

## 3. Methodology

Data was collected between October 2023 and January 2024 on the Ghent University and Leiden University campuses. Ethical clearance was given by the European Commission, while ethics approval was granted by the Ethics Committee of the Faculty of Arts and Philosophy at Ghent University. The anonymised dataset and some of the experiment materials are freely available at this link: https://zenodo.org/communities/dual-t/records?q=&l=list&p=1&s=10&sort=newest.

### *3.1. Participants*

Potential participants were contacted via professional translator associations in Flanders and in the Netherlands. A total of 24 professional English > Dutch literary translators took part in the study. However, three participants had to be excluded from the analysis, so the final dataset presented here comprises data from 21 participants. In two cases, the exclusion from the dataset was due to keylogging files becoming corrupted during data collection, while one participant was excluded due to not following the guidelines for one of the tasks. Participants were paid a flat fee of 250 euros to participate in the study and their travel expenses were reimbursed. On average, a session lasted between four and five hours. Participants were assigned codes from P01 to P24.

Of the 21 participants, five were aged 26-36, five were 36-45, three were 46-55, five were 56-65, and three were older than 65. The average age of participants was 48 years old. In terms of gender distribution, 13 identified as female and eight identified as male. Participants had an average of 15 years of experience translating professionally, and a specific average of 12 years in literary translation. Twelve participants had between four and 10 years of experience in literary translation, while four had between 15 and 28 years of experience. The least experienced participant had been translating for one year, while the most experienced participant had been translating for 43 years.

In evaluating their English proficiency, 12 participants self-reported being at C2 level (highly advanced), eight at C1 level (advanced), and one participant identified themselves as native/bilingual. Thirteen participants hold an academic qualification in Translation, 11 of these at postgraduate level. Additionally, two participants indicated holding an MA in Languages and Literature, and one participant reported holding an MA in Comparative Literature.

Eight participants reported that they had used postediting in their professional translation practice, with half of them employing it for literary translation. These participants were almost equally distributed among all age groups. The same number of participants (eight) indicated that they had used CAT tools in their translation practice, with four of them using them for literary translation. The majority of CAT tool users (six) were under the age of 45.

### *3.2. Texts*

The texts chosen for the experiment were three short stories from the 2014 short story collection *One More Thing* by US actor, comedian, and writer B. J. Novak. The stories in the book often have a satirical edge and explore various aspects of modern life and societal norms with humour and occasional absurdity. The three texts participants worked on were *Rome* (T01; 306 words, 29 sentences, average words per sentence = 11.4, type-token ratio (TTR) = 55%), *The Beautiful Girl in the Bookstore* (T02; 349 words, 26 sentences, average words per sentence = 14.5, TTR = 53%), and *They Kept Driving Faster and Outran the Rain* (T03; 290 words, 27 sentences, average words per sentence = 11.2, TTR = 51%). The short story format was chosen in order to have three texts that were self-contained and short enough to be translated in one experimental session using all three conditions, while also being relatively challenging for professional literary translators. The short stories also had to be part of an English-language collection that had already been translated into Dutch to enable the compilation of a translation memory (TM) for the Trados Studio 2022 task which would contain similar texts by the same author (see section on experimental setup).

### *3.3 Experimental setup and procedure*

The experimental sessions took place either at Ghent University or at Leiden University. All experimental sessions were conducted using the same setup, and in the same room on each campus (Figure 1 shows the room and setup at Leiden University). Before starting the experiment, participants were asked to read an information letter and sign a consent form. They were also provided with a translation brief introducing the short story collection they were going to be working on, and instructing them to translate to the best of their abilities and provide a translation as close to publishable quality as possible, given the experimental constraints.

*Figure 1. Setup for the experiment at Leiden University.*

The first task was to complete a pre-task questionnaire asking for information on age, gender, language proficiency, qualifications, years of experience, levels of confidence with the tools used in the experiment, prior experience of postediting and using CAT tools, and current use of and attitudes towards translation technology. The participants were then asked to translate three short stories using a different tool for each of the stories. The tools they were asked to use were the word processor Microsoft Word (WF01), the CAT tool Trados Studio 2022 (WF02), and a proprietary web-based MTPE platform provided by he Swedish company Nuanxed, the industry partner of the project (WF03). For all conditions, translators were allowed to use a web browser to consult online resources as they felt appropriate. More details on each translation condition are provided below:

- Microsoft Word
- For the WF01 tasks, participants had to type their translation in a Microsoft Word file. The source text was provided to them in a PDF file.
- Trados Studio 2022
- Prior to translating in Trados Studio, participants were given a tutorial on the main features of the software. The tutorial was accompanied by a warm-up task for participants to familiarise themselves with the software. For the main task, a TM and a termbase (TB) were attached to the Trados Studio project. The TM contained the official Dutch translation of *One More Thing* (*Onverzameld Werk*, translated by Jevgenia Lodewijks, Lydia Meeder, and Maarten van der Werf), with the exception of the three short stories used for this study. The TB was compiled using Sketch Engine to automatically extract key terms from the whole short story collection.
- MTPE platform
- Participants were given a tutorial on the main features of the software and a warm-up task to familiarise themselves with the interface prior to starting the main task. The platform is web-based and it shows the source text and machine-translated target text side by side. All participants were provided with the same version of the machine-translated texts, which was generated using a commercially available MT provider in July 2023.

Participants translated each text in one of the three translation conditions. The Microsoft Word task was used as a benchmark and always appeared first, while the Trados and MTPE conditions alternated between being the second and third conditions presented to participants. In order to obtain a balanced dataset and control for task order and text difficulty, the study was designed so that each text-condition combination appeared the same number of times across the whole dataset. Furthermore, each combination involving the use of Trados and the MTPE platform appeared the same number of times as the second and third conditions respectively. The balance could not be fully retained due to having to exclude participants from the analysis. However the differences across the sample are minimal.

Following the three translation tasks, participants were asked to complete a post-task questionnaire requesting that they rank the three translation conditions and the tools' features, as well as re-share their attitudes toward the tools. Finally, participants took part in an in-depth interview about the tasks. Interview data will not be part of the analysis and discussion presented in this article.

For each translation task, keystrokes were recorded using Inputlog 8.0 (Leijten and Van Waes, 2013) and participants' screens were recorded using OBS Studio (version 29.1.3). The questionnaires were administered using Qualtrics.

A total of 63 keystroke logs were analysed to determine translation time, time spent outside the tool, number of keystrokes and pauses, and pause duration. The data was extracted by running source and pause analyses in Inputlog.

To verify if workflow had a statistically significant impact on temporal, technical, and cognitive effort, linear mixed-effect models were fitted in R (R Core Team, 2024), using the packages lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) for significance testing. Workflow was always used as a predictor, and participant codes and text were included as random effects. The model with the predictor was tested against a null model without a predictor. Model results are only reported when the model with the predictor outperformed the null model.

## 4. Results

### *4.1 Temporal effort*

An overview of temporal effort for each participant can be seen in Figure 2. In terms of temporal effort, it took participants an average of 52 minutes to translate using Word, 45 minutes to translate using Trados, and 43 minutes to translate using the MTPE platform. Overall, using MTPE reduced translation time by 17% when compared to Word, and by 4% when compared to Trados. The linear mixed-effect model confirmed that there was an impact of workflow on total time in seconds, with the Word condition leading to an estimated increase in time of 534 seconds (standard error = 164, $p$ = 0.002) compared to the MTPE condition. Large variability was observed among participants, as the highest value was 109 minutes (P23, Word) and the lowest was 13 minutes (P04, MTPE). Participants took between 20 and 109 minutes to translate using Word, 16 and 77 minutes to translate using Trados, and 13 and 76 minutes to translate using the MTPE platform. MTPE was the fastest condition for 12 participants (57%), while Trados was the fastest condition for seven participants (33%). Word was the slowest condition for 10 participants (48%), while Trados and MTPE were the slowest conditions for six and five participants respectively. In terms of time spent outside the tool, participants spent an average of 13 minutes outside Word, and seven minutes outside each of Trados and the MTPE platform. For the Word task, only an average of one minute was spent on the PDF file containing the source text. However, it is worth noting that participants were working with the Word and PDF files side by side, which means part of the time spent inside Word might have been spent on the PDF file instead, which would increase the amount of time spent outside the tool for the Word condition.

Figure 2 also shows how participants ranked each workflow in terms of translation speed, with 1 being the fastest and 3 the slowest. Overall, 10 participants (48%) ranked the MTPE platform workflow as the fastest. The same number of participants ranked Trados second, while Word was ranked first by six participants, second by seven participants, and third by eight participants. While perceived temporal effort broadly matches measured effort, especially for MTPE, it is worth noting that only two participants were faster under the Word condition, yet this workflow was perceived as the fastest by seven participants. In contrast, Trados was perceived as slower than measured, being ranked first by four participants.

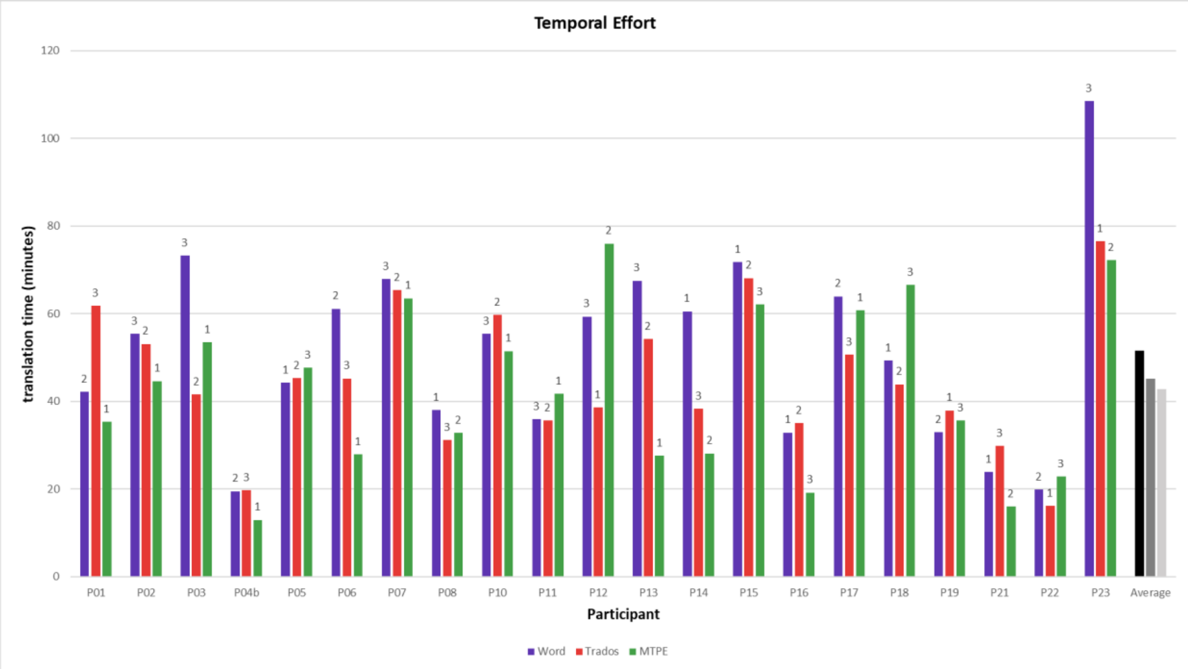


*Figure 2. Translation time (minutes) for each workflow and participants' workflow preferences.*

## 4.2 Technical effort

Number of keystrokes was used as a measure of technical effort. Figure 3 shows the number of keystrokes for each participant and translation condition. On average, literary translators used 5087 keystrokes when translating in Word, 3750 keystrokes when translating in Trados, and 2327 keystrokes when translating in the MTPE platform. Overall, the number of keystrokes decreases by 54% with MTPE when compared to Word, and by 38% with MTPE when compared to Trados. The linear mixed-effect model confirmed that there was an impact of workflow on the number of keystrokes per second, with both the Word condition (estimate = 0.81, standard error = 0.1, $p < 0.001$) and the Trados condition (estimate = 0.62, standard error = 0.1, $p < 0.001$) leading to a significant increase in the number of keystrokes per second compared to the MTPE condition. The highest number of keystrokes was 12922 (P03, Word), while the lowest number of keystrokes was 461 (P04, MTPE). Large variability was observed for each condition, as participants used between 12922 and 2507 keystrokes to translate in Word, between 6474 and 1973 keystrokes to translate in Trados, and between 4998 and 461 keystrokes when translating in the MTPE platform. Overall, almost all participants (20) used more keystrokes when translating with Word, while 17 used fewer keystrokes when translating with the MTPE workflow.

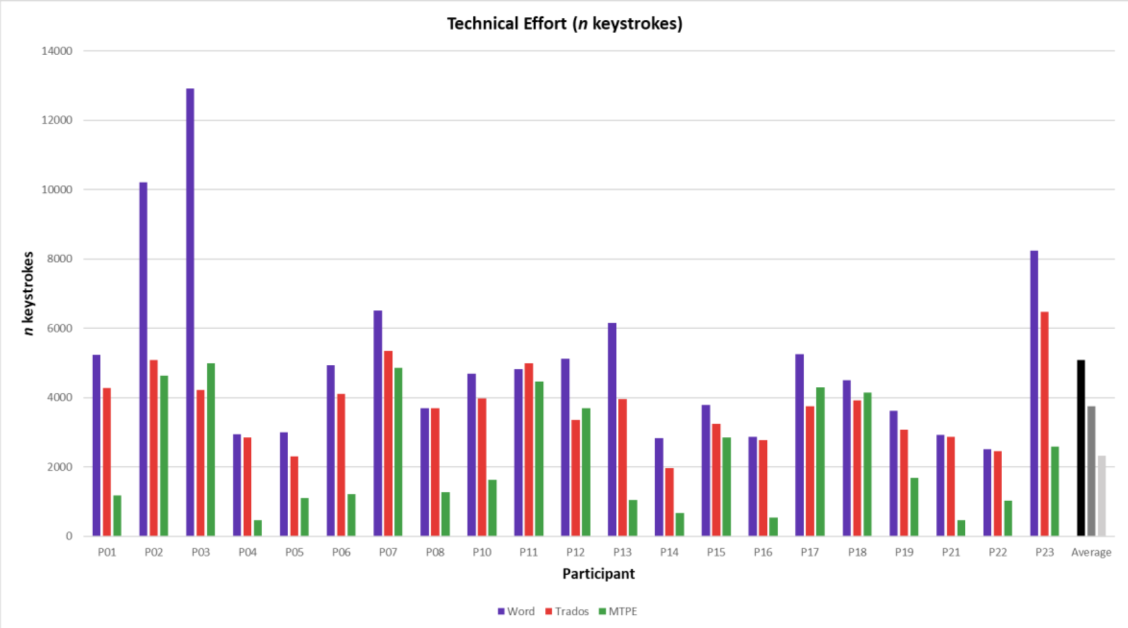


*Figure 3. Number of keystrokes for each participant and translation condition.*

## 4.3 Cognitive effort

The number and duration of pauses were used as measures of cognitive effort. The pause threshold was set at 300 ms (Lacruz et al., 2014). As shown in Figure 4, there was an average of 1200 pauses with the Word workflow, 1016 pauses with the Trados workflow, and 848 pauses with the MTPE workflow. Using the MTPE platform translates to a 29% decrease in the number of pauses when compared with Word, and a 16% decrease when compared to the Trados condition. Furthermore, using Trados resulted in a 15% decrease in the number of pauses in comparison with Word. Linear mixed-effects analysis confirmed that workflow was a significant predictor of the number of pauses, with Trados (estimate = 175, standard error = 56, $p = 0.003$) and Word (estimate = 341, standard error = 56, $p < 0.001$) leading to a significantly higher number of pauses compared to the MTPE workflow. Here, too, large variability among participants was observed, as the highest number of pauses recorded was 2002 (P23, Word), while the lowest was 260 (P04, MTPE). The number of pauses was between 454 and 2002 for the Word condition, between 411 and 1605 for the Trados condition, and between 260 and 1632 for the MTPE condition. Overall, the highest number of pauses occurred in the Word condition for 13 participants (62%), while the lowest occurred in the MTPE platform for 15 participants (71%). The Trados condition saw a somewhat even split among participants, as five participants paused the most and six participants paused the least while using Trados.

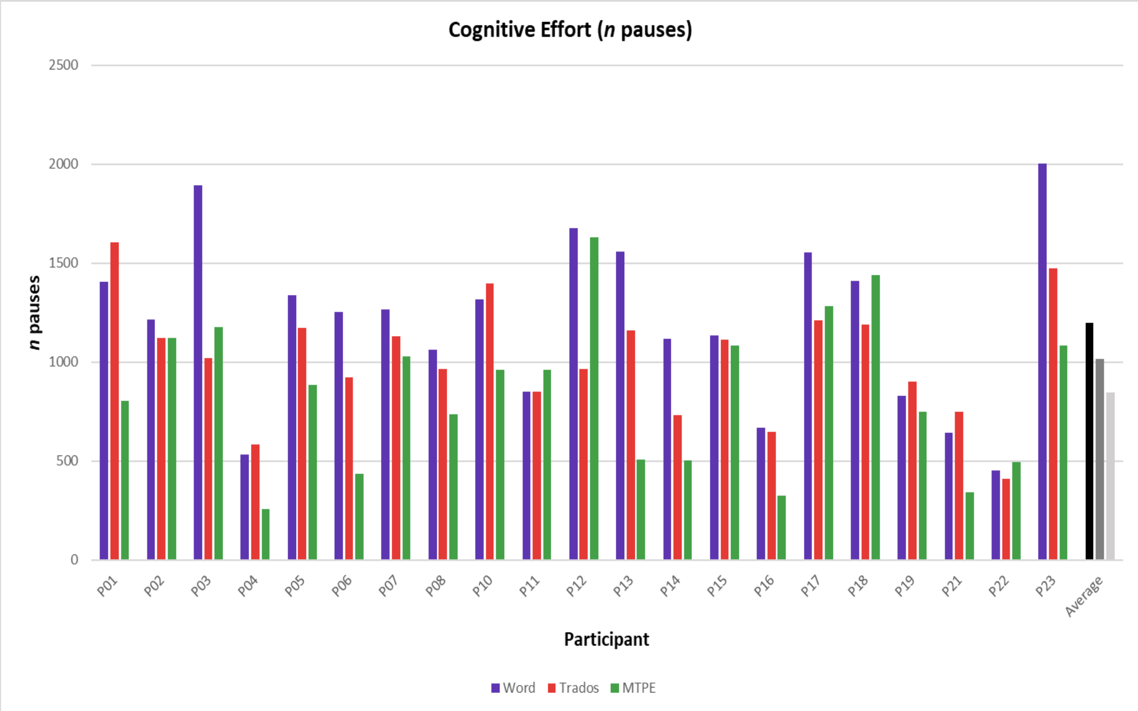


*Figure 4. Number of pauses for each participant and condition.*

The total duration of pauses amounted, on average, to 31 minutes for the Word condition, 28 minutes for the Trados condition, and 25 minutes for the MTPE condition. Linear mixed-effects analysis confirmed that workflow was a significant predictor of the total pause duration, with Word (estimate = 354, standard error = 95, $p < 0.001$) leading to a longer total pause duration compared to the MTPE workflow. The mean duration of pauses (Figure 5) was 1.566 seconds for Word, 1.644 seconds for Trados, and 1.805 seconds for the MTPE platform. Linear mixed-effects analysis confirmed that workflow was a significant predictor of mean pause duration, with Trados (estimate = -0.15, standard error = 0.06, $p = 0.018$) and Word (estimate = -0.23, standard error = 0.06, $p < 0.001$) leading to shorter mean pause durations compared to the MTPE workflow. The mean duration of pauses was the highest with MTPE for 13 participants (62%), while it was the lowest with the Word condition for 11 participants (52%). The average pause ratio (obtained by dividing the total pause time by the total translation time) was 60% for Word, 61% for Trados, and 58% for the MTPE platform. The linear mixed-effects model with workflow as a predictor did not outperform the null model, so workflow was not found to significantly impact average pause ratio. Thus, while the total duration of pauses was slightly less for MTPE than for the other workflows, the mean duration was slightly higher for the MTPE workflow, and the pause ratio was between 58 and 60% across the three conditions.

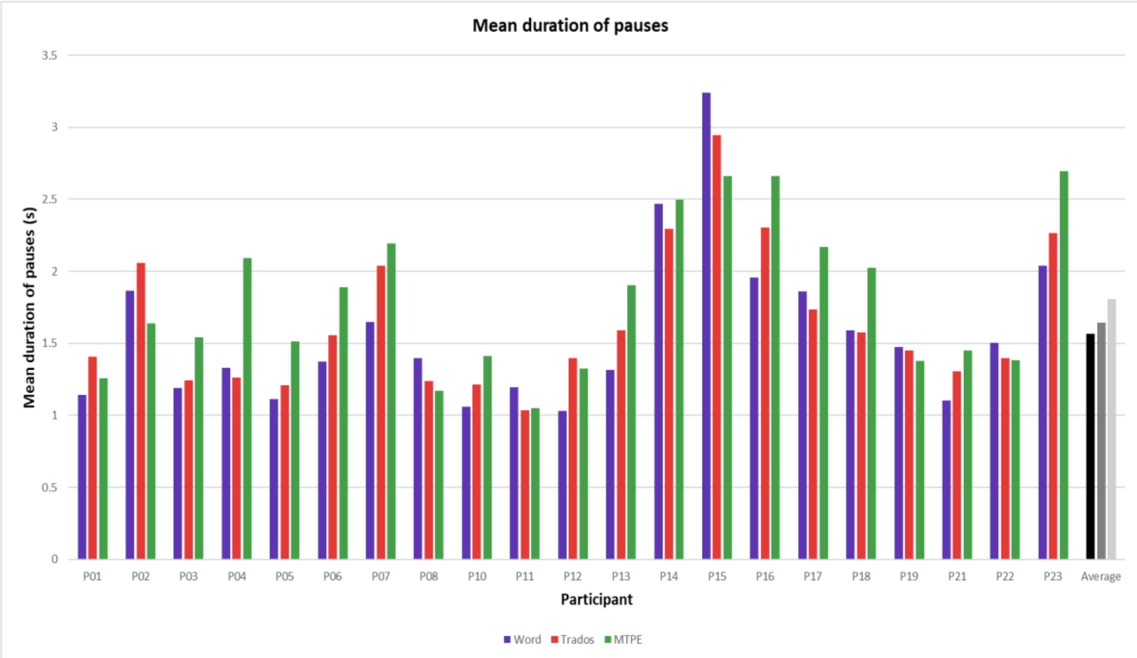


*Figure 5. Mean duration of pauses.*

## *4.4 Perceived effort*

In the post-task questionnaire, participants were asked to rank the three workflows in terms of effort required, with 1 being the condition requiring the most effort and 3 the condition requiring the least effort. Word was perceived as the condition requiring the most effort by more than half of participants (11), followed by the MTPE workflow (eight). Nine participants indicated that the MTPE condition required the least effort. Trados was ranked second by most respondents (12). Results show that, on average, participants perceive MTPE to require more effort than was measured, and Word to require less effort than was measured. This is almost double when looking at technical effort for MTPE, where 17 participants used the smallest number of keystrokes, and almost half when looking at Word, where 20 participants used the highest number of keystrokes.

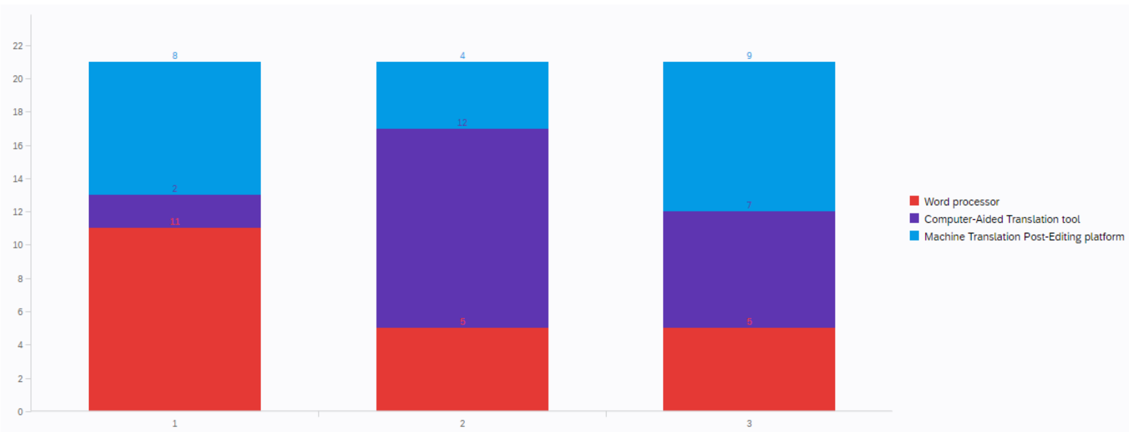


*Figure 6. Participants' perceived effort for each condition.*

## 5. Discussion and conclusion

The section above has presented data on literary translators' temporal, technical, and cognitive effort, as well as their perceptions of effort, when using Microsoft Word, Trados Studio 2022, and a proprietary MTPE platform to translate.

A comparison of the three workflows has revealed that the MTPE condition reduced translation time by 17% when compared to Word, and that MTPE was the fastest condition for 57% of participants, while Word was the slowest for almost half of participants. Technical effort was reduced by 54% when working with MTPE instead of Word. Working with MTPE also resulted in a 29% decrease in cognitive effort (measured using number of pauses) when compared to Word. For 62% of participants, the highest number of pauses occurred while working in Word, while the lowest number of pauses occurred when working with the MTPE platform. Analysing pause behaviour by looking at mean pause duration and pause ratio has highlighted similarities in the proportion of time participants spent pausing in relation to overall translation time (between 58 and 60% for all workflows). Furthermore, the mean duration of pauses while using the MTPE platform was slightly longer than when using Word or Trados, and it was the highest in this translation condition for 62% of participants.

In terms of participants' perceptions of effort, a tendency was observed for participants to rank Word as faster and as requiring less effort than was measured. This is in contrast to the findings of Moorkens et al. (2018), where measurements and perceptions coincided in terms of temporal effort. Similarly, MTPE tended to be ranked as slower and as requiring more effort than was observed for all tasks. This divide becomes particularly evident when looking at technical effort, as the number of participants who perceived Word as requiring the most effort is almost half the number of participants for whom Word was the condition with the highest number of keystrokes. The inverse was observed for MTPE, which was perceived as the method requiring the least effort by nine participants, while keystroke measures show that this was the case for 17 participants.

Gains in terms of temporal effort are lower than those found in other studies on literary translation so far (e.g. slightly more than half the 36% reduction in temporal effort found by Toral et al., 2018). This might be due to the use here of a generally available MT system rather than a highly customised one, as well as to having worked on a standalone full text for each task, rather than segments, which is arguably closer to how a literary translator would complete a similar translation task in real life. In this respect, it is worth noting that the type of text used might also have influenced the amount of time required to complete the translation task, and that higher gains in translation time might be achieved when working with less creative texts.

Using MTPE more than halved technical effort for almost all participants and reduced cognitive effort by slightly less than a third for 71% of participants. However, pause behaviour remained consistent across all three conditions, and pauses lasted slightly longer with MTPE, which could indicate a change in the translation process in terms of how literary translators relate to the machine-translated text and approach the postediting task when compared to translation from scratch. This might also explain why more literary translators tended to perceive MTPE as requiring more effort than shown by the measurements. Toral et al. (2018) also found pauses to be fewer but longer when postediting literary texts. Possible reasons for this might include not trusting the MT output, needing to do less typing but to spend more time checking and revising the output, pre-existing attitudes towards the practice of MTPE, as well as not being familiar with the tool, or particularly challenging segments requiring more postediting effort. The latter factor was mitigated here by providing a tutorial and a warm-up task before the translation task. Overall, the analysis of participants' perceptions highlighted the polarisation between the Word and the MTPE workflows, which emerge as the two ends of a continuum where Trados occupies an intermediate position.

In addition to the above, due to the high variability of speed, number of keystrokes and pauses, the effects of each workflow on the literary translation process might be highly dependent on each translator's specific way of working, as well as the type of text being translated.

Ultimately, there is still more to be done in terms of understanding how different types of technology affect the literary translation process and product. In particular, gaining more insights into how literary translators relate to and view each of these tools might help to explain differences between measured and perceived effort, and specific tasks for which human translation, CAT, postediting or other technological resources might be useful. In this respect, saving time or effort might not be a priority for literary translators, who might deem it more important to retain control over their translation and workflow. It might also be interesting to conduct similar studies on longer texts (e.g. a full novel) to understand, for example, whether gaining more familiarity with the tool used has any impact on effort and perception. Going forward, it is also important to consider the specific ways in which MTPE might affect the translation product in terms of quality and creativity, as well as reader reception.

The next phase of this project will address some of the aspects highlighted above by relating user testing results to participants' attitudes towards technology before and after the task, their narratives of each translation condition as emerging from the post-task interviews, and what effect each condition had on translation quality and creativity.

## Acknowledgements

This project has received funding from the European Union's Horizon Europe (HORIZON) research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 101062428.

## References

Aslanyan, Anna. (2024). AI translation: how to train "the horses of enlightenment". *The Guardian*, 15th March. Available at https://www.theguardian.com/books/2024/mar/15/ai-translation-literature [Accessed: 20240515].

Bates Douglas; Mächler Martin; Bolker Ben; Walker Steve (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, v. 67, n. 1, pp. 1–48. <https://doi.org/10.18637/jss.v067.i01>. [Accessed: 20241223].

CEATL (2023). Statement on Artificial Intelligence.. <https://www.ceatl.eu/tools-of-the-trade/statement-on-artificial-intelligence>. [Accessed: 20240515].

Daems, Joke (2022). Dutch literary translators' use and perceived usefulness of technology. In: Hadley, James; Taivalkoski-Shilov, Kristiina; Teixeira, C. S. Carlos; Toral, Antonio (eds.). *Using Technologies for Creative-Text Translation*. New York: Routledge, pp. 40–65. <https://doi.org/10.4324/9781003094159-3>. [Accessed: 20241223].

Declercq, Christophe; van Egdom, Gys-Walt (2023). No more buying cats in a bag? Literary Translation in the age of language automation. *Revista Tradumàtica*, v. 21, pp. 49–62. <https://doi.org/10.5565/rev/tradumatica.407>. [Accessed: 20241223].

European Commission, Directorate-General for Education, Youth, Sport and Culture (2022). Translators on the cover: multilingualism & translation : report of the Open Method of Coordination (OMC) working group of EU Member State experts. *Publications Office of the European Union*. <https://data.europa.eu/doi/10.2766/017> [Accessed: 20240515].

Guerberof-Arenas, Ana; Toral, Antonio (2022). Creativity in translation: Machine translation as a constraint for literary texts. *Translation Spaces*, v. 11, n. 2, pp. 184–212. <https://doi.org/10.1075/ts.21025.gue>. [Accessed 20240515].

Hadley, James Luke (2024). MT and CAT: Challenges, Irrelevancies, or Opportunities for Literary Translation? In: Rothwell, Andrew; Way, Andy; Youdale, Roy (eds.). *Computer-Assisted Literary Translation*. New York: Routledge, pp. 91–105.

Horenberg, Lisa (2019). How CAT may support literary translators in source-text analysis, translation and retranslation: a case study of "Mr loveday's Little Outing". [Master's thesis]. Utrecht University. Utrecht. <https://dspace.library.uu.nl/handle/1874/383930>. [Accessed: 20240515].

Kenny, Dorothy; Winters, Marion (2020). Machine Translation, Ethics and the Literary Translator's Voice. *Translation Spaces*, v. 9, n. 1, pp. 123–149. https://doi.org/10.1075/ts.00024.ken. [Accessed: 20240515].

Klemin, Jeremy (2024). The Last Frontier of Machine Translation. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2024/01/literary-translation-artificial-intelligence/677038/>. [Accessed: 20240515).

Kolb, Waltraud (2017). "It was on my mind all day": Literary translators working from home – some implications of workplace dynamics. *Translation Spaces*, v. 6, n. 1, pp. 27–43. <https://doi.org/10.1075/ts.6.1.02kol>. [Accessed: 20240515].

Kolb, Waltraud (2024). “I Am a Bit Surprised”. In: Rothwell, Andrew;., Way, Andy;. and Youdale, Roy. (eds.). *Computer-Assisted Literary Translation.*, New York: Routledge, pp. 53–68. <https://doi.org/10.4324/9781003357391-4>. [Accessed: 20240515].

Kolb, Waltraud; Miller, Tristan (2022). Human–computer interaction in pun translation. In: Hadley, James; Taivalkoski-Shilov, Kristiina; Teixeira, C. S. Carlos; Toral, Antonio (eds.). *Using Technologies for Creative-Text Translation*. New York: Routledge, pp. 65–88. <https://doi.org/10.4324/9781003094159-4>. [Accessed: 20240515].

Krings, Hans P. (2001). *Repairing texts : empirical investigations of machine translation post-editing processes*. Kent, Ohio & London: The Kent State University Press.

Kuzman, Taja; Vintar, Špela; Arčan, Mihael (2019). Neural Machine Translation of Literary Texts from English to Slovene. In: Hadley, James; Popović, Maja; Afli, Haithem; Way, Andy (eds.). *Proceedings of the Qualities of Literary Machine Translation*. European Association for Machine Translation, pp. 1–9. <https://aclanthology.org/W19-7301>. [Accessed: 20240515].

Kuznetsova Alexandra; Brockhoff Per B.; Christensen, Rune H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, v. 82, n. 13, pp. 1–26. <https://doi.org/10.18637/jss.v082.i13>. [Accessed: 20240515].

Lacruz, Isabel; Denkowski, Michael; Lavie, Alon (2014). Cognitive demand and cognitive effort in post-editing., In: O'Brien, Sharon; Simard, Michael; Specia, Lucia (eds.). *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*. Association for Machine Translation in the Americas, pp. 73–84. <https://aclanthology.org/2014.amta-wptp.6>. [Accessed: 20240515].

Leijten, Mariëlle; Van Waes, Luuk (2013). Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication*, v. 30, n. 3, pp. 358–392. <https://doi.org/10.1177/0741088313491692>. [Accessed: 20240515].

Matusov, Evgeny (2019). The Challenges of Using Neural Machine Translation for Literature. In: Hadley, James; Popović, Maja; Afli, Haithem; Way, Andy (eds.). *Proceedings of the Qualities of Literary Machine Translation*. European Association for Machine Translation, pp. 10–19. <https://www.aclweb.org/anthology/W19-7302>. [Accessed: 20240515].

Moorkens, Joss; Toral, Antonio;, Castilho, Sheila; Way, Andy (2018). Translators' Perceptions of Literary Post-editing Using Statistical and Neural Machine Translation. *Translation Spaces*, v. 7, n. 2, pp. 240–262. < https://doi.org/10.1075/ts.18014.moo>. [Accessed: 20240515].

Noriega-Santiáñez, Laura; and Corpas Pastor, Gloria (2023). Machine vs Human Translation of Formal Neologisms in Literature: Exploring E-tools and Creativity in Students. *Revista Tradumàtica*, v. 21, pp. 233–264. <https://doi.org/10.5565/rev/tradumatica.338>. [Accessed: 20240515].

Novak, B. J. (2014). *One more thing: stories and other stories*. New York: Alfred A. Knopf.

Novak, B. J. (2014). *Onverzameld werk*. Translated by Lodewijks, Jevgenia; Meeder, Lydia; van der Werf, Maarten. Houten-Antwerpen: Agathon.

Oliver, Antoni (2024). Author-Tailored Neural Machine Translation Systems for Literary Works. In: Rothwell, Andrew; Way, Andy; Youdale, Roy (eds.). *Computer-Assisted Literary Translation*. New York: Routledge, pp. 126–141. <https://doi.org/10.4324/9781003357391-9>. [Accessed: 20240515].

R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. [Accessed: 20240515].

Reedz (2023). Swedish AI startup Reedz launches Reedz Books - The world's first fully AI powered publisher. <https://www.mynewsdesk.com/se/reedz-ab/pressreleases/swedish-ai-startup-reedz-launches-reedz-books-the-worlds-first-fully-ai-powered-publisher-3241999>. [Accessed: 20240515].

Rothwell, Andrew (2024). Retranslating Proust Using CAT, MT, and Other Tools. In: Rothwell, Andrew;., Way, Andy;. and Youdale, Roy. (eds.). Computer-Assisted Literary Translation. New York: Routledge, pp.106–125. <https://doi.org/10.4324/9781003357391-8>. [Accessed: 20240515].

Rothwell, Andrew; Way, Andy; Youdale, Roy (eds.) (2024). *Computer-assisted literary translation*. New York: Routledge.

Rudan, Sasha Mile; Kelbert, Eugenia; Kovacevic, Lazar; Reynolds, Matthew; Rudan, Sinisha (2024). Augmenting and Informing the Translation Process through Workflow-Enabled CALT Tools. In: Rothwell, Andrew;., Way, Andy;. and Youdale, Roy. (eds.). Computer-Assisted Literary Translation. New York: Routledge, pp. 258–181. <https://doi.org/10.4324/9781003357391-19>. [Accessed: 20240515].

Ruffo, Paola (2022). Collecting literary translators' narratives. In: Hadley, James; Taivalkoski-Shilov, Kristiina; Teixeira, C. S. Carlos; Toral, Antonio (eds.). *Using Technologies for Creative-Text Translation*. New York: Routledge, pp. 18–39. <https://doi.org/10.4324/9781003094159-2>. [Accessed: 20240515].

Ruffo, Paola (2024). Literary translators in-between: An exploration of their self-imaging discourse and relationship to technology. *Translation in Society*, v. 3, n. 1, pp. 87–103. <https://doi.org/10.1075/tris.23015.ruf>. [Accessed: 20240515].

Şahin, Mehmet; Gürses, Sabri (2021). English-Turkish Literary Translation Through Human-Machine Interaction. *Revista Tradumàtica*, v. 19, pp. 171–203. <https://doi.org/10.5565/rev/tradumatica.284>. [Accessed: 20240515].

Slessor, Stephen (2020). Tenacious Technophobes or Nascent Technophiles? A survey of the Technological Practices and Needs of Literary Translators. *Perspectives*, v. 28,

n. 2, pp. 238–252. <https://doi.org/10.1080/0907676X.2019.1645189>. [Accessed: 20240515].

Terribile, Silvia (2023). Is post-editing really faster than human translation?. *Translation Spaces*. <https://doi.org/10.1075/ts.22044.ter>. [Accessed: 20240515].

The Society of Authors (2023). Artificial Intelligence. <https://www2.societyofauthors.org/where-we-stand/artificial-intelligence/>. [Accessed: 20240329].

Toral, Antonio (2019). Post-editese: an Exacerbated Translationese. In: Forcada, Mikel; Way, Andy; Haddow, Barry; Sennrich, Rico (eds.). *Proceedings of Machine Translation Summit XVII: Research Track*. European Association for Machine Translation, pp. 273–281. <https://aclanthology.org/W19-6627>. [Accessed: 20240515].

Toral, Antonio; van Cranenburgh, Andreas; Nutters, Tia (2024). Literary-adapted machine translation in a well-resourced language pair: Explorations with More Data and Wider Contexts. In: Rothwell, Andrew;., Way, Andy;. and Youdale, Roy. (eds.). Computer-Assisted Literary Translation. New York: Routledge, pp. 27–52. <https://doi.org/10.4324/9781003357391-3>. [Accessed: 20240515].

Toral, Antonio; Wieling, Martijn; Way, Andy (2018). Post-editing Effort of a Novel with Statistical and Neural Machine Translation. *Frontiers in Digital Humanities*, v. 5, n. 9, pp. 1–11. <https://doi.org/10.3389/fdigh.2018.00009>. [Accessed: 20240515].

Van Egdom, Gys-Walt; Kosters, Onno; Declercq, Christophe (2023). The Riddle of (Literary) Machine Translation Quality. *Revista Tradumàtica*, v. 21, pp. 129–159. <https://doi.org/10.5565/rev/tradumatica.345>. [Accessed: 20240515].

Vanmassenhove, Eva; Shterionov, Dimitar; Way, Andy (2019). Lost in Translation: Loss and Decay of Linguistic Richness in Machine Translation. In: Forcada, Mikel;., Way, Andy;., Haddow, Barry;., and Sennrich, Rico. (eds.)., Proceedings of Machine Translation Summit XVII: Research Track. European Association for Machine Translation, pp. 222–232. <https://aclanthology.org/W19-6622>. [Accessed: 20240515].

Vieira, Lucas N., Zelenka, Natalie; Youdale, Roy; Zhang, Xiaochun; Carl, Michael (2023). Translating science fiction in a CAT tool: machine translation and segmentation settings. *Translation & Interpreting*, v. 15, n. 1, pp. 216–235. <https://doi.org/10.12807/ti.115201.2023.a11>. [Accessed: 20240515].

Wang, Hongtao (2023). Defending the last bastion: A sociological approach to the challenged literary translation. Babel. Revue internationale de la traduction / International Journal of Translation. <https://doi.org/10.1075/babel.00330.wan>. [Accessed: 20230515].

Way, Andy; Rothwell, Andrew; Youdale, Roy (2023). Why Literary Translators should embrace Translation Technology. *Revista Tradumàtica*, v. 21, pp. 87–102. <https:/doi.org/10.5565/rev/tradumatica.344>. [Accessed: 20240515].

Webster, Rebecca; Fonteyne, Margot; Tezcan, Arda; Macken, Lieve; Daems, Joke; (2020). Gutenberg Goes Neural: Comparing Features of Dutch Human Translations with Raw Neural Machine Translation Outputs in a Corpus of English Literary Classics. *Informatics*, v. 7, n. 3, 32, pp. 1–21. <https://doi.org/10.3390/informatics7030032>. [Accessed: 20240515].

Youdale, Roy; Rothwell, Andrew (2022). Computer-Assisted Translation (CAT) tools, translation memory and literary translation. In: Deane-Cox, Sharon; Spiessens, Anneleen (eds.). *The Routledge Handbook of Translation and Memory*. New York: Routledge, pp. 381–402. <https://doi.org 10.4324/9781003273417-28>. [Accessed: 20240515].