

## 2 ‘I Am a Bit Surprised’

### Literary Translation and Post-Editing Processes Compared

*Waltraud Kolb*

#### 2.1 Introduction

At the International Literature Festival in Berlin in 2019, a panel composed of two authors/translators and a computer linguist<sup>1</sup> discussed the future role of machine translation (MT) in the literary domain. It was a remarkable early effort to draw the reading public’s attention to recent developments in the field. Towards the end of the event, moderator Gregor Dotzauer, a literary editor at the German newspaper *Tagesspiegel*, singled out Ernest Hemingway as an obvious candidate for MT because he ‘has this Latin clarity’ (Internationales Literaturfestival Berlin, 2019). Hemingway is indeed well known for his straightforward prose, and one of his early short stories also serves as the source text for the explorations of literary post-editing that will be presented in this chapter. Post-editing lies at the very heart of any investigation of the opportunities and challenges engendered by digital technologies in the literary-translation domain as MT is still widely considered ‘unsuitable for translating literature’ (European Commission, Directorate-General for Education, Youth, Sport and Culture, 2022, p. 60), unless followed by human post-editing. So far, we do not know very much about what post-editing actually entails for literary translation, be it literary translation as a process or as a product.

The aim of this chapter is to examine the differences between translation from scratch and post-editing processes in the literary domain from a primarily cognitive perspective and identify correlations with the resulting target texts. In which ways is post-editing a short story by Hemingway different from translating it from scratch? Is it faster, does it involve less typing? Do post-editors reflect on the same issues as translators and go through the same decision-making routines? Do both engage in the same way with Hemingway’s choices? Are post-editors primed by MT suggestions? To explore these questions, I will compare findings from two

1 Karin Tidbeck, Sylvain Neuvel, and Aljoscha Burchardt.

empirical studies: in the first study, five professional literary translators translated Hemingway’s story from scratch and their working and decision-making processes were captured by keylogging and think-aloud (Kolb, 2011, 2013, 2017, 2021); in the second study, using the same research design, five different professional literary translators post-edited a raw MT version of the same short story generated by DeepL.<sup>2</sup>

## 2.2 Related Research

Post-editing (PE) is ‘the activity of revising a text that has been translated automatically by a Machine Translation (MT) system’ (O’Brien, 2021, p. 177), usually with the aim of increasing productivity. In this context, productivity can be measured as the effort that goes into post-editing a machine-translated text, compared to the effort that goes into translating the same text from scratch, with lower effort signifying higher productivity. Krings (2001) proposed three levels to study PE effort that have since been widely applied in research, namely the temporal level (time spent on the task), the technical level (number of keystrokes), and the cognitive level (e.g. number and duration of pauses). Regarding literary texts, Moorkens et al. (2018) and Toral et al. (2018), reporting on the same experiment, found substantial productivity gains for PE (for both statistical MT and neural MT) compared to human translation (HT) for their (English–Catalan) literary-adapted MT systems. Their participants, six professional translators with experience in literary translation, were faster post-editing than translating, generating fewer keystrokes, and PE resulted in fewer but longer pauses (see Moorkens et al., 2018, for the participants’ perception of the task). The same parameters were also used in a later study by Guerberof-Arenas and Toral (2022) involving the language pairs English–Catalan and English–Dutch with four literary translators as participants (see Chapter 1, this volume); their results were mixed, with a lower technical effort for PE for both language pairs but a higher temporal effort for PE compared to HT for Dutch. For both target languages, the average number of pauses was also lower for PE than for HT, which the authors interpreted as an indicator of lower cognitive effort, adding, however, that the higher cognitive effort in HT might indicate that translators “‘think harder” for a more creative solution’ (p. 13). Both studies used keylogging as a tool and based their analysis of cognitive effort on pause patterns. In order to gain more insight into the cognitive aspects of translatorial processes as well as what happens during pauses and how translators and post-editors make their decisions, Krings (2001) and Vieira (2015)—both

2 [www.deepl.com/en/translator](http://www.deepl.com/en/translator).

dealing with non-literary texts—used think-aloud protocols (TAPs). In the literary domain, TAPs have occasionally been used to explore HT processes (Jones, 2011; Kolb, 2011, 2013, 2017, 2021; Borg, 2022), but, to the best of my knowledge, the present study is the first to use such verbal data for the study of literary PE processes.

Effort and productivity need to be assessed against questions of quality. Regarding non-literary texts, numerous studies have found PE to produce comparable, or sometimes even better, quality than HT (Koponen, 2016; Daems et al., 2017). A central issue in exploring PE quality is the question of whether post-edited texts exhibit characteristics of ‘post-editeese’ (see Section I.3, Introduction, this volume), a term introduced by Daems et al. (2017) to describe ‘the expected unique characteristics of a PE text that set it apart from a translated text’ (p. 90). So far, research has yielded mixed results. While Daems et al. (2017) have not found any indication of post-editeese in their study of (English–Dutch) translated and post-edited newspaper articles, a number of other studies did identify post-editeese in PE output. Čulo and Nitzke (2016) compared the two modes in terms of terminology (English–German) and found that there was less variation in PE than in HT. Farrell (2018) also found lower variation in PE as well as normalization and homogenization phenomena (English–Italian). In a study involving five languages (English, German, French, Spanish, and Chinese), Toral (2019) also reported normalization of the target language in PE, in addition to simplification and more interference from the original than in HT, so that he even speaks of post-editeese as ‘exacerbated translationese’ (p. 273). Volkart and Bouillon (2022) confirmed some of these features for an authentic corpus of human-translated and post-edited press releases issued by the European Investment Bank.

Regarding literary PE, Castilho et al. (2019) compared the presence of post-editeese in the news and literary domains for the language-pair English–Brazilian Portuguese and found the level of post-editeese to be higher for literature. In a subsequent study, Castilho and Resende (2022a) analyzed features of post-editeese in literary texts, using a range of parameters such as lexical richness (see e.g. Vanmassenhove et al. (2019) on the loss of lexical diversity in MT), lexical density, sentence length, punctuation, explicitation, personal pronouns, and convergence between modalities. They found that PE output was similar to MT output in terms of lexical density, use of pronouns, and sentence length, with mixed results for the other categories, also depending on the type of literary text. They interpreted their results as an indication of priming effects through MT (see also Castilho and Resende, 2022b). That MT priming has a substantial impact on a post-editor’s textual voice has also been shown by Kenny and Winters (2020; see also Chapter 3 in the present volume). They asked Hans-Christian Oeser, a well-known German translator of English

literature, to post-edit a chapter of a novel that he had translated from scratch two decades earlier and found that his voice—as reflected in his stylistic choices—is less manifest in his post-edited text than in his HT. Macken et al. (2022) designed a three-stage study in which an English source text was first machine-translated into Dutch, the raw MT was then post-edited by a professional literary translator, and, as a third step, this version was revised (mainly monolingually) by a different translator. The results show that the MT and the post-edited version were more similar to each other than the post-edited version and the revision, and that more editing occurred during revision than during PE, with the reviser's edits primarily aiming at improving readability and acceptability for target readers.

Guerberof-Arenas and Toral (2020) measured creativity as an indicator of quality in HT, MT, and PE for the language pair English–Catalan, with two professional literary translators as participants. To determine creativity, they looked at acceptability (number of errors) and novelty (creative shifts). Regarding acceptability, they found that PE scored best for accuracy and HT for fluency; regarding novelty, HT scored higher than PE, indicating that PE is constrained by the MT output. In a subsequent study, Guerberof-Arenas and Toral (2022) used a similar design but added the language pair English–Dutch. They describe their results as a ‘cautionary tale’ for using MT, ‘because the translator becomes the evaluator and not the creator’ (p. 207). That the post-editor's role as evaluator (see also Chapter 7 in the present volume) has a profound impact on decision-making processes in PE will also become apparent below.

### 2.3 Research Design

In a study of literary translation processes carried out in 2009/2010 (Study 1), five literary translators (four female, one male; henceforth referenced as T–1 to T–5) translated a short story by Ernest Hemingway into German. Their processes were captured by keylogging and think-aloud, product data comprised first drafts, interim, and final versions; information about the participants' educational and professional backgrounds and work routines was collected through questionnaires (Kolb, 2011, 2013, 2017, 2021). To explore literary PE and compare it to HT, a follow-up study was carried out in 2020/2021 (Study 2), in which five different literary translators (all female: PE–1 to PE–5) post-edited a DeepL version of the same story, using the same type of tools. DeepL was selected as a general-domain neural MT system as it is widely used in German-speaking countries and would, in practice, be a likely choice for literary translators (who would usually not be in a position to train their own MT systems; see Section I.2,

Introduction, this volume, for more on MT system training, tuning, and testing).

The participants of both groups were experienced literary translators who had been working as such between eight and 29 years at the time of the respective study. It is important to note that none of the participants in Study 2 had any previous post-editing experience, which is still typical of the literary translation field at this point. The participants of both groups worked in their usual work environments: in most cases, a space in their homes, and in three cases an office away from home. They were asked to follow their usual work routines as far as possible and work on the target text until they felt it would be ready to be sent to a publisher. Both groups were asked to think aloud as much as possible while working on their task, verbalizing their thoughts and deliberations; Audacity<sup>3</sup> was used as the voice-recording software. They were also asked to record retrospectively any thoughts they had while they were not at their desk but, for example, were loading the dishwasher. To capture their keystrokes, Group 1 used Translog (Jakobsen & Schou, 1999), and Group 2 Inputlog (Leijten & Van Waes, 2013). Inputlog was used in Study 2 because some participants in Group 1 had experienced technical problems with Translog. Both tools log the same type of data of interest, most notably, insertions, deletions, and pauses. Conditions for the participants were comparable so that the choice of tool did not affect their work: Translog users worked in a target text window that resembles common text editors, Inputlog users in an MS Word window; and both tools have the advantage that texts are not split into segments, making it easier to keep one's eye on document-level narrative and stylistic features.

The source text was a short story by Hemingway appropriately entitled 'A Very Short Story' (1925) about a man and a woman who meet in a hospital in Padua in northern Italy at the end of WWI. The unnamed male protagonist, an American soldier, is a patient there; the female protagonist Luz is (probably) a nurse. The story is 637 words long, including the title, which makes it manageable in a research setting and still has the advantage of being a complete literary piece so that the participants could take into account document-level features. Despite Hemingway's straightforward language (low lexical variation, preference for simple and common words, short, paratactic sentences), the raw MT output contained a number of errors. The types of errors are well-known problems in MT research: disambiguation of polysemous words, orthographically similar words, co-reference resolution, verb tense, register, literality, and omission of

source text material. The purpose of this chapter, however, is not to evaluate DeepL's performance, but to explore PE processes.

## 2.4 Data and Discussion

In my analysis of HT and PE processes, I will first briefly compare the temporal and technical effort (i.e. task time and number of keystrokes) associated with the two modalities, and then focus on the underlying cognitive dimension. Unlike some studies mentioned above, I will not look at pauses but will explore decision-making processes and the role played by MT priming. Priming effects have been shown to occur on multiple levels (Farrell, 2018; Toral, 2019; Kenny & Winters, 2020; Castilho & Resende, 2022a); in this chapter, examples of MT priming will be discussed that relate to terminology, interpretation, normalization, and literalism.

### 2.4.1 *Temporal and Technical Effort*

For the purposes of this study, temporal effort is measured as the total time of the verbal record. Participants were asked to switch on Audacity whenever they worked on the task so that the verbal record includes time spent working at the computer (including silent periods) as well as periods they spent on manually revising the printout of an interim target version. It does not cover some preliminary task-related activities of unknown length, such as setting up the tools, or periods in which they reflected on task-related issues away from their desk, e.g. during breakfast. However, whenever participants at a later point summarized such deliberations for the record, these brief sequences of retrospective verbalization are included in their total time count. On three occasions, translators put on record their desire to pause the recording and read the target text silently; these periods are also not included. Since the participants worked in their authentic working environments rather than in a supervised lab, some technical hiccups occurred which also impacted time measurements: three participants from Group 1 (T-1, T-2, T-3) had to re-type passages that had not been saved by Translog; these periods have been subtracted from their total times (see Kolb, 2017 for more details) so as to not skew correlations with the respective keystroke numbers. Given the above limitations, the time measurements in Table 2.1 are approximate values only but still serve to indicate general tendencies for the ten participants.

As shown in Table 2.1, the post-editors spent on average less time on the task than the translators (approx. 24% less, a number that is slightly lower than results from other studies, e.g. Toral et al. (2018)), with the median being only approx. 10% lower. However, a closer look reveals that there are significant differences between individual participants. Inter-subject variation is particularly great in Group 1, ranging from 1 hour

*Table 2.1* Total time of verbal record and total number of keystrokes for translators and post-editors, CHARCUT for post-editors.

<i>Part.</i>	<i>Total time of verbal record (hrs:min)</i>	<i>Total number of keystrokes</i>	<i>CHARCUT</i>
T-1	5:00	9750	
T-2	2:22	6091	
T-3	2:29	6991	
T-4	3:54	10247	
T-5	1:35	6379	
Average	3:04	7892	
Median	2:29	6991	
PE-1	2:03	7931	0.2816
PE-2	2:14	3464	0.3137
PE-3	2:40	5486	0.3195
PE-4	2:55	4583	0.2577
PE-5	1:46	5787	0.3656
Average	2:20	5450	0.3076
Median	2:14	5486	0.3137

and 35 minutes (T-5) to 5 hours (T-1). The participants in Group 2 spent between 1 hour and 36 minutes (PE-5) and 2 hours and 55 minutes (PE-4) on the PE task, their temporal effort thus varying to a lesser extent, with no outliers.

The total number of keystrokes was approx. 31% lower on average for the PE modality, the median approx. 21.5%. Comparing individual task times and keystroke numbers shows that working styles differed significantly between participants in both groups. For instance, T-1 spent by far the longest time, but did not produce the highest number of keystrokes; the greatest technical effort in terms of keystrokes was T-4's, whose total working time was more than an hour shorter than T-1's. In Group 2, PE-1 and PE-5 spent the shortest time but produced the most keystrokes, and PE-4 spent the longest time, while her number of keystrokes is in the lower range. This indicates that some participants spent more time deliberating without typing than others, or typed and then deleted more candidate solutions for a word or phrase than others. In all, individual speed and working style seem to have had less of an impact in PE than in HT.

Another measure to assess PE effort is the difference (or similarity) between the raw MT output and the final post-edited version, which can be determined by a number of different metrics. For the purpose of this chapter, CHARCUT (Lardilleux & Lepage, 2017; see also Macken et al., 2022) was used, which is a character-based metric that counts the cost of PE in terms of deletions, insertions, and shifts. As Table 2.1 shows, the

differences between MT and post-edited versions range from 0.2577 (PE-4) to 0.3656 (PE-5). As above, the results reflect different working styles: for instance, PE-4's final target text is the most similar to the raw MT output, even though her technical effort was not the lowest and she spent the longest time on the task; on the other end of the spectrum is PE-5, whose target text is the most different from the MT output, even though she was the fastest and her number of keystrokes was not the highest. PE-1, who did by far the most typing, submitted a target text that ranges second in terms of similarity with the MT output.

If we look at the sentence level, the participants edited between 35 and 38 out of the story's total of 40 sentences (87.5%—95%); even more significantly, there was not a single sentence where all five post-editors agreed that it could be accepted without any edits. Based on my own evaluation of accuracy (correct content) and fluency (orthographic, grammatical, idiomatic correctness), 15 out of the 40 sentences (37.5%) would not have needed any editing, a percentage that is in line with findings in other studies (e.g. Macken et al., 2022). This seems to indicate that the achievement of accuracy and fluency is not enough for professional literary translators to feel comfortable putting their name to the final product (see also, e.g. Moorkens et al., 2018). What goes beyond issues of accuracy and fluency are, for instance, cultural considerations, stylistic preferences, the interpretation of ambiguous language, and considerations of loyalty towards the source text author's choices or the target readers.

#### 2.4.2 *Cognitive Processes*

The quantitative data described above, i.e. temporal and technical effort or inter-text similarity, are shaped by the underlying cognitive processes. In the case of PE, these processes relate to either the source text, the raw MT output, or the target text, whereas with HT the cognitive load is distributed between two texts only rather than three. Studies on non-literary PE have shown that the distribution of attention tends to differ between the two modalities, with the source text usually receiving less attention in PE than in HT (Krings, 2001; Mesa-Lao, 2014). This is hardly surprising, given the fact that in HT drafting the first version is usually the phase in which translators most intensely engage with the source text, its interpretation, its style, and the author's choices; in PE, the first draft—and with it, the first interpretation of the source text—is supplied by the MT engine.

The following discussion of cognitive processes and the role played by MT priming will draw on data from the participants' verbal reports. TAPs include various kinds of verbalization, ranging from emotional reactions (sighing, laughing) to reading out loud snippets of source or target text or raw MT, from disjointed sequences of thinking aloud to explanations or rationalizations addressed to the researcher (Ericsson & Simon, 1993).



They contain directly or indirectly articulated indicators of uncertainty, i.e. a cognitive state of indecision on the part of the translator or post-editor and subsequent problem-solving and decision-making behavior (Angelone, 2010).

Examples 1 and 2 show in which ways the participants' engagement with the source and target texts differed between HT and PE, how their problem-solving routines differed, and how the post-editors were primed by DeepL's suggestions. The resulting post-edited target texts are more similar to each other than those translated from scratch. Both examples are taken from the opening paragraph of Hemingway's story (relevant passages in italics):

One hot evening in Padua they carried him up onto the roof and he could look out over the top of the town. There were *chimney swifts* [Example 1] in the sky. After a while it got dark and the searchlights came out. The others went down and took the bottles with them. He and Luz could hear them below on the balcony. *Luz sat on the bed* [Example 2]. She was cool and fresh in the hot night.

(Hemingway, 1925, p. 83)

Example 1 relates to terminology. The German translation of *chimney swift* provided by DeepL was *Schornsteinsegler*, one of two correct German terms for the bird in question (Lat. *chaetura pelagica*). All post-editors retained the MT suggestion. As a first step, they all double-checked it in a Google/Wikipedia search or an online dictionary. Three of the post-editors were quickly satisfied that DeepL had supplied the correct German term. Two of the post-editors (PE-1 and PE-5) realized that there was a discrepancy between the birds' habitat and the story's setting, as *chimney swifts* live in the Americas, but not in Europe. PE-5 quickly laughed off her initial surprise, proceeding immediately to the next segment. PE-1 was the only one in Group 2 who briefly reflected on the implications (translations from the German TAPs are mine, the words in italics are English in the original TAPs):

PE-1: it is somehow written from the point of view of an American soldier who maybe thinks they are chimney swifts, and it fits as they are sitting on the roof ... so I think we can leave it as it is ... also because I think the image is nice

PE-2: I have to look up those Schornsteinsegler, what exactly that is, *chimney swifts in the sky*, if I am honest I can't imagine what that is in English [...] let's see if DeepL is actually right here [...] so I just put it into Google [...] *chimney swift*, Schornsteinsegler, aha, they are indeed Schornsteinsegler

PE-3: these *chimney swifts*, Schornsteinsegler, I can't imagine that this is correct and therefore I now open dict.cc on my computer [...] *chimney swifts*, ok, they exist and they are indeed Schornsteinsegler, which means this is a bird

PE-4: let's look up *chimney swifts*, I have never heard Schornsteinsegler, but this does not have to mean anything, *chimney swifts*, I am on dict.cc now, let's see [...] good, so DeepL has identified this all right, this is a bird

PE-5: should I check now if this is indeed a bird [...] yes [...] but I am a bit surprised now that it flies around in Italy, even though it is an American swift, but ok [laughs]

While the post-editors were primarily focused on verifying the terminological correctness of the MT suggestion, the translators engaged in extended research when drafting their first target versions. Having described these deliberations in some detail previously (Kolb, 2021), I will simply summarize salient points here. With one exception, the translators based their decisions on either a careful reflection on the implications of the habitat/setting discrepancy or the different images that would be evoked by different German options. T-1 and T-2, for instance, both spent considerable time weighing the pros and cons of putting an American bird into an Italian story; eventually, T-1 decided to replicate Hemingway's choice (*Schornsteinsegler*), arguing that Hemingway 'imported ... his images, and that he simply saw in the sky the birds that he knew, no matter whether they were there or not'. In contrast, T-2 concluded that 'an American swift ... in Padua ... this is nonsense, Hemingway knew chimney swifts from back home' and therefore decided to use a different bird from the swift family that lives in Europe (*Mauersegler*). For T-3 and T-4, the determining factor was the image (briefly touched upon by only one of the post-editors, PE-1). T-3 reflected at length on the different images evoked by the two German terms that exist for the American bird, i.e. *Kaminsegler* and *Schornsteinsegler* (like the English term, they are compounds with the first element denoting different kinds of chimneys that evoke different images) and then opted for *Kaminsegler*. T-4 also based her decision on the visual image, appreciating the 'beautiful image' created by *Mauersegler*. T-5 found the term *Rauchschwalbe* (*barn swallow* in English) in a dictionary search and was quickly satisfied, given that 'a swift is a bird'; coincidentally, she chose a bird that lives on both continents. We thus encounter four different birds (two European, two American, and one European-American) in the five translations, compared to just one (the American) in the five post-edited versions.

In Example 2, the post-edited target texts are also identical, while the translations again vary. Here, however, priming occurred on the level of interpretation. Ambiguity is a typical feature of literary writing, including Hemingway's, notwithstanding his 'Latin clarity' (Internationales Literaturfestival Berlin, 2019), highlighting the translator's role in interpretation and meaning construction. As Hermans (2007, p. 30) phrased it, different translations 'flesh out the interpretive potential of a given text'. The original sentence *Luz sat on the bed* is ambiguous in that the verb *sat* can be read in two ways, either as the action of sitting down (*setzte sich* in German) or as the state of being seated (*saß* in German). Both interpretations make perfect sense in the context of the story, and indeed we find both options in the translations (*setzte sich* in two, *saß* in three translations), whereas all five post-edited target texts present the same scene, in all cases the one suggested by DeepL (*setzte sich*). PE-1 was the only post-editor who considered the possibility of an alternative reading but ultimately found DeepL's version 'acceptable'; the verbal data and keylogs of the other four post-editors do not contain any indication that they were aware of this interpretative potential or questioned the MT suggestion in any way. In Group 1, two translators automatically, without any obvious hesitation, chose the second option, while T-1 and T-2 explicitly reflected on the ambiguity, T-1 even revising her choice twice before settling on her final interpretation. In T-4's audio-recording we can hear a slight hesitation before she typed in the sentence, which may be taken as an indicator of some state of cognitive uncertainty, though it was obviously not strong enough to prompt further deliberations.

Example 3 is taken from a later part of the story and serves to illustrate two more instances of priming, now regarding syntax normalization and literalism, both indicative of post-editeese (italics added):

Original: When they had to say good-bye, *in* the station at Milan, they kissed good-bye, but were not finished with the quarrel. (Hemingway, 1925, p. 84)

DeepL: Als sie sich *im* Bahnhof von Mailand verabschieden mussten, verabschiedeten sie sich, waren aber mit dem Streit noch nicht fertig. [English gloss: When they *in* the station of Milan had to say good-bye, they said good-bye, but were not yet done with the quarrel.]

DeepL normalized the syntax of the first part of the sentence, shifting the phrase *in the station at Milan* to the front of *had to say good-bye*. In German, retaining the original's syntax is possible but results in a slightly unusual rhythm; four of the five translators opted for it, while four of the five post-editors retained the normalized German syntax suggested by

DeepL, again, without any hesitation. Of some interest is also the preposition *im* (*in the*). Both in English and German, one could use either *in the station* (*im Bahnhof*) or *at the station* (*am Bahnhof*), and in both languages, the second option leaves more room for interpretation (where exactly do they say good-bye<sup>4</sup>: in front of the building, inside the building, by the train's door on the platform?). In both languages, the second option would also be the more common way to describe the scene (we can only speculate that Hemingway might have opted for *in* here to avoid another *at* in the same phrase). The MT suggestion (*im*) is a literal translation and was retained by four post-editors (the exception being PE-5), while three of the five translators opted for the less literal and more common (and ambiguous) *am* (*at the*). The following excerpt from PE-3's TAP is particularly interesting (italics highlight the words of interest and do not indicate any change in intonation):

PE-3: [reading German MT:] als sie sich *am, im* Bahnhof von Mailand verabschiedeten, mussten ... verabschiedeten sie sich [laughs], [reading English original:] they kissed good-bye, [translating:] küssten sie sich [...] [reading German MT:] *im Bahnhof*, well, ok ... well now we have one, two, three saying good-bye, one of them completely wrong

We can hear how PE-3 stumbles when she first reads out loud the German MT version, subconsciously correcting the MT output into the more common *am* (*at the*); she then immediately corrects herself and reads what is actually on her screen (*im*); a second later she seems to experience another moment of uncertainty ('*im Bahnhof*, well, ok'), but does not react in any way to it, probably because the MT error (omission of *kissed*) monopolizes her attention. As Krings (2001) has shown, post-editors 'tend to intervene correctively in the machine translation subconsciously during reading and to recognize the correct (and thus familiar) well-formed nature of a text even where it is not present at all' (p. 364). And indeed, when reading out loud the raw MT output or their various interim or final versions, all post-editors (with the exception of PE-4) subconsciously switched several times between *im* (*in the*) and *am* (*at the*).

## 2.5 Conclusion

Post-edited target texts are synthetic texts that are part human-generated and part machine-generated, containing both edited and unedited material. In the examples discussed above, PE resulted in target texts that contain distinct traces of priming through MT and are more similar to

4 Hemingway's spelling.

each other than the translations from scratch. Even though the post-editors in this study went well beyond securing accuracy and fluency, their target texts still do not 'flaunt [their] identity through the difference with other translative interpretations' (Hermans, 2007, p. 31) the same way translations do; put differently, the voices of the post-editors are less manifest in the target texts than those of the translators (see also Kenny and Winters, 2020).

On average, PE was found to be faster than HT and involved fewer keystrokes; however, inter-subject variation regarding speed and working style was considerable in both groups of participants. Regarding the underlying cognitive processes associated with the two modalities, different patterns emerged. In line with the post-editors' role as evaluators (Guerberof-Arenas & Toral, 2022), their attention was primarily on the raw MT output, e.g. verifying terminology (*chimney swifts*) or identifying and correcting MT errors (omission of *kissing*); at the same time, their engagement with the source text and its author's choices (and on occasion also their final target texts, e.g. regarding imagery) was less extensive and remained more superficial than in the case of the translators. This is closely linked to priming effects, which occurred on multiple levels in the PE modality, be it on the level of terminology (*chimney swifts*), interpretation (*sat*), normalization or literalism (*in the station*). Some of these priming effects were more subtle than others, but in each case, several of the post-editors were primed by the same MT suggestion. Together with the fact that priming occurred on so many levels despite Hemingway's rather straightforward language, this clearly points to the great impact of MT stimuli in a PE task, an impact which can be assumed to be even greater with stylistically more demanding source texts. While individual post-edited target texts are not necessarily of lower quality, MT priming seems to be an intrinsic element of PE processes. Priming effects are here shown to constrain the post-editor's agency and lead to more homogeneous target texts. It is likely that, in the long run, they will also lead to various impoverishments of translated literary language (see also Toral, 2019; Vanmassenhove et al., 2019).

As the participants in this study had no prior PE training or experience, further studies will be needed to ascertain whether priming would have less impact for more experienced post-editors. As far as research tools go, TAPs provide very rich data for learning more about how HT and PE differ in terms of translatorial cognition and decision-making; in future studies, eye-tracking might yield additional insights into patterns of attention distribution and cognitive rhythms. We can assume that at least in some sectors of the book market, MT with subsequent human PE will become more widespread before long (see also Chapter 7 in the present volume). Therefore, it seems of paramount importance that both literary translators

who take on PE work and publishers who commission it are aware of the potential implications entailed by the specificities of the cognitive processes associated with literary PE.

## References

- Angelone, E. (2010). Uncertainty, uncertainty management and metacognitive problem solving in the translation task. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 17–40). Benjamins. <https://doi.org/10.1075/ata.xv.03ang>
- Borg, C. (2022). *A literary translation in the making: A process-oriented perspective*. Routledge. <https://doi.org/10.4324/9781003150909>
- Castilho, S., & Resende, N. (2022a). Post-editeuse in literary translations. *Information* 13(2), 66. <https://doi.org/10.3390/info13020066>
- Castilho, S., & Resende, N. (2022b). MT-pese: Machine translation and post-editeuse. *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation* (pp. 305–306). <https://lt3.ugent.be/media/uploads/eamt2022/proceedings-eamt2022.pdf>
- Castilho, S., Resende, N., & Mitkov, R. (2019). What influences the features of post-editeuse? A preliminary study. *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology* (pp. 19–27). [https://doi.org/10.26615/issn.2683-0078.2019\\_003](https://doi.org/10.26615/issn.2683-0078.2019_003)
- Čulo, O., & Nitzke, J. (2016). Patterns of terminological variation in post-editing and of cognate use in machine translation in contrast to human translation. *Baltic Journal of Modern Computing* 4(2), 106–114. <https://aclanthology.org/W16-3401/>
- Daems, J., De Clercq, O., & Macken, L. (2017). Translationese and post-editeuse: How comparable is comparable quality? *Linguistica Antverpiensia* 16, 89–103. <https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/434>
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis. Verbal reports as data*. MIT Press.
- 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>
- Farrell, M. (2018). Machine translation markers in post-edited machine translation output. *Proceedings of the 40th Conference Translating and the Computer* (pp. 50–59). <https://www.asling.org/tc40/wp-content/uploads/TC40-Proceedings>
- 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. <https://doi.org/10.1075/ts.20035.gue>
- 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>

- Hemingway, E. (1925). A very short story. In E. Hemingway (Ed.), *In our time* (pp. 83–85). Scribner.
- Hermans, T. (2007). *The conference of the tongues*. Routledge. [https://www.academia.edu/31205772/Theo\\_Hermans\\_2007\\_The\\_Conference\\_of\\_the\\_Tongues\\_Manchester\\_UK\\_St\\_Jerome\\_Publishing\\_Pages\\_190](https://www.academia.edu/31205772/Theo_Hermans_2007_The_Conference_of_the_Tongues_Manchester_UK_St_Jerome_Publishing_Pages_190)
- Internationales Literaturfestival Berlin. (2019). *Potentialities of machine translation*. Retrieved December 9, 2022, from <https://www.youtube.com/watch?v=7MTHjYvXR8>
- Jakobsen, A. L., & Schou, L. (1999). Translog documentation. In G. Hansen (Ed.), *Probing the process in translation: Methods and results* (pp. 151–186). Samfundslitteratur. <https://research.cbs.dk/en/publications/translog-documentation-version-10>
- Jones, F. R. (2011). *Poetry translating as expert action: Processes, priorities and networks*. Benjamins. <https://doi.org/10.1075/btl.93>
- 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>
- Kolb, W. (2011). The making of literary translations: Repetition and ambiguity in a short story by Ernest Hemingway. *Across Languages and Cultures* 12(2), 259–274. <https://akjournals.com/view/journals/084/12/2/article-p259.xml>
- Kolb, W. (2013). 'Who are they?': Decision-making in literary translation. In C. Way, S. Vandepitte, R. Meylaerts, & M. Bartlomiejczyk (Eds.), *Tracks and treks in translation studies* (pp. 207–221). Benjamins. <https://doi.org/10.1075/btl.108.11kol>
- Kolb, W. (2017). 'It was on my mind all day': Literary translators working from home — Some implications of workplace dynamics. *Translation Spaces* 6(1), 27–43. <https://doi.org/10.1075/ts.6.1.02kol>
- Kolb, W. (2021). 'Hemingway's priorities were just different': Self-concepts of literary translators. In K. Kaindl, W. Kolb, & D. Schlager (Eds.), *Literary translator studies* (pp. 107–121). Benjamins. <https://doi.org/10.1075/btl.156.05kol>
- Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *Journal of Specialised Translation* 25, 131–148. [https://www.jostrans.org/issue25/art\\_koponen.pdf](https://www.jostrans.org/issue25/art_koponen.pdf)
- Krings, H. P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent State University Press. <https://doi.org/10.7202/008026ar>
- Lardilleux, A., & Lepage, Y. (2017). CHARCUT: Human-targeted character-based MT evaluation with loose differences. *Proceedings of the 14th International Workshop on Spoken Language Translation*, 146–153. <https://aclanthology.org/2017.iwslt-1.20/>
- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze writing processes. *Written Communication* 30(3), 358–392. <https://doi.org/10.1177/0741088313491692>
- Macken, L., Vanroy, B., Desmet, L., & Teczan, A. (2022). Literary translation as a three-stage process: Machine translation, post-editing and revision. *Proceedings*

- of the 23rd Annual Conference of the European Association for Machine Translation (pp. 101–110). <https://aclanthology.org/2022.eamt-1.13>
- Mesa-Lao, B. (2014). Gaze behaviour on source texts: An exploratory study comparing translation and post-editing. In S. O'Brien, L. Winther Balling, M. Carl, M. Simard, & L. Specia (Eds.), *Post-editing of machine translation: Processes and applications* (pp. 219–245). Cambridge Scholars Publishing. <https://research.cbs.dk/en/publications/gaze-behaviour-on-source-texts-an-exploratory-study-comparing-tra>
- 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.18014.moo>
- O'Brien, S. (2021). Post-editing. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of translation studies* (vol. 5, pp. 177–183). Benjamins. <https://doi.org/10.1075/hts.5.pos3>
- Toral, A. (2019). Post-editeese: An exacerbated translationese. *Proceedings of MT Summit XVII, Volume 1: Research Track, Dublin, Ireland* (pp. 273–281). European Association for Machine Translation. <https://aclanthology.org/W19-6627/>
- Toral, A., Wieling, M., & Way, A. (2018). Post-editing effort of a novel with statistical and neural machine translation. *Frontiers in Digital Humanities*, 5 (Art. 9), 1–11. <https://doi.org/10.3389/fdigh.2018.00009>
- Vanmassenhove, E., Shterionov, D., & Way, A. (2019). Lost in translation: Loss and decay of linguistic richness in machine translation. *Proceedings of MT Summit XVII, Volume 1: Research Track, Dublin, Ireland* (pp. 222–232). European Association for Machine Translation. <https://aclanthology.org/W19-6622.pdf>
- Vieira, L. N. (2015). *Cognitive effort in post-editing of machine translation: Evidence from eye movements, subjective ratings, and think-aloud protocols* [Doctoral dissertation]. Newcastle University]. <https://core.ac.uk/download/pdf/173983801.pdf>
- Volkart, L., & Bouillon, P. (2022). Studying post-editeese in a professional context: A pilot study. *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation* (pp. 71–79). <https://lt3.ugent.be/media/uploads/eamt2022/proceedings-eamt2022.pdf>