Katharina WALTER
University of Innsbruck

Creativity in Human and AI-Enhanced Literary Translation: A Keylogging Experiment

Human-Centred AI in the Translation Industry.
Questions on Ethics, Creativity and Sustainability

Katharina Walter,
Marco Agnetta
[eds.]

5.1/2025
Yearbook of Translational Hermeneutics
Jahrbuch für Übersetzungshermeneutik

Journal of the Research Center
Zeitschrift des Forschungszentrums

Hermeneutics and Creativity, University of Leipzig
Hermeneutik und Kreativität, Universität Leipzig

DOI: 10.52116/yth.vi1.106

Cite this article:
Walter, Katharina (2025): "Creativity in Human and AI-Enhanced Literary Translation: A Keylogging Experiment." In: *Yearbook of Translational Hermeneutics* 5.1: *Human-Centred AI in the Translation Industry. Questions on Ethics, Creativity and Sustainability* (ed. by Katharina Walter, Marco Agnetta), pp. 197–224.
DOI: <10.52116/yth.vi1.106>.

Yearbook of Translational Hermeneutics 5.1/2025
ISSN: 2748-8160 | DOI: 10.52116/yth.vi1.106

Creativity in Human and AI-Enhanced Literary Translation: A Keylogging Experiment

Katharina WALTER
Universität Innsbruck

Abstract: Based on a recent experiment with MA students at the University of Innsbruck, Austria, this article analyzes the benefits and drawbacks of literary post-editing. Using the keylogging software Inputlog, six experiment participants documented not only the product but also the step-by-step process that led to their final German translation of a short prose poem by Virginia Woolf entitled "Green," which was originally published in 1921. Some students worked only with monolingual and bilingual online dictionaries, while others post-edited an automated translation generated with the DeepL next-generation language model, which was launched in September 2024 and is based on a large language model infrastructure. Echoing previous research on literary post-editing, the experiment results show that AI-enhanced literary translation may entail a loss of creativity. At the same time, as the translation of Woolf's prose poem with its novel images and expressions requires translation strategies that move underneath the textual surface level, the time gains that come with partial automation are negligible at best.

Keywords: Literary post-editing, Creativity, Keylogging, Translator training, DeepL next-generation language model.

1 Introduction

This article compares creativity in literary post-editing and human translation based on a keylogging experiment carried out in January 2025 with six MA students at the University of Innsbruck, Austria. The key aim of this experiment was to examine whether the post-edited German versions were as creative as the human translations of "Green," a short literary text by Virginia Woolf originally published in 1921 as part of a collection of short prose texts called *Monday or Tuesday*. In this context, translatorial creativity is defined as manifesting in "creative shifts" or deliberate changes made to the structure of a source text in translation (for further information about this term, see Bayer Hohenwarter 2011: 669; Guerberof-Arenas/Toral 2022: 186 and 2024: 220), as opposed to "reproductive" translations (see Bayer-Hohenwarter 2011: 674).

The use of keylogging data collected with the open-source software Inputlog (see Leijten/Van Waes 2013) as part of this experiment allows for an examination not only of the outcomes but also of the respective processes underlying each post-edited version and human translation of Woolf's literary sketch. During the experiment, Inputlog recorded each single keystroke, deletion and mouse click, together with a time stamp for the beginning and end of each of these actions respectively. Additionally, the students' screens were recorded with OBS to back up and complement the keylogging data. While particularly Kolb (2021 and 2023) has recurrently used keylogging data to examine both machine-enhanced and human translation, other potentially useful research methods to document translation processes include think-aloud protocols, screen recording and questionnaires (see Angelone 2025). Due to its comprehensiveness and neutrality, keylogging is a particularly effective tool for analyzing creative processes, especially

when it is combined with other documentation methods, such as screen recording or questionnaires.

Overall, the results from this experiment echo previous research on post-editing compared to human translation, showing that time gains in AI-enhanced literary translation, if they exist at all, may come at the costs of a decline in translatorial creativity and in the joyfulness of the activity of translation. It will also be demonstrated that Woolf's text manifests a high density of "units of creative potential" (UCPs), defined by Guerberof-Arenas/Toral (2022: 191) as "units that are expected to require translators to use problem-solving skills, as opposed to those that are regarded as routine units." The keylogging data indicate that, overall, the post-editing group spent less time actively working on their translation and produced less creative translations than the group that worked without machine assistance. This suggests that for texts rich in UCPs, such as Woolf's prose poem, automated pre-translation may ultimately be counterproductive.

In chapter two of this article, the technological framework and key terminology for this project are outlined before the machine output is compared to a recently published human reference translation by Christel Kröning (Woolf 2021). Afterwards, this chapter zooms in on three segments from each post-edited and human-translated version of Woolf's prose poem produced in class. In chapter three, the insights gained from this keylogging experiment will be evaluated with an eye toward potential future avenues for translator training.

2 Experiment Design

Before comparing Woolf's 1921 prose poem to Kröning's 2021 German version and the automated translation used in

the 2025 classroom experiment, this article outlines some preliminary technological and terminological information.

2.1 Technology: The DeepL Next-generation Language Model

In a course on literary translation from English to German, six MA students pursuing a degree in translation studies at the University of Innsbruck transferred Woolf's literary sketch "Green" (1921) from English into German, with the overall aim of achieving equivalent effect in translation. Three of the experiment participants were not allowed to rely on automated translation and had to limit themselves to using specific monolingual and bilingual online dictionaries. *Duden* (n.d.) was recommended for monolingual queries regarding the German language, while for searches on bilingual and monolingual English entries the experiment participants were allowed to alternate between *Cambridge* (n.d.) and *Collins* (n.d.). The other three MA students had to work with both the source text and a pretranslated machine draft that was generated with the DeepL next-generation language model (henceforth shortened to DeepL next-gen). In addition, the three dictionaries the human translators were supposed to use were also recommended for this group.

DeepL next-gen was launched in September 2024 for some language combinations, including English and German. It has since been expanded to include all the languages the DeepL classic language model (DeepL classic) supports, 30 in total, as well as Hebrew, Vietnamese, Latin American Spanish and Thai. Apparently, this new language model "is powered by a large language model (LLM) infrastructure" (DeepL n.d.) and achieves higher quality in translation than DeepL classic, particularly for longer texts. Although the text analyzed in this ar-

ticle is not representative of this tendency, when using DeepL next-gen for the English–German language combination, often a slightly higher error rate seems to occur in the machine output compared to the classic language model, while at the same time the automated translations generated with DeepL next-gen also seem to contain more "creative shifts" (Bayer-Hohenwarter 2011: 669) or translation solutions that significantly transform the source text. This has been shown by Walter (2025a) in her analysis of the DeepL next-gen output for a section from a contemporary Irish short story by Mike McCormack entitled "Beyond." The translation of the same section generated with DeepL's classic language model, by contrast, was more "reproductive" (see Bayer-Hohenwarter 2011: 674), meaning that structures and expressions from the source text were transferred into the target language, sometimes at the cost of idiomaticity. Generally speaking, the classic language model also seems to be more reliable than DeepL next-gen, meaning that translation errors occur less frequently.

Given the limited information available on the precise technology underlying the DeepL Translator, any statements about reasons for the next-generation language model's less reliable and more frequently paraphrased translation output have to remain speculative. What can be established, however, is that DeepL next-gen manifests similar tendencies in translation as general-purpose LLMs such as ChatGPT when they are used without task-specific prompting (see Walter 2025a). Hendy et al. (2023) argue that, in order to achieve strong, in-context multilingual capabilities, general-purpose LLMs such as Chat GPT require multiple times more training data than conventional neural machine translation (NMT) systems such as DeepL. One potential reason for the higher error rate in LLM translation output is thus a larger randomization of the results due to much more substantial training datasets.

To date, the term "creative shifts" (Bayer-Hohenwarter 2011: 669) as discussed in this article has usually been associated with deliberate choices made by humans, either in translation or post-editing. However, since machine output with new translation technology such as DeepL next-gen manifests patterns that sometimes *look like* the choices conventionally made by humans, a pragmatic decision has been made to also use this term for phenomena observed in the machine output. At the same time, this article does not seek to anthropomorphize when discussing technologies based on artificial intelligence (AI). Nor does the appropriation of a term conventionally used for human language indicate that humans are now less important as agents than before in literary translation workflows. In fact, the comparison of the post-edited versions to the human translations in this article shows that—particularly for texts that contain a higher percentage of UCPs—literary post-editing may not be a good option. This is because a translator's choices are constrained by the machine output, which may negatively interfere with their creativity and the joyfulness of the activity of translating. How these constraints occur will be described in more detail in the next section of this article. However, before some key terms associated with literary post-editing are outlined, it should be noted that Orel Kos (2024) and Agnetta (2025) have gained similar insights regarding counterproductive effects of AI-enhanced workflows on translation quality and translator motivation in the context of audiovisual translation.

2.2 Key Terms

A selection of phenomena that frequently occur in post-edited literary translations will be briefly outlined here, starting with "priming" (see Hamm 2024: 16; Kolb 2022: 20 and 2023: 55).

Literary post-editors may be limited in the variety of translation options they can conceive due to their excessive reliance on machine output. On the other hand, some post-editors also find it difficult to accept any machine output, often due to negative or ambivalent attitudes towards the technologization of translation. This phenomenon, which is diametrically opposed to priming, has been described as "Distinktionswille" (a "desire to be different"; Förster et al. 2023: n.p.).

In fact, the cognitive influence of machine output on post-editors known as priming may not be the only reason for the limited creativity and appeal of many post-edited translations. Particularly in specialized translation, language service providers that have been assigned post-editing tasks are often encouraged to use as much of raw machine output as possible due to time constraints. This is another reason why post-edited target texts tend to retain many features of the original automated translations they are based on, especially when they are the product of minimally invasive "light" post-editing (see Nitzke/Hansen-Schirra 2021: 30–31). Apart from priming, post-editors have also witnessed a "fatigue" effect (Hamm 2024: 16) as a result of the exhaustion that may stem from having to base their final version on two drafts instead of just one, a source text and an automated translation. Finally, an "obstacle" effect (ibid.) has also been described. It occurs when post-editors find detecting errors in the machine output particularly difficult due to the fact that the algorithmic, probabilistic language of AI-based translation tools only *resembles* human language but operates on very different principles.

For reasons outlined above, post-edited translations are often said to contain examples of "post-editese" (Toral 2019). This somewhat vague but conveniently encompassing term refers to the understanding that post-edited translations tend to be simpler and show more interference from the source lan-

guage than human translations. Compared to human translations, post-edited target texts therefore tend to include fewer creative shifts, particularly if they were produced with the help of systems such as the DeepL classic language model or Google Translate. The more imprecise but also more flexible LLM-based translation output of recent years tends to be less reproductive of the source text structures and may thus be more promising with regard to achieving idiomaticity in the target language.

2.3 Source Text

The source text used for this experiment is a literary sketch or prose poem called "Green," which was originally published by Virginia Woolf in 1921 as part of a collection of short prose called *Monday or Tuesday*. Like the entire collection of which this text forms part, "Green" has been associated with literary impressionism. One of the author's aims with this collection was to reconcile writing with other art forms, in the case of "Green" specifically painting. The text is part of a literary diptych called "Blue & Green," where "Green" represents day, while "Blue" depicts a symbolic image of evening (see Louvel 2023). Both texts are largely descriptive, more like images than narratives with conventional storylines.

2.4 DeepL Output

In Table 1, Woolf's original literary sketch (1921/2021: 174) is juxtaposed with the DeepL output from January 10, 2025, which the post-editing group had to base their final translations on. There were two significant errors in the translation generated with DeepL next-gen that are represented in bold letters,

which made this draft particularly interesting for this classroom experiment, and which will be analyzed in the next paragraph.

"Green" by Virginia Woolf (1921/2021)	**DeepL next-gen, January 10, 2025**
GREEN The pointed fingers of glass hang downwards. The light slides down the glass, and drops a pool of green. All day long the ten fingers of the lustre drop green upon the marble. The feathers of parakeets—their harsh cries—sharp blades of palm trees—green, too; green needles glittering in the sun. But the hard glass drips on to the marble; the pools hover above the dessert sand; the camels lurch through them; the pools settle on the marble; rushes edge them; weeds clog them; here and there a white blossom; the frog flops over; at night the stars are set there unbroken. Evening comes, and the shadow sweeps the green over the mantelpiece; the ruffled surface of ocean. No ships come; the aimless waves sway beneath the empty sky. It's night; the needles drip blots of blue. The green's out.	GRÜN Die spitzen Finger aus Glas hängen nach unten. Das Licht gleitet das Glas hinunter und lässt eine grüne Lache fallen. Den ganzen Tag lang fallen die zehn Finger des **Glanzes** grün auf die **Murmel**. Die Federn der Sittiche – ihre schrillen Schreie – scharfe Klingen von Palmen – auch grün; grüne Nadeln glitzern in der Sonne. Aber das harte Glas tropft auf den Marmor; die Pfützen schweben über dem Wüstensand; die Kamele taumeln durch sie hindurch; die Pfützen setzen sich auf dem Marmor ab; Binsen säumen sie; Unkraut verstopft sie; hier und da eine weiße Blüte; der Frosch plumpst hinüber; nachts sind die Sterne ungebrochen dort gesetzt. Der Abend kommt, und der Schatten streift das Grün über den Kaminsims; die gekräuselte Oberfläche des Ozeans. Keine Schiffe kommen; die ziellosen Wellen schwanken unter dem leeren Himmel. Es ist Nacht; die Nadeln tropfen blaue Flecken. Das Grün ist aus.

Table 1: Original text versus translation with DeepL next-gen

One challenge in the source text that both the post-editors and the human translators struggled with is the term "lustre," which is frequently used in the sense of "brilliance" or "Glanz" but which in British English is also a synonym of "chandelier" ("lustre"). Due to its inbuilt probabilistic approach to translation, DeepL next-gen generated the most likely equivalent of "lustre" in German, which is "Glanz," rather than correctly

rendering the term as "Kronleuchter" or "Luster" in the given context. In the same sentence, "marble" is also mistranslated as "Murmel," one of many small colored glass balls that children use to play Marbles, the rarely used singular form of which is also "marble" in English ("marble"). This mistranslation is somewhat surprising, as elsewhere in the text "marble" is correctly rendered as "Marmor," a type of limestone, but the lack of training data for "lustre" in the British English sense of "chandelier" probably made a correct translation of other elements in this sentence more difficult.

While AI-enhanced translation tools occasionally generate erroneous translation output, sometimes errors in a source text are also corrected. There is a spelling error resulting in a different word from the one intended in Woolf's original text, who used "dessert" (something you eat at the end of a meal and that is often sweet) instead of "desert" (an extremely dry area of land where it is usually hot). Based on the context, DeepL next-gen correctly translated "dessert" as part of the compound "Wüstensand" ("desert sand"), rather than generating the lexical equivalent, "Nachspeise" ("dessert"). While a potential correction of errors in a source text is certainly desirable in automated translation, this example highlights that human control of machine output remains key to ensuring high-quality, correct and adequate translations.

Even though for another source text translated with DeepL next-gen the output tended to be less reproductive than with DeepL classic (see Walter 2025a), the automated translation of Woolf's text only manifests very few creative shifts. This is probably due to the fact that the collocations used in the source text are mostly very unusual, so that the training datasets did not yield any probabilistic combinations in the target language.

2.5 DeepL Output versus Human Reference Translation

In Table 2, the DeepL next-gen machine output from January 10, 2025, is juxtaposed with a recently published translation of "Green" by Christel Kröning (see Woolf 2021: 175). Subsequently, the key differences between both German versions will be briefly outlined.

DeepL next-gen, January 10, 2025	**"Green" by Virginia Woolf, translated by Christel Kröning (2021)**
GRÜN Die spitzen Finger aus Glas hängen nach unten. Das Licht gleitet das Glas hinunter und lässt eine grüne Lache fallen. Den ganzen Tag lang fallen die zehn Finger des Glanzes grün auf die Murmel. Die Federn der Sittiche – ihre schrillen Schreie – scharfe Klingen von Palmen – auch grün; grüne Nadeln glitzern in der Sonne. Aber das harte Glas tropft auf den Marmor; die Pfützen schweben über dem Wüstensand; die Kamele taumeln durch sie hindurch; die Pfützen setzen sich auf dem Marmor ab; Binsen säumen sie; Unkraut verstopft sie; hier und da eine weiße Blüte; der Frosch plumpst hinüber; nachts sind die Sterne ungebrochen dort gesetzt. Der Abend kommt, und der Schatten streift das Grün über den Kaminsims; die gekräuselte Oberfläche des Ozeans. Keine Schiffe kommen; die ziellosen Wellen schwanken unter dem leeren Himmel. Es ist Nacht; die Nadeln tropfen blaue Flecken. Das Grün ist aus.	GRÜN Die spitzen Glasfinger hängen herab. Das Licht läuft am Glas hinunter und bildet Tropfen um Tropfen eine Lache Grün. Den ganzen Tag lang tropft von den zehn Kronleuchterfingern Grün auf den Marmor. Federn von Papageien – ihre schrillen Rufe – messerscharfe Palmblätter – grün auch sie. Grüne Nadeln, glitzernd im Sonnenlicht. Doch das Kristallglas tropft weiter auf den Marmor. Die Lachen schweben über dem Wüstensand. Die Kamele taumeln hindurch. Die Lachen werden heimisch auf dem Marmor. Schilf umsäumt sie. Algen durchdringen sie. Hier und dort eine weiße Blüte. Der Frosch springt plumpsend hinein. Nachts liegen ungebrochen die Sterne darin. Der Abend naht und der Schatten wischt das Grün über den Kaminsims. Aufgewühlter Ozean. Kein Schiff in Sicht. Die ziellosen Wellen wiegen sich unter leerem Himmel. Es ist Nacht. Von den Nadeln tropfen Kleckse Blau. Das Grün ist erloschen.

Table 2: DeepL next-generation language model versus human reference translation

The most important differences between Kröning's version and the DeepL output result from creative shifts that occur due to the translator's deliberate interventions in the source text structure. Unlike Woolf, Kröning uses many compounds, a frequent word formation strategy in German that normally features much less in English, thus naturalizing the target text. For instance, Kröning uses "Kronleuchterfinger(n)," rather than "Finger des Glanzes," and "messerscharfe Palmblätter" instead of "scharfe Klingen von Palmen," taking advantage of the flexible word combinations the German language facilitates. In general terms, compared to the DeepL output, Kröning's translation could be described as hermeneutic, as a product of intellectually engaging with the meaning of the source text, rather than reorganizing algorithmic patterns found at the textual surface level. Among other things, this difference manifests in the translator's choice of a number of target terms that are far removed from the textual surface constructed in the source language, whose meaning they interpret. For example, Kröning uses "werden heimisch" or "make themselves at home" for "settle on," "aufgewühlt" or "upset" for "ruffled," and "erloschen" or "extinguished" for "out." By contrast, the DeepL output for these passages remains much closer to the source text. For instance, target language equivalents used in the machine output include "sich absetzen" for "settle on," "gekräuselt" for "ruffled" and "aus" for "out."

In recent years, the Bilingual Evaluation Understudy (BLEU) score and other metrics have been developed to measure the quality of automated translations compared to a human reference translation. Using the open-source online evaluation tool MATEO (see Vanroy et al. 2023: 499–500), the BLEU metric is determined by rating the machine output on a scale from 0 to 100, 0 meaning that there is no similarity, while 100 indicates that an automated translation is identical to a hu-

man reference translation. Each translation is divided into n-grams, which may be words, syllables, a group of adjacent letters or punctuation marks, and quality is measured "in terms of surface n-gram matching" between the machine output and the human reference translation (Saadany/Orăsan 2021: 50). Although the BLEU score is mainly used to measure the progress made in the training of automated translation tools, the result will be provided for the machine output analyzed in this article, which in comparison to Kröning's translation scored a value of 16.6. This fairly low score indicates that the differences between the machine output and the human reference translation are more substantial than they might seem at first glance.

2.6 Post-edited Versions versus Human Translations

In this section, three segments from the source text will be compared to the six different versions in the target language produced by the participants in the literary post-editing/translation experiment discussed in this article. For each segment, in a first step an overview of the post-edited versions and human translations will be shown and analyzed in general terms. For segments one and two, the post-edited versions and human translations respectively will then be explored in more detail. To substantiate the analysis, some keylogging data will also be presented to examine not only the final translations but also how they were produced.

2.6.1 Segment One

The first segment analyzed in this article is the opening sentence in the source text. The following overview shows the three post-edited and human-translated versions of this segment.

The pointed fingers of glass hang downwards.
(Woolf 1921/2021: 174)

Die spitzen Finger aus Glas hängen nach unten.
(DeepL next-gen)

Die spitzen Finger aus Glas hängen nach unten.
(DeepL next-gen post-edited, PE-A)

Die spitzen Finger aus Glas hängen nach unten.
(DeepL next-gen post-edited, PE-B)

Die spitzen Finger aus Glas hängen herab.
(DeepL next-gen post-edited, PE-C)

Die gläsernen Finger langen nach unten.
(Human translation, HT-D)

Die spitzen Glasfinger hängen von oben herab.
(Human translation, HT-E)

Die Fingerspitzen aus Glas zeigen nach unten.
(Human translation, HT-F)

Overall, the three human-translated target texts show more variety and more creative shifts than the post-edited versions. Interestingly, in two out of three human translations the anthropomorphism that is already present in the source segment is enhanced, as Woolf's "pointed fingers of glass" are given agency. In version D, the glass fingers "reach down," while in version F they "point downwards." This tendency towards giving the animated chandelier agency is not present in the post-edited versions, which are almost identical to each other and to the DeepL output, with the exception of version C, in which the end of the sentence was changed slightly.

Table 3 again shows the post-edited versions of the opening sentence and, in the second column, gives an overview of the keylogging data from the general analysis generated with Inputlog, which recorded each keystroke, mouse click and deletion made by the experiment participants. Given the homogeneity of the results for this segment, it comes as no

surprise that the keylogging data for this sentence also showed little variety.

DeepL next-gen PE	Keylogging
A: Die spitzen Finger aus Glas hängen nach unten.	*No words are looked up. The sentence is not edited.*
B: Die spitzen Finger aus Glas hängen nach unten.	*No words are looked up. The sentence is not edited.*
C: Die spitzen Finger aus Glas hängen herab.	*After 7:38 minutes, "nach unten" is replaced with "herab."*

Table 3: Segment 1: observations from keylogging (post-edited versions)

This overview testifies to the fact that two out of three participants in the post-editing group never questioned the machine output for this sentence. Only in version C, the adverb at the end of the sentence ("nach unten") was replaced with a synonymous expression ("herab"). This marks the first change made to the machine output in version C and is the only one recorded for this sentence.

By contrast, the keylogging data for the more varied human translations of the first sentence in Woolf's story also show the participants' more diverse approaches to its translation, as Table 4 indicates.

HT	Keylogging
D: Die gläsernen Finger langen nach unten.	*After 12:12 minutes, the beginning of the sentence is written down. The final version of the sentence is produced straight away.*
E: Die spitzen Glasfinger hängen von oben herab.	*All other sentences are typed first first. After 27:24 minutes, the beginning of the sentence is written down, starting with "Die langen Finger … ." Then, the noun is replaced with the compound "Glasfinger." After 32:31 minutes, the adjective "langen" is replaced with "spitzen."*
F: Die Fingerspitzen aus Glas zeigen nach unten.	*After 9:46 minutes, the beginning of the sentence is written down, starting with "Die Fingerspitzen …." Then, the* Duden *entry on "Finger" is checked. After 11:03 minutes, this participant continues "… aus Glas zeigen nach unten" and also notes down*

	a second version of this sentence: "Die gestreckten Finger aus Glas hängen nach unten." After 36:57 minutes, the second version is deleted.

Table 4: Segment 1: observations from keylogging (human translations)

The keylogging data for the human translations of this segment indicate that the underlying translation processes are as different as their outcomes. An examination of the full keylogging data shows that participant D in this group has a tendency to first think through their translation and start writing later than the other participants. At the same time, once they have written down a target language version, they are unlikely to change it. Participant F recurrently notes down different translations of a segment, out of which they subsequently choose their preferred version. Participant E wrote down all other sentences in translation before they rendered the opening sentence into German. Perhaps they wanted to make sense of the overall meaning of the text before providing a German version of the alienating, anthropomorphized description of a chandelier in this segment. The translation process underlying version F of this sentence is particularly multi-layered, showing several replacements. Such complexity frequently characterizes human translation. Post-editing, by contrast, tends to be less analytical and more superficial. While no one from the post-editing group looked up any terms before translating this segment, the keylogging files show that participant F in the human translation group checked a monolingual German dictionary entry on "Finger," which also points towards human translation being more thorough than post-editing.

2.6.2 Segment Two

The second sentence in Woolf's prose poem deserves special attention. As before, the original English sentence is followed

by the raw machine output, as well as the post-edited and human-translated German versions produced during the classroom experiment.

> All day long the ten fingers of the lustre drop green upon the marble. (Woolf 1921/2021: 174)
>
> Den ganzen Tag lang fallen die zehn Finger des Glanzes grün auf die Murmel. (DeepL next-gen)
>
> Den ganzen Tag lang fallen die zehn Finger des Glanzes grün auf den Marmor. (DeepL next-gen, PE-A)
>
> Den ganzen Tag lang tropfen die zehn Finger des Glanzes grün auf den Marmor. (DeepL next-gen, PE-B)
>
> Den Tag hindurch fällt Grün von den zehn Fingern des Kronleuchters auf den Marmor. (DeepL next-gen, PE-C)
>
> Den ganzen Tag lang tropft es an den zehn Fingern des Lusters grün auf den Marmor. (Human translation, HT-D)
>
> Den ganzen Tag werfen die zehn glänzenden Finger ein Grün über den Mamor. (Human translation, HT-E)
>
> Den ganzen Tag lang lassen die zehn Finger des Glanzes Grün auf den Marmor tropfen. (Human translation, HT-F)

This sentence was expected to be difficult for both the post-editors and the human translators. As has been mentioned before, "lustre" is used in the comparatively little known sense of "chandelier." Accordingly, it comes as no major surprise that in the DeepL output the more familiar target term "Glanz" ("brilliance") is used. Additionally, "marble" is mistranslated as "Murmel," a colored type of glass ball frequently used in Marbles, a game mainly played by children.

Before the overall tendencies in the post-edited and human-translated versions respectively are outlined, some interesting observations about individual versions will be briefly summarized. In version PE-C, the adjective "green" replaces "fingers" as a verb complement, and both errors are corrected.

In version HT-E, there is a typing mistake in "Marmor," which is wrongly spelt "Mamor," probably due to the fact that spell-checking does not work when the keylogging software Inputlog is used. In the same version, "lustre" is mistranslated as "glänzend," which is the adjective form of the noun "Glanz." In version HT-F, the main verb is changed to "lassen" or "let," which is a subtle but significant intervention that attributes intentionality to the lustre fingers.

As before, there is less variety in the post-edited target texts compared to the human translations made without machine assistance, but the difference in variety is less pronounced than for segment one. In all final versions, "marble" was correctly rendered as "Marmor." Thus, none of the post-editors were misled by the inconsistency in the machine output for this term, which features three times in total in the source text. However, the mistranslation of "lustre" as "Glanz" in the machine output was only corrected by participant C in the post-editing group. Furthermore, this error also features in two out of three human translations. A look at the keylogging data for this segment in Table 5 will shed light on the reasons for this error.

DeepL next-gen PE	**Keylogging**
A: Den ganzen Tag lang fallen die zehn Finger des Glanzes grün auf den Marmor.	*After 09:15 minutes, the term "lustre" is looked up in* Cambridge *(English), which contains no reference to the use of this term in the sense of "chandelier." After 10:36 minutes, "Murmel" ("glass ball") is replaced with "Marmor" ("limestone").*
B: Den ganzen Tag lang tropfen die zehn Finger des Glanzes grün auf den Marmor.	*After 10:06 minutes, "lustre" is looked up in the bilingual* Cambridge *English–German dictionary. After 11:56 minutes, this Participant notes down "den Marmor" alongside "die Murmel." After 17:42 minutes, German translations of "drip" are looked up in* Collins. *After 30:17 minutes, "die Murmel" is deleted. After 30:55 minutes, "tropfen" replaces "fallen."*

C: Den Tag hindurch fällt Grün von den zehn Fingern des Kronleuchters auf den Marmor.	*After 09:29 minutes, this participant looks up "lustre" in* Cambridge *(English), then in* Collins *(English). They scroll down to the British English definitions, which include "chandelier." They google images for "lustre" and "lustre finger." After 11:00 minutes, they look up "Kronleuchter" in* Cambridge *(English–German). After 13:08 minutes, images for "lustre" are googled. After 13:53 minutes, this participant looks up "lustre" in* Collins (*English*). *After 16:17 minutes, they return to the definition of "lustre" in* Collins. *At 16:21 minutes. they look up "finger" in* Collins (*English*). *After 19:23 minutes, this participant replaces the original sentence with the new version.*

Table 5: Segment 2: observations from keylogging (PE versions)

As can be expected due to a number of translation challenges that occurred in this segment, combined with the wrong DeepL translation output, all participants in the post-editing group looked up terms related to this sentence. Furthermore, neither participant left the machine output unchanged. However, while participant A only corrected the obvious error related to the mistranslation of "marble" as "Murmel," participant B additionally changed a verb. Participant C, whose translations can be regarded as the most successful overall for this group, dedicated ten minutes to searching terms in bilingual and monolingual dictionaries and looking up images related to these terms. This shows that a thorough post-editing process that leads to a successful translation may take nearly as long as translating without machine assistance. This is particularly the case for texts such as Woolf's prose poem, the translation of which is challenging, partly because of the occurrence of UCPs, exemplified by the anthropomorphized image of the lustre.

For participants E and F, the keylogging data for the human translations manifest complex, multi-layered research and editing processes preceding the final versions for the second segment (see Table 6).

HT	Keylogging
D: Den ganzen Tag lang tropft es an den zehn Fingern des Lusters grün auf den Marmor.	*As before, this participant immediately writes down the final version of this sentence after an extended period without any keylogging activity, starting after 17:49 minutes.*
E: Den ganzen Tag werfen die zehn glänzenden Finger ein Grün über den Mamor.	*After 7:53 minutes, this participant starts writing down their version of the sentence, beginning with "Den ganzen Tag …." Then they look up "lustre" in* Cambridge *(English), then* Collins *(English). While in the* Collins *entry for British English "chandelier" is listed, the American English dictionary entry does not include a reference to "chandelier," and the participant does not scroll down to check the British English entry, as the screen recording for this segment shows. After 9:01 minutes, this participant continues writing: "… werfen die zehn Finger des Scheins Grün über den Marmor." They replace "des Scheins" with "des Glanzes." They change the word category to "scheinenden," replacing the noun with an adjective. After 43:31 minutes, they replace "scheinenden" with "glänzenden."*
F: Den ganzen Tag lang lassen die zehn Finger des Glanzes Grün auf den Marmor tropfen.	*After 14:20 minutes, this participant starts typing "Den ganzen Tag lang Grün tropft von den Fingern des …." They look up "lustre" in* Collins *(English–German). They write down "Schimmers" and "Glanzes" as translation options. They replace the original verb with a construction with "lassen" ("let") and write "lassen die zehn Finger des Schimmers/Glanzes Grün auf den Marmor." They look up "tropfen" and "perlen" in the assigned monolingual and bilingual dictionaries. After 39:24 minutes, "Schimmer" is deleted and "tropfen/perlen" are added as potential replacements. "Perlen" is subsequently deleted again.*

Table 6: Segment 2: observations from keylogging (HT versions)

For two out of three participants in this group, the translation of this segment involved examining dictionary entries and several rounds of revision. Only participant D did not revise this sentence and followed their habit of thinking the translation through in advance and then immediately finalizing it.

2.6.3 Segment Three

A simple three-word sentence, "The green's out," ends Woolf's literary sketch. While the keylogging data for this segment do not yield any interesting insights, a comparison of the final translation outputs for the two groups involved in this experiment shows fundamentally different tendencies. The human translators chose three different target language equivalents for the complement in the original sentence, whereas in the post-editing group only one out of three participants changed the complement used in the machine output, which was based on a very literal translation of the source segment, as the following overview shows.

> The green's out. (Woolf 1921/2021: 174)
>
> Das Grün ist aus. (DeepL next-gen)
>
> Grün ist aus. (DeepL next-gen post-edited, PE-A)
>
> Das Grün ist aus. (DeepL next-gen post-edited, PE-B)
>
> Grün ist vorbei. (DeepL next-gen post-edited, PE-C)
>
> Das Grün ist leer. (Human translation, HT-D)
>
> Das Grün verblasst. (Human translation, HT-E)
>
> Das Grün ist weg. (Human translation, HT-F)

The different target language versions chosen for this simple three-word sentence highlight that literary post-editing, even though it may well increase productivity, especially for simpler source texts than Woolf's prose poem, will inevitably curtail linguistic diversity in translation. If post-editors find the machine output acceptable, the motivation to choose alternative translations seems to be limited at best. The three human translations, in contrast with the almost identical post-edited versions, are not only superficially different from each other, but they are in fact all based on a fundamentally different understanding of what "out" means in the source text. In this re-

spect, the experiment results echo Mjölsnes (2022: 41), who has criticized automated translation for deceiving its readers, as a painted window on the façade of a house might do. If post-editing becomes a standard workflow in literary translation, there is a real risk that too little attention might be given to understanding a source text, as too much energy might be taken up by producing a synthetic draft, based on a combination of machine output and more or less superficial human interventions.

2.7 Comparison

A comparison of the six target language versions of each segment from Woolf's literary sketch presented in this article shows that the human translations differ substantially from each other. In contrast, the post-edited versions tend to strongly resemble both the machine output and each other. In addition, the human translations are fundamentally hermeneutic, more so than the post-edited automated translations, which frequently reproduce source language structures, thus overemphasizing the textual surface of the original. Finally, an analysis of how target language version C of the second segment was produced suggests that a thorough post-editing process that leads to high-quality translation output may be as time-consuming as a thorough human translation process. If translation quality is prioritized, the time gains that come with post-editing are therefore insignificant or inexistent for a text like Woolf's "Green."

3 Synthesis and Outlook

Consistent with previous research, this article shows that while literary post-editing may increase productivity for some but not

all source texts, it certainly poses several challenges. First, priming effects may limit translatorial agency and creativity when post-editors are overly influenced by machine output and struggle to make a translation their own (see Kolb 2022: 21). Additionally, post-editing processes may be complicated by a fatigue effect, which is the result of an exhaustion that may stem from working with two drafts instead of one (see Hamm 2024: 16). Finally, post-editors' work is sometimes also constrained by an obstacle effect, meaning that detecting errors in machine output, which resembles human language but operates on fundamentally different principles, can be difficult (see ibid.).

Although there is no explicit record in the keylogging data testifying to a fatigue effect, the limited effort made by two out of three participants in the post-editing group to personalize the machine output may well be due to priming combined with fatigue. Furthermore, two out of three post-edited versions of segment two—"Den ganzen Tag lang fallen die zehn Finger des Glanzes grün auf den Marmor." (DeepL next-gen, January 10, 2025)—retain an awkward construction from the DeepL output, which could be a manifestation of Hamm's obstacle effect. Specifically, in the automated translation the verb "drop" is mistranslated as "fallen" ("fall") instead of "tropfen" ("drop") and then wrongly associated with the noun "Finger." This creates an image of green fingers falling down, which is difficult to reconcile with the image in Woolf's original text, where green color drips down from the anthropomorphized chandelier. For this section, the DeepL next-gen translation output, which overemphasizes the textual surface in the original literary sketch and is oblivious to the underlying meaning, was not changed by a majority of the post-editors. A comparison shows that a similarly awkward rendering does not appear in the human translations of the same segment.

Above all, these potential problems suggest that human language mediators must develop new skills (see Agnetta/Walter 2025, as well as other articles in the thematic *trans-kom* issue outlined in this introductory text). For this reason, the training of future translators must consider the cognitive effects of machine output from the beginning and determine how and where to address them. While standardized machine output may not be perceived as a disadvantage when translating an instruction manual, for texts with a high level of UCPs such as Woolf's prose poem, this homogenization of translation is likely to entail at least some degree of impoverishment, as the experiment analyzed in this article indicates. In different contexts, a potential reduction in linguistic diversity that may result from automated literary translation is also addressed by Kolb (2025) and Walter (2025b) respectively.

If automation is to be used for literary translation without harming translation quality, new workflows for post-editing creative texts must be developed. For instance, these workflows may include new CAT tools with larger context windows that provide access to both NMT and LLM output as required. In order to maximize translatorial creativity and limit priming (see Hamm 2024: 16; Kolb 2022: 20 and 2023: 55), it would probably be best to make machine output available on demand, rather than automatically (see Brenner/Koponen 2025: 54). If machine output offers several translation options, its potentially unwanted, excessive influence on post-editors might be reduced. However, this advantage would have to be weighed against the likelihood of increasing the fatigue effect that may stem from working with several translation drafts (see Hamm 2024: 16). Ultimately, literary post-editing demands a recalibration of both tools and training to better support creative agency. If approached proactively, as well as with critical awareness, the evolving interplay between human creativity

and machine assistance may facilitate flexible and empowering new practices for translators working at the intersection of art and automation.

4 References

AGNETTA, Marco / WALTER, Katharina (2025): "Künstliche Intelligenz in der Ausbildung zum Sprachmittler: Einführung in das Themenheft." In: AGNETTA, Marco / WALTER, Katharina [eds.]: *trans-kom* 18/1: *Künstliche Intelligenz in der Sprachmittlung und im Fremdsprachenerwerb*, pp. 1–25. URL: <https://www.trans-kom.eu/bd18nr01/trans-kom_18_01_01_Agnetta_Walter_Einleitung.20250707.pdf> (08.08.2025).

AGNETTA, Marco (2025, forthcoming): "Investigating Students' Attitude toward the Use of AI in Interlingual Subtitling Process." In: YAHIAOUI, Rashid [ed.]: *Transl-AI-tion 2.0. Embracing the AI Revolution*. Berlin / Bern: Peter Lang Publishing.

ANGELONE, Erik (2025): "Using Think-aloud, Screen Recording, and Questionnaire Data to Gauge Translation Student Application and Perception of ChatGPT." In: *Lebende Sprachen* 70/1, pp. 70–91. DOI: <10.1515/les-2025-0020>.

BAYER-HOHENWARTER, Gerrit (2011): "'Creative Shifts' as a Means of Measuring and Promoting Translational Creativity." In: *Meta* 56/3, pp. 663–692. DOI: <10.7202/1008339ar>.

BRENNER, Judith / KOPONEN, Maarit (2025): "Video Game Translation and Creativity in the Age of Artificial Intelligence." In: WALTER, Katharina / AGNETTA, Marco [eds.]: *Applying Artificial Intelligence in Translation: Possibilities, Processes and Phenomena*. London / New York: Routledge, pp. 46–60.

CAMBRIDGE – URL: <https://dictionary.cambridge.org/de/worterbuch/deutsch-englisch/> (27.11.2024).

COLLINS – URL: <https://www.collinsdictionary.com> (27.11.2024).

— s. v. "lustre." URL: <https://www.collinsdictionary.com/dictionary/english/lustre> (29.07.2025).

— s. v. "marble." URL: <https://www.collinsdictionary.com/dictionary/english/marble> (29.07.2025).

DeepL (2024): "About the Next-Generation Language Model." URL: <https://support.deepl.com/hc/en-us/articles/14241705319580-About-the-next-generation-language-model> (27.11.2024).

DeepL (n. d.): "About DeepL Language Models." URL: <https://support.deepl.com/hc/en-us/articles/14241705319580-About-DeepL-language-models#h_01JCNJ5K0A7A62EWGDRQSQY071> (28.07.2025).

DUDEN – URL: <https://www.duden.de/woerterbuch> (27.11.2024).

FÖRSTER, Andreas G. / FRANCK, Heide / HANSEN, André (2023): "Übersetzungsmaschinen und Literatur" (online seminar, 30.11.2023–01.12.2023). URL: <kollektive-intelligenz.de> (28.07.2025).

GUERBEROF-ARENAS, Ana / TORAL, Antonio (2022): "Creativity in Translation." In: *Translation Spaces* 11/2, pp. 184–212. DOI: <10.1075/ts.21025.gue>.

GUERBEROF-ARENAS, Ana / TORAL, Antonio (2024): "To Be or Not to Be. A Translation Reception Study of a Literary Text Translated into Dutch and Catalan Using Machine Translation." In: *Target. International Journal of Translation Studies* 36/2, pp. 215–244. DOI: <10.1075/target.22134.gue>.

HAMM, Claudia (2024): "Das Blaue vom grünen Himmel." In: HAMM, Claudia [ed.]: *Automatensprache*. Munich: Hanser, pp. 9–31.

HENDY, Amr / ABDELREHIM, Mohamed / SHARAF, Amr / RAUNAK, Vikas / GABR, Mohamed / MATSUSHITA, Hitokazu / KIM, Young Jin / AFIFY, Mohamed / AWADALLA, Hany Hassan (2023): "How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation." In: *arXiv*. URL: <https://arxiv.org/abs/2302.09210> (28.07.2025).

KOLB, Waltraud (2021): "'Hemingway's Priorities Were just Different.' Self-concepts of Literary Translators." In: KAINDL, Klaus / KOLB, Waltraud / SCHLAGER, Daniela [eds.]: *Literary Translator Studies*. Amsterdam: John Benjamins, pp. 107–121. DOI: <10.1075/btl.156.05kol>.

KOLB, Waltraud (2022): "Welche Rolle können Maschinen in der Literaturübersetzung spielen?" In: *UNIVERSITAS* 1/22, pp. 19–23.

KOLB, Waltraud (2023): "'I am a bit surprised': Literary Translation and Post-editing Processes Compared." In: ROTHWELL, Andrew / WAY, Andy / YOUDALE, Roy [eds.]: *Computer-Assisted Literary Translation: The State of the Art*. London / New York: Routledge, pp. 53–68. URL:

<https://www.taylorfrancis.com/chapters/edit/10.4324/9781003357391-4/bit-surprised-waltraud-kolb> (08.08.2025).

KOLB, Waltraud (2025): "Creativity in Literary Post-editing: Reworking a German DeepL Translation of a Text by Lima Barreto." In: WALTER, Katharina / AGNETTA, Marco [eds.]: *Applying Artificial Intelligence in Translation: Possibilities, Processes and Phenomena*. London / New York: Routledge, pp. 15–28.

LEIJTEN, Mariëlle / VAN WAES, Luuk (2013): "Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes." *Written Communication* 30/3, pp. 358-392. DOI: <10.1177/0741088313491692>.

LOUVEL, Liliane (2023): "'Blue & Green': One of Virginia Woolf's 'Short Things.'" In: *Journal of the Short Story in English* 80/81, n.p. URL: <https://journals.openedition.org/jsse/4107> (28.07.2025).

MJÖLSNES, Ettore (2022): *Der stumme Text: Eine Kritik der maschinellen Übersetzung*. Küsnacht: Digiboo.

NITZKE, Jean / HANSEN-SCHIRRA, Silvia (2021): *A Short Guide to Post-Editing*. Berlin: Language Science Press. DOI: <10.5281/zenodo.5646896>.

OREL KOS, Silvana (2024): "Introduction of Machine Translation into Audiovisual Translation Teaching." In: *ELOPE. English Language Overseas Perspectives and Enquiries* 21/1: English Studies at the Dawn of the 21st Century, pp. 185–208.

SAADANY, Hadeel / ORĂSAN, Constantin (2021): "BLEU, METEOR, BERTScore: Evaluation of Metrics Performance in Assessing Critical Translation Errors in Sentiment-oriented Text." In: *Translation and Interpreting Technology Online*, pp. 48–56. DOI: <10.26615/978-954-452-071-7_006>.

TORAL, Antonio (2019): "Post-Editese: An Exacerbated Translationese." In: FORCADA, Mikel / WAY, Andy / HADDOW, Barry / SENNRICHS, Rico [eds.]: *Proceedings of the Machine Translation Summit XVII: Research Track*. Dublin: European Association for Machine Translation, p. 273–281. URL: <https://aclanthology.org/W19-6627> (28.07.2025).

VANROY, Bram / TEZCAN, Arda / MACKEN, Lieve (2023): "MATEO: Machine Translation Evaluation Online." In: NURMINEN, Mary / BRENNER, Judith / KOPONEN, Maarit / LATOMAA, Sirkku / MIKHAILOV, Mikhail / SCHIERL, Frederike / RANASINGHE, Tharindu / VANMASSENHOVE, Eva / ALVAREZ VIDAL, Sergi / ARANBERRI, Nora / NUNZIATINI, Mara / PARRA ESCARTÍN, Carla / FORCADA, Mikel /

POPOVIC, Maja / SCARTON, Carolina / MONIZ, Helena [eds.]: *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*. Tampere, Finland: European Association for Machine Translation, pp. 499–500. URL: <https://aclanthology.org/2023.eamt-1.52/> (28.07.2025).

WALTER, Katharina (2025a, forthcoming): "Literarisches Übersetzen im Zeitalter der künstlichen Intelligenz: Möglichkeiten und Grenzen teilautomatisierter Arbeitsprozesse." In: CZULO, Oliver / KAPPUS, Martin / HOBERG, Felix [eds.]: *Digitale Translatologie*. Berlin: Language Science Press.

WALTER, Katharina (2025b): "Jeopardizing Linguistic Diversity: How AI-generated Translations Neutralize Vernacular Irish English." In: WALTER, Katharina / AGNETTA, Marco [eds.]: *Applying Artificial Intelligence in Translation: Possibilities, Processes and Phenomena*. London / New York: Routledge, pp. 29–45.

WOOLF, Virginia (1921/2021): "Green." In: *Meistererzählungen. Collected Stories. Zweisprachige Ausgabe*. Munich: Anaconda, p. 174.

WOOLF, Virginia (2021): "Grün." Translated by Christel Kröning. In: *Meistererzählungen. Collected Stories. Zweisprachige Ausgabe*. Munich: Anaconda, p. 175.

About the author: Katharina Walter (https://orcid.org/0009-0007-8042-1145) has been working as a lecturer in English, Translation and Cultural Studies at the University of Innsbruck, Austria, since 2012. Previously, she completed a concurrent Master's degree in English and Spanish at the University of Innsbruck (2002), as well as a Master's degree in Women's Studies (2006) and a PhD in English (2011) at the University of Galway, Ireland. Her current research mainly focuses on artificial intelligence in the field of literary translation, on translating children's literature and on translatorial creativity. Recent publications include *Artificial Intelligence in Translation, Interpreting, and Foreign Language Learning: Educational and Vocational Changes* (2025, *trans-kom* special issue) and *Applying Artificial Intelligence in Translation: Possibilities, Processes and Phenomena* (2025, Routledge), both co-edited with Marco Agnetta.

Contact: Katharina.Walter@uibk.ac.at