Effects of Domain-adapted Machine Translation on the Machine Translation User Experience of Video Game Translators

Judith Brenner[1], **Julia Othlinghaus-Wulhorst**[2]
[1] University of Eastern Finland
[2] Native Prime
jbrenner@uef.fi

Abstract

In this empirical study we examine three different translation modes with varying involvement of machine translation (MT) post-editing (PE) when translating video game texts. The three modes are translation from scratch without MT, full PE of MT output in a static manner, and flexible PE as a combination of translation from scratch with PE of only those automatically translated sentences deemed useful by the translator. In a mixed-methods approach, quantitative data was generated through keylogging, eyetracking, error annotation, and user experience questionnaires as well as qualitative data through interviews. Results for 12 freelance translators show a negative perception of PE and indicate that translators' user experience is positive when translating from scratch, negative with static PE of generic MT output, and neutral with a positive tendency with flexible PE of domain-adapted MT output.

1 Introduction

Video games are multifarious, and translating them blends software localization, technical translation, literary translation, and multimodal translation (Jimenez-Crespo, 2024). The creative field of video game localization, where the act of translating is only one part of localizing a video game for a target market, is under-researched (Zoraqi and Kafi, 2024). At the time the statistical machine translation (SMT) paradigm was prevalent, consensus in the game localization industry was that machine translation (MT)[1] was not useful for game material. With the shift to neural MT (NMT), this notion changed gradually, following behind the general adoption of NMT in the translation industry (just as it followed behind the adoption of computer-aided translation tools; Moorkens et al., 2024). By now, some game publishers and game localization service providers have adopted processes that include post-editing (PE) of NMT output (Akhulkova, 2021; Anselmi and Rubio, 2020; Lionbridge Games, 2024), the main drivers for PE in game translation being the increase of translation speed and the reduction of translation costs (Moorkens et al., 2024).

Although PE is in demand by game localization buyers (Rivas Ginel and Theroine, 2022), freelance translators who provide the translations needed for game localization push back on this practice. In several manifestos published by game translators personally or together with other media translators represented by professional associations, they argue against the use of machine-generated translations, claiming it dismisses human expertise (En Chair et en Os, 2023) and that PE reduces translation quality (Danilov, 2023), inhibits creativity, and is slower than translation from scratch (Deryagin et al., 2021). Based on these points, the comprehensive Machine Translation Manifesto published by the Audiovisual Translators Europe association argues to consider MT as a tool that contributes to the notion of the augmented translator, where the translator is "front and centre and uses technology to enhance their capabilities" (Deryagin et al., 2021, p. 1).

With contradictory claims about the usefulness of PE in game translation between buyers and providers, our research aims to better understand the PE process when translating video games. By focusing on the video game translator's point of view, we contribute to a shift in MT research

[1] In this article we use "MT" as an umbrella term for all forms of automatic translation, regardless of the underlying technological framework. We specify "SMT" and "NMT" when the technological framework is important for specificity.

Proceedings of the Second Workshop on Creative-text Translation and Technology (CTT), pages 27–43, June 24, 2025

proposed by O'Brien (2023) toward human-centered augmented translation, where one aspect is the impact of translation tools on the experience of their users (Briva-Iglesias et al., 2023).

We conducted a study with video game translators who translated parts of a video game under three different translation conditions with varying involvement of PE. The participants were divided into two groups, where one group used generic NMT, and the other group used domain-adapted NMT. Visiting the translators in their home offices, we observed the translation process over a full day using eyetracking and keylogging and captured the user experience with pre- and post-task questionnaires. The study was conducted in collaboration with the game localization service provider Native Prime. Native Prime supported the research by recruiting the study participants and compensating them for their time spent on the study according to a rate negotiated individually between each participant and Native Prime, by providing the game material to translate, resources such as access to the MT system ModernMT, to the terminology database (TB) and translation memory (TM) for the game translation project, and by setting the project up in the translation management system memoQ TMS. Native Prime is interested in the research results as a basis for their strategic decisions about if and how to offer PE as a service while ensuring a good translator experience. Nonetheless, technical and organizational measures were taken to ensure they cannot directly connect participants to the results.

In this article, we first review the current state of research on PE with professional translators. We then explain the experimental setup, followed by presenting the experiment's results concerning perception of PE and user experience. The results are then discussed in relation to other research findings, leading to some conclusions.

2 Related Work

As no publications appear to exist yet on the use of MT or PE when translating video games – a gap also observed by Hansen and Houlmont (2022) and Zhang (2022) –, we informed our study based on PE research in information-centered fields as well as in creative fields such as literary translation and multimodal fields such as subtitling.

Studies on PE in information-centered fields such as banking and finance or software localization report a general increase in productivity compared to translation from scratch (Läubli et al., 2019; Macken et al., 2020), although this increase does not happen for all translators and has a high variability (Macken et al., 2020; Terribile, 2024). On the contrary, in creative fields such as literary translation, productivity can be noticeably decreased (Guerberof-Arenas and Toral, 2022). Quality of translations in banking and finance seems not to be affected whether or not they were produced by post-editing MT output (Läubli et al., 2019). Yet, a quality analysis of post-edited subtitles for TV series showed inferior quality compared to subtitles translated without PE (Hagström and Pedersen, 2022). Additionally, the level of creativity can be decreased in literary translations produced with PE (Guerberof-Arenas and Toral, 2022). With contradictory results for both productivity and quality between PE studies on information-centric texts and on literary texts and subtitles, the question remains how productivity and quality are affected when PE is involved in translating video games, where informative texts and creative texts merge.

PE productivity relates to the translation process, whereas quality relates to the translation product. For a human-centered take on PE, the effects on the translators are of interest as well. Regarding cognitive effort when post-editing, a study on general texts found PE to be cognitively less demanding for the translators than translation from scratch (Daems, 2016). Also, studies with literary translators found PE to cause less cognitive effort than translating from scratch (Ruffo et al., 2024; Toral et al., 2018). However, when asked about their opinion about post-editing, game translators with a diverse set of backgrounds find it inconvenient and not useful (Rivas Ginel, 2023). A survey among German audio-visual translators revealed that those respondents who are specialized in game translation speak out more negatively against the use of MT in general and the most negative aspect in their opinion is the MT output quality (Jaki et al., 2024). In a study with literary translators that compared PE of SMT and NMT with translation from scratch, participants still preferred translation from scratch, even when they learned that their PE productivity was higher than their translation productivity (Moorkens et al., 2018). Such negative points of views can be better understood by studying the user experience of the translator as user of machine translation (Briva-Iglesias, 2024). As part of a human-centered

approach in MT research, Briva-Iglesias (2024) proposes to examine the user experience before and after PE tasks and introduces the concept of machine translation user experience (MTUX), which encompasses usability, satisfaction, and perception both when anticipating to post-edit as well as after the PE task was performed.

Many PE studies involve traditional PE, where entire documents are pretranslated with MT output and consequently everything has to be reviewed and edited (Briva-Iglesias et al., 2023). Some PE studies combine MT output with translation memories (TM) (Macken et al., 2020; Terribile, 2024), where the TM is used to pretranslate sentences with partial TM matches above a certain threshold and the remaining sentences below the threshold are pretranslated using MT. Even fewer studies look at the use of MT for translation support without pretranslation of entire documents (see Vieira 2020 for different ways of employing MT while translating), for example by using adaptive or interactive MT systems where the translator starts typing and the MT system suggests how to complete the sentence (Briva-Iglesias et al., 2023). Some authors propose to study alternatives to PE, such as Hansen and Houlmont (2022) who in the context of game translation propose to use MT not for post-editing but as an additional resource to the TM, in order to not constrain translators' creativity; or Rothwell (2023) who questions the usefulness of pretranslation with MT for literary translation and recommends to explore alternative forms of applying MT in creative text translation.

3 Experimental Setup

The group of interest in our study is professional freelance translators as this is by far the largest group among game translators (Rivas Ginel, 2023; Zoraqi and Kafi, 2024). We decided to employ an experiment in the field because these experiments render results closer to the real working conditions than laboratory-based experiments (Teixeira and O'Brien, 2019). A field experiment also allows us to observe each participant over a longer period, in our case, one full working day. We decided to conduct experiments instead of surveys as we are interested in obtaining measurements from the PE process which are then put into perspective by participant-declared perceptions of it. Lastly, we are interested in comparing a generic NMT system with a domain-adapted NMT system as we want to understand if compiling material for domain adaptation is advisable, considering that game localization poses specific challenges for NMT models (Anselmi and Rubio, 2020; Díaz Montón, 2024), or if PE of NMT can be a viable option for game translation projects where no previous resources for domain adaptation are available.

3.1 Participants

For this study, 15 professional game translators were recruited for four language pairs. They all translate from English into one of the European languages French, Italian, German, or Spanish, the main languages the industry partner provides services for. The participants were recruited from the industry partner's pool of translators and selected based on the following criteria: a) specialization in video game localization, b) language pair, c) place of residence, d) availability, and d) willingness to participate in the study.

The 15 translators gave their consent for participation and data processing in writing. The data generated with their help is kept anonymous from the industry partner. For 2 participants, the experiment sessions could not be completed due to technical problems, so in the end 13 valid data sets could be obtained. Of these 13 participants, 12 were freelance translators with translation being their main occupation and 1 participant with a salaried position in a game localization role and some game translation experience. As the focus of this article is the user experience of professional freelance translators, we remove this participant from the analysis in this article. For one freelancer participant, one of three post-task questionnaires is missing. As this is only a small portion of the entire data set of that participant, and all other data provided by them is of high quality, we contain this participant in the analysis. This results in an even distribution of language pairs, with three freelance participants per target language in the analysis.

The 12 participants under investigation in this article include 6 male and 6 female translators (0 with other gender and 0 prefer not to say), 5 of them between 25 and 34 years, 3 between 35 and 44 years, and 4 between 45 and 54 years of age. The majority is highly educated, with 8 participants holding at least 1 master-level degree in translation or languages, 2 participants with a bachelor's degree in translation, out of which 1 participant has incomplete master's studies, and 2 participants without university education. Their experience working as game translator ranges for 3

participants between 1 and 5 years, for 4 participants between 6 and 10 years, for 2 participants between 11 and 15 years, and for 3 participants between 16 and 20 years. Their work experience translating in other domains ranges from 0 to 30 years. The higher number of years working in other domains suggests a specialization on video game translation occurred over the course of the participants' careers, which is also backed by 10 participants indicating that they mainly or exclusively translate video games, whereas only 2 participants also translate in other domains.

Regarding the participants' experience with PE, 9 participants report prior PE experience. 8 out of these have PE experience when translating games, ranging from 5 months to 2 years. 1 participant has no PE experience in games but 10 months in other domains. 5 participants have PE experience both in games and other domains, and their PE experience in other domains ranges from 1 to 10 years.

3.2 Translation Resources

The aim was to let participants work on a realistic game translation project. We simulated a project where a game that has been translated before gets updated with new content to translate. In such a scenario, typically several resources are available.

The game as object of investigation was chosen by the industry partner. It is a match-3 and hidden-object mobile game revolving around the themes of mysteries and crime-solving. It was picked because it comprises different content types (descriptive text, user interface, dialogs, system messages, etc.) and the translation demands creativity and context sensitivity, while not being too complex. Furthermore, the industry partner had extensive project resources available (comprehensive TMs and a TB in the respective language pairs, reference material for familiarization, etc.).

All participants were provided with a 17-minute gameplay video prepared by the developers of the game for the original localization process. The video served as an introduction to the game so the participants could understand what the game is about and how it is played. In addition, they were presented with 2 pages of localization guidelines from the developers, including information about the game, requested tone and style, formatting specifics, important terminology, and length limitations. The familiarization material was handed to the participants at the beginning of the experiment session to ensure that they all take the time to review it. They could refer to the material anytime during the translation tasks.

Each participant was provided with their own, dedicated memoQ project and anonymized user account with exact same project settings, including a copy of the project's original TM and TB with some adjustments: The TM was cleaned from the segments that were selected for the translation tasks of the study. Moreover, all high fuzzy matches above 90% between the TM and the documents to translate were removed from the TM to make sure the translators had to come up with their own translations by preventing the TM to present a complete translation. As there were no matches between 81% and 89%, this resulted in a TM containing only matches with an 80% match rate or lower. This was relevant for preparing the static PE task. If there had been TM matches between 85% and 89%, we would have used these during the pretranslation step instead of pretranslating entirely with ModernMT. The modified TM included 7,067 segments with a total of around 74,500 words per language. The same TM was also used to create the domain-adapted MT version.

The TB had been created during the original translation of the game by the translators and reviewers in charge of the localization at that time. To make sure that all participants work under the same conditions, the industry partner made sure that the TB contained a similar number of terms in each language. The final TB for the experiment contained 292 terms. The translation documents for the experiment sessions had been prepared in Excel and then imported into each of the memoQ projects with preconfigured settings and filter configurations to make sure all participants worked under the exact same conditions.

ModernMT was included in the memoQ projects in form of an integrated memoQ plugin. According to its website[2], ModernMT is a self-learning MT technology based on neural networks. It was picked for this study because a) it offers document-level adaptation (i.e., it elaborates the translation based on the whole document and not just single sentences), b) it learns from corrections in real time, c) its engine can be further adapted through existing TMs, and d) it can be used with different workflows (static and dynamic).

[2] https://www.modernmt.com

Participants had internet access to perform research while translating.

3.3 Translation Tasks

The translation tasks were designed to be as close to a typical game translation job as possible. Each participant translated three documents that were compiled by manually selecting strings from the game. All three documents were structured the same, with a similar number of strings for each content type, such as dialogues, user interface, descriptive texts, and the game's description for an online shop. The documents were between 820 and 850 words long, comprising of 2,496 words in total. All participants translated the same documents, but in different orders (see Table 1).

The participants were tasked to translate under three different conditions: C1 – translation from scratch, C2 – static post-editing, and C3 – flexible post-editing. In C1, translation from scratch, participants worked in a document where at the start all target segments were empty. They had the TB and the TM available as resources, but not MT. In C2, static PE, the entire document was pretranslated with ModernMT. Again, participants had access to the TB and the TM. In C3, flexible PE, the target segments again were empty at the start and participants had the TB and the TM available. Additionally, they also received suggestions from ModernMT. These were generated at the time of activating a segment for translation and shown alongside the TB and TM matches in the Translation Results pane to the right of the target text column in the memoQ translation editor. The pretranslation for C2 was generated on the morning of each participant session to ensure that the ModernMT output quality for C2 would be similar to the ModernMT output quality in C3 on the same day for each participant.

Moreover, participants were divided into two groups, with each group receiving MT suggestions from a different MT version. ModernMT can be either used in its generic state off the shelf (group A), or tuned with a TM, creating a domain-adapted version (group B). For group B, the same TM of around 74,500 words that was also available during translation was added to ModernMT to create a version adapted specifically to the game under translation. During the study, participants did not know which ModernMT version they used.

Conditions as well as documents were rotated among the 15 participants to account for learning and fatigue effects (see Table 1). Participants only had access to their own copy of the TM to ensure they do not see TM entries originating from other participants before them. From one task to the next, each participant used a fresh copy of the TM, meaning they had no access to what they had translated in the previous translation task. This was to make sure that the productivity data for all three tasks remained comparable.

Parti.	Lang.	Task1	Task2	Task3	Group
P01	DE	C1T2	C2T3	C3T1	B
P02	DE	C1T1	C3T3	C2T2	B
P03	DE	C2T1	C1T3	C3T2	A
P04	ES	C1T2	C2T1	C3T3	A
P05	ES	C3T1	C2T2	C1T3	B
P06	ES	C2T3	C1T1	C3T2	A
P07	FR	C1T3	C3T2	C2T1	B
P08	FR	C3T3	C1T1	C2T2	B
P09	FR	C2T3	C3T1	C1T2	A
P10	IT	C3T3	C1T1	C2T2	A
P11	IT	C3T2	C2T1	C1T3	A
P12	IT	C3T1	C1T2	C2T3	B

Table 1: Rotated assignment of conditions (C1, C2, C3), documents (T1, T2, T3), and MT version (group A, B).

Translation took place in memoQ 11.2, a computer-aided translation tool commonly used for game translations (Rivas-Ginel, 2023) and the industry partner's tool of choice.

3.4 Data Generation

Data was generated at the home offices of participating freelance game translators in France, Italy, Germany, and Spain in December 2024 and January 2025. Most of the translators worked at their usual desks, sitting in their usual chairs and were operating their own mice and keyboards. To fulfill the hardware requirements for the eyetracker, a Tobii Pro X3-120, we brought a 22" screen to which the eyetracker was attached, along with the eyetracker's external processing unit (EPU) and a laptop on which the translation tasks were performed. With his setup participants worked in their usual environment instead of in a lab and their data privacy was protected as no software needed to be installed on participants' PCs.

Each participant session lasted for a full working day. A session began with setting up the

workstation and ensuring the eyetracker was able to track the participant's eye movements. While the researcher prepared the memoQ project and the eyetracking project for the first task, the participant used the familiarization material to get acquainted with the game project. They could take as much time for this as needed. Then they were introduced to the first task to set expectations, followed by answering the pre-task questionnaire. After the questionnaire, they started with their first translation task of the day. For each task, they were expected to translate the entire document including self-revision to create a translation to the same level of quality they typically deliver to the industry partner. All translators were encouraged to take breaks to counter fatigue effects. They all took short breaks during the translation tasks and longer breaks between the tasks. After each task, they filled in the post-task questionnaire for capturing their user experience. The day was concluded with a short interview, followed by dismantling the workstation and packing the research equipment.

3.5 Measurements

The eyetracker logged the time taken for the translation tasks, the keystrokes and mouse actions, the gaze data such as fixations, and created a screen recording of everything happening on the main screen the participants were working on.

The resulting translations of each task were saved for a subsequent error analysis, which will be reported on in a future publication.

Before the first translation task and after each of the three translation tasks, participants responded to questionnaires. The pre-task questionnaire asked for demographic data, professional experience, their translation process, experience with machine translation and post-editing as well as their perception of MT and PE. We compiled this questionnaire by combining items relevant for our study from the validated Translation and Interpreting Competence Questionnaire (TICQ; Schaeffer et al., 2019), Briva-Iglesias' (2024) pre-task questionnaire, and a questionnaire directed at literary translators about their use of translation technology (Ruokonen and Salmi, 2024).

Immediately after each of the translation tasks, participants responded to a self-reported user experience questionnaire (UEQ; Laugwitz et al., 2008). The UEQ is a validated questionnaire developed to measure the user experience and usability of an interactive software product. In the UEQ, participants use a 7-step scale to indicate their level of agreement with 26 items. The items fall under 6 dimensions that cover aspects such as the overall impression, how easy it is to learn, how efficient the usage is, if the user feels in control, if the usage is exciting and motivating, and if it is innovative and creative (Schrepp et al., 2014). The 26 items represent pairs of adjectives that are extreme opposites to each other, such as "1 annoying – 7 enjoyable". The order in which the 26 items are presented is randomized for each individual post-task questionnaire. Half of the items are presented with the negative adjective on the left of the 7-step scale and the positive adjective on the right, while it is the other way around for the other half of the items. The full list of adjective pairs is available in appendix A. Although the UEQ is available in different language versions, the post-task questionnaires were provided to all study participants in English only. This contributed to language conformity between all questionnaires.

After the translation tasks and the respective questionnaires were completed, participants gave a short interview in which they were asked about their overall impression of the three different tasks, about the types of PE they had done before, whether they would consider using MT if they could choose freely, what needed to improve for MT to be useful in their work, and what potential they saw in large language models and generative AI applications for supporting their work as video game translators. The interviews were conducted in English with all participants. This created equal conditions for all participants as otherwise the German participants might have had an advantage in being interviewed in their native language.

4 Results

Here we report on the perception of PE and the user experience observed in this study. A presentation and discussion of productivity and eyetracking data as well as translation quality is out of scope of this article (see Brenner 2024 and Forthcoming for further analysis plans). Statistics were calculated and tables and figures were generated with jamovi version 2.6.26 and the vijPlots module.

4.1 PE Perception before Translation Tasks

Four items in the pre-task questionnaire were designed to measure the participants' perception of PE. These items were taken from the PE pre-task questionnaire by Briva-Iglesias (2024).

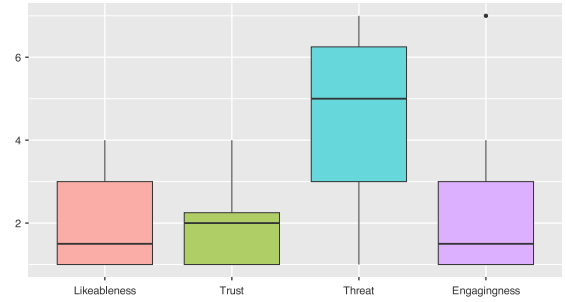

Figure 2: Pre-task perception of PE; n = 12.

The four items were as follows:

a) On a scale of 1–7, where 1 is "Strongly Dislike" and 7 is "Strongly Like", please rate your perception of doing MTPE tasks in professional game translation projects.
b) On a scale of 1–7, where 1 is "Not trustworthy at all" and 7 is "Very trustworthy", please rate if you can trust MTPE to help you successfully deliver a professional game translation project.
c) Please rate how much you agree or disagree with this statement: Machine Translation is a threat to the sustainability of the profession of game translators. (scale 1–7, fully disagree to fully agree)
d) Please rate the following statement: When I am doing MTPE tasks, I find them… (In case you have never done MTPE, answer what you think how you might find it.) (scale 1–7, boring to engaging)

Most participants do not like PE in game translation: Out of 12 participants, 6 assessed this item with the lowest rating of 1, "strongly dislike". No participant gave a higher rating than 4.

The trust in PE to be a good help in game translation is also low, with 9 participants assessing this item with 1 or 2. No participant gave a higher trust rating than 4.

Most participants regard machine translation as a threat to the sustainability of the profession of game translators, with 7 participants assessing this item with 5, 6, or 7. One participant neither agrees nor disagrees, rating this item with 4. The remaining 4 participants do not see a large threat, rating this item with 1, 2, or 3.

With 6 participants giving the lowest possible rating of 1, 2 participants giving a rating of 2, and another 2 participants a rating of 3, most participants find PE boring. One participant finds PE neither boring nor engaging, and only one participant finds this type of translation engaging.

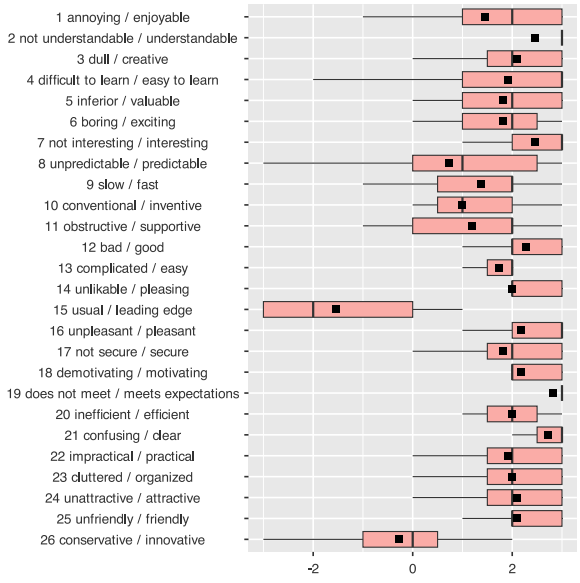

Figure 1: User experience measured after C1 – translation from scratch; n = 11.

Interestingly, this participant had never done PE before becoming a participant in this study. This is also the participant who gave the lowest threat rating. During the interview, it was confirmed that these ratings reflect the participant's opinion.

Figure 2 illustrates the participants' perceptions of PE in game translation as reported at the beginning of the experiment sessions.

4.2 User Experience after Translation Tasks

The UEQ is designed by its original authors with half of the items showing the positive adjective on the left while answering (see appendix A). For analysis, these adjective pairs need to be switched so that all 26 items show the negative adjective on the left. For this, the 7-step scale is converted to a scale from −3 to +3. According to the UEQ authors, all values between −0.8 and +0.8 are interpreted as a neutral experience, whereas values smaller than −0.8 represent a negative experience and values above +0.8 a positive experience (Schrepp, 2023).

4.2.1 User Experience of Translation from Scratch

For the translation condition C1 – translation from scratch, there are 11 post-task questionnaires available for analysis as 1 questionnaire is missing. On average all aspects of user experience when translating from scratch were rated positively, except for "8 unpredictable / predictable", which on average was rated neutral with a mean of 0.73, but a large standard deviation of 2.20. The item "26 conservative / innovative" on average was rated

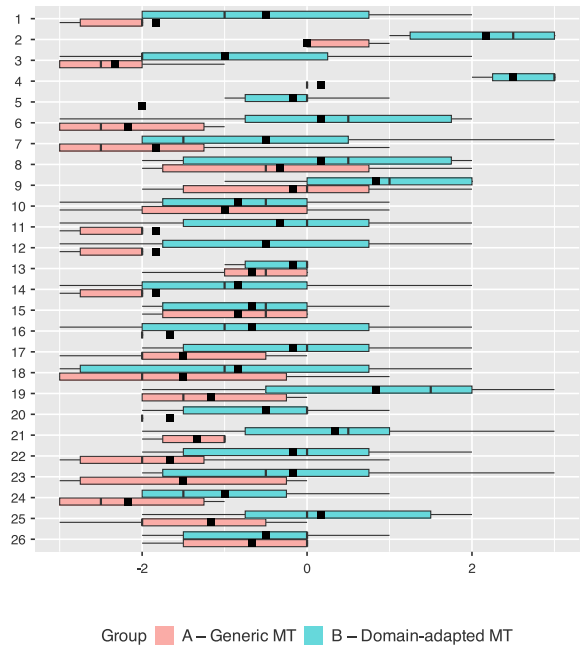

Figure 3: User experience measured after C2 – static PE; n = 12.

neutral as well (M: −0.27; SD: 1.49). The item "15 usual / leading edge" (M: −1.55; SD: 1.57) is the only item with average negative ratings. See Figure 1 for a visualization of all ratings for the user experience of translation from scratch. It clearly shows how almost all aspects were rated on the positive side of the scale.

4.2.2 User Experience of Static PE

For translation condition C2 – static PE, the participants were divided into two groups. Group A (n = 6) used the generic ModernMT output for PE, while group B (n = 6) post-edited the domain-adapted ModernMT output. Figure 3 shows the user experience ratings after task C2 – static PE, with groups A and B side by side for comparison. To save space in Figure 3, items are numbered from 1 to 26. The order of the items is the same as in Figure 1. Appendix A lists the item numbers with their corresponding adjective pairs for reference.

As can be seen in Figure 3, those participants who performed static PE with the generic version of ModernMT (group A) rated the user experience negatively for almost all items (mean, indicated by the black square, below −0.8). Exceptions are items 2 (not understandable / understandable; M: 0.00; SD: 1.67), 4 (difficult to learn / easy to learn; M: 0.17; SD: 0.41), 8 (unpredictable / predictable; M: −0.33; SD: 1.63), 9 (slow / fast; M: −0.17; SD: 1.60), 13 (complicated / easy; M: −0.67; SD: 0.82), and 26 (conservative / innovative; M: −0.67; SD:1.03). With means between −0.8 and +0.8, these 6 items were rated neutral. Based on the mean values, for static PE with generic MT, no user experience aspect was rated positively. Values used in Figure 3 are given in appendix B.

	Group	N	Mean	SD	Minimum	Maximum
11	A – Generic MT	6	-1.83	1.47	-3	1
	B – Domain-adapted MT	6	-0.33	1.86	-3	2
12	A – Generic MT	6	-1.83	1.47	-3	1
	B – Domain-adapted MT	6	-0.50	1.87	-3	2
14	A – Generic MT	6	-1.83	1.47	-3	1
	B – Domain-adapted MT	6	-0.83	1.83	-3	2
16	A – Generic MT	6	-1.67	1.37	-3	1
	B – Domain-adapted MT	6	-0.67	1.97	-3	2

Table 2: Comparison between groups A and B in C2 for items 11 (obstructive / supportive), 12 (bad / good), 14 (unlikable / pleasing), and 16 (unpleasant / pleasant).

Compared to this, group B, who used the domain-adapted version of ModernMT for static PE, rated the user experience more positively for each of the 26 items. The only items rated negatively are items 3 (dull / creative; M: −1.00; SD: 2.00) and 24 (unattractive / attractive; M: −1.00; SD: 1.26), whereas the only items rated positively are items 2 (not understandable / understandable; M: 2.17; SD: 0.98), and 4 (difficult to learn / easy to learn; M: 2.50; SD: 0.84). The remaining 22 items are rated neutral by group B. However, group B shows a wider variance in their user experience ratings compared to group A. This can be most clearly seen in items 11 (obstructive / supportive), 12 (bad / good), 14 (unlikable / pleasing), and 16 (unpleasant / pleasant). Table 2 shows the mean values, the standard deviation, the minimum, and the maximum values for these exemplary items. It is noticeable that the minimum value is −3 for all items, regardless of the group, whereas the maximum value is lower for all items in group A compared to group B. The mean indicates that, on average, group A rated all these items negatively, whereas group B rated them neutral. However, the standard deviation in group B is higher than in group A for all four items.

4.2.3 User Experience of Flexible PE

In translation condition C3 – flexible PE, participants belonged to the same group as in condition C2. Group A (n = 6) received suggestions from the generic ModernMT version, while group B (n = 6) saw the domain-adapted ModernMT

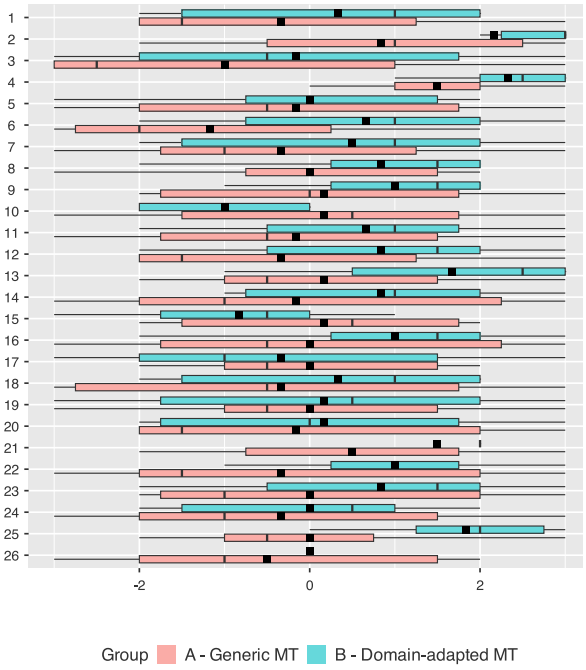

Figure 4: User experience measured after C3 – flexible PE for groups A (n = 6) and B (n = 6).

output. Figure 4 shows the user experience ratings after task C3 – flexible PE, with groups A and B side by side for comparison. Item numbering and ordering is the same as in section 4.2.2.

In C3 (flexible PE), participants who worked with generic MT (group A) rated their experience lower than participants who worked with domain-adapted MT (group B). This matches the same tendency in C2 – static PE. The group difference is especially profound in items 2 (not understandable / understandable), 6 (boring / exciting), 13 (complicated / easy), 22 (impractical / practical), and 25 (unfriendly / friendly), where the differences in mean values between the two groups are larger than 1. Table 3 shows the mean values, the standard deviation, the minimum, and the maximum values for these exemplary items. There are only two items that on average were rated negatively by group B and neutral by group A, item 10 (conventional / inventive; M_A: 0.17; M_B: −1.00) and item 15 (usual / leading edge; M_A: 0.17; M_B: −0.83). However, looking at the standard deviation (item 10: SD_A: 2.32; SD_B: 1.10; item 15: SD_A: 1.83; SD_B: 1.47), we see that the ratings are more varied in group A than in group B.

Overall, in C3 – flexible PE, the ratings are varied within both groups. This can be seen in Figure 4, where for each of the 26 items the horizontal lines reach either the left or the right extreme of the boxplot for at least one of the two

	Group	N	Mean	SD	Minimum	Maximum
2	A - Generic MT	6	0.83	2.04	-2	3
	B - Domain-adapted MT	6	2.17	1.60	-1	3
6	A - Generic MT	6	-1.17	2.14	-3	2
	B - Domain-adapted MT	6	0.67	1.97	-2	3
13	A - Generic MT	6	0.17	1.94	-2	3
	B - Domain-adapted MT	6	1.67	1.75	-1	3
22	A - Generic MT	6	-0.33	2.66	-3	3
	B - Domain-adapted MT	6	1.00	1.41	-1	3
25	A - Generic MT	6	0.00	1.79	-2	3
	B - Domain-adapted MT	6	1.83	1.17	0	3

Table 3: Comparison between groups A and B in C3 for items 2 (not understandable / understandable), 6 (boring / exciting), 13 (complicated / easy), 22 (impractical / practical), and 25 (unfriendly / friendly).

groups. Also, the boxes representing the lower and upper quartiles are elongated. Values used in Figure 4 are given in appendix C.

5 Discussion and Conclusion

The PE perception measured at the beginning of the experiment sessions shows that participants do not like PE. They do not trust it being helpful and they find the task of performing PE boring. Most of the participants regard machine translation to be a threat for the sustainability of their profession. This negative view on PE is in line with other studies where participants were asked about their perception of PE. In a survey conducted by Alvarez-Vidal et al. (2020) the satisfaction level with PE was considerably lower than with translation from scratch. In a pre-PE questionnaire used in a study by Ciobanu et al. (2024) less than half of the participants with previous PE experience agreed to liking PE. In a study with literary translators, preference also goes to translation from scratch (Guerberof-Arenas and Toral, 2022).

Briva-Iglesias and O'Brien (2024) point out that "past experiences and perceptions have a great impact and are a determinant for future beliefs, attitudes and behaviours" (p. 445). We see this reflected in our findings in two ways. First, participants in our study, on average, had the most positive user experience with translation from scratch compared to static PE and flexible PE. Translation from scratch is what the highly educated participants have been trained for and what they mostly do in their daily work. Thus, this is the past experience that shapes their user experience while performing the task. Second, we see a negative attitude towards PE uttered before

the task reflected in a negative user experience when having done the PE task.

The negative attitude towards PE found in our study is in line with the findings of a survey among game translators world-wide, where the majority of survey respondents finds MT "inconvenient", "not important", and "not useful" (Rivas Ginel, 2023, p. 256). The same study also revealed that despite their negative perception, game translators do use MT tools, a finding in line with the results of our pre-task questionnaire.

Measuring the user experience after each of the three translation conditions showed that C1 – translation from scratch is a highly positive experience for all participants. Compared to that, C2 – static PE, the type of PE commonly employed in the translation industry, is a negative experience for those translators who used a generic MT version. For group B, who used the domain-adapted MT version, the user experience of static PE leans more towards neutral, with some translators finding the experience positive. Translation condition C3 – flexible PE is a combination of translation from scratch with PE only of those sentences where the MT output is considered useful by the translator. On average, participants had a better user experience with C3 than with C2. This difference is similar to the results of the MTUX study by Briva-Iglesias et al. (2023), who compared the traditional PE approach (similar to our static PE) with interactive PE, where translators have more control over the use of MT suggestions. In this study, the user experience of interactive PE was significantly higher than that of traditional PE, and the authors attribute this to empowerment and control for the translators.

In our study, user experience was improved when PE was combined with domain-adapted MT, showing that domain adaptation can have benefits for the translators who perform PE. Our results about perception of PE and the MT user experience are mostly in line with studies relating to other domains. In future publications we will analyze the generated data regarding productivity, quality, and translator's cognitive effort. This will contribute to a wider picture of the PE process when translating video games and potentially show differences to or commonalities with other fields of translation.

6 Limitations

In this study, we simulated a real-world game translation project. As this could only approximate what actually happens when freelance game translators translate and post-edit, there naturally are some limitations.

Study participants were recruited from the pool of freelance translators who have an established business relationship with the industry partner. It was important that an established trust relationship before taking part in the experiment could continue after the experiment. Therefore, candidates with interest in participating who had not worked with the industry partner before could not be considered, limiting the opportunities for anyone to participate.

The limited number of 12 participants with 4 language pairs does not allow a generalization for all video game translators worldwide. Instead, it highlights individual differences between translators that should be considered when assessing PE for video game translation.

As each participant was only exposed to either generic or domain-adapted MT, it might be possible that group A happens to consist of more people with a general negative tendency.

Participants were aware of taking part in an experiment, as the technical setup on their desks was different than usual.

According to Finnish national guidelines for ethical review, a formal ethics review was not required for this type of research (Finnish National Board on Research Integrity TENK, 2019). Still, ethical advisors at the University of Eastern Finland were consulted regarding participants' remuneration. It was agreed that compensation would not add pressure to participate but, on the contrary, promote research feasibility and quality. During the experiment it was clear that participants treated it like a regular assignment and aspired to deliver their typical quality. Without compensation, it would not have been possible for them to dedicate a full working day to the research.

Acknowledgments

This research was funded by the Finnish Kone Foundation (2023–2025, project number 202202303) and the European Association for Machine Translation (EAMT), Sponsorship of Activities, Students' Edition 2023, and supported by Native Prime. We wish to thank the 15 study participants for their participation.

References

Yulia Akhulkova. 2021. Machine Translation for Games – An interview with Mikhail Gorbunov.

Sergi Alvarez-Vidal, Antoni Oliver, and Toni Badia. 2020. Post-editing for Professional Translators: Cheer or Fear? *Revista Tradumàtica*(18):49–69. https://doi.org/10.5565/rev/tradumatica.275

Cristina Anselmi and Inés Rubio. 2020. The Future is Here: Neural Machine Translation for Games.

Vicent Briva-Iglesias. 2024. Fostering human-centered, augmented machine translation: Analysing interactive post-editing. Ph.D. thesis, Dublin City University, Dublin, Ireland.

Vicent Briva-Iglesias and Sharon O'Brien. 2024. Pre-task perceptions of MT influence quality and productivity: the importance of better translator-computer interactions and implications for training. In Carolina Scarton, Charlotte Prescott, Chris Bayliss, Chris Oakley, Joanna Wright, Stuart Wrigley, Xingyi Song, Edward Gow-Smith, Rachel Bawden, Víctor M Sánchez-Cartagena, Patrick Cadwell, Ekaterina Lapshinova-Koltunski, Vera Cabarrão, Konstantinos Chatzitheodorou, Mary Nurminen, Diptesh Kanojia, and Helena Moniz, editors, *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)*, pages 444–454, Sheffield, UK. European Association for Machine Translation (EAMT).

Vicent Briva-Iglesias, Sharon O'Brien, and Benjamin R. Cowan. 2023. The impact of traditional and interactive post-editing on Machine Translation User Experience, quality, and productivity. *Translation, Cognition & Behavior*, 6(1):60–86. https://doi.org/10.1075/tcb.00077.bri

Judith Brenner. 2024. The MTxGames Project: Creative Video Games and Machine Translation – Different Post-Editing Methods in the Translation Process. In Carolina Scarton, Charlotte Prescott, Chris Bayliss, Chris Oakley, Joanna Wright, Stuart Wrigley, Xingyi Song, Edward Gow-Smith, Mikel Forcada, and Helena Moniz, editors, *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 2)*, pages 47–48, Sheffield, UK. European Association for Machine Translation (EAMT).

Judith Brenner. Forthcoming. MTxGames: Machine Translation Post-Editing in Video Game Translation – Findings on User Experience and Preliminary Results on Productivity. Accepted at MT Summit 2025. Geneva, Switzerland.

Dragoș Ciobanu, Miguel Rios, Alina Secară, Justus Brockmann, Raluca-Maria Chereji, and Claudia Wiesinger. 2024. The impact of using text-to-speech (TTS) in post-editing machine translation (PEMT) workflows on translators' cognitive effort, productivity, quality, and perceptions. *Revista Tradumàtica*(22):323–254. https://doi.org/10.5565/rev/tradumatica.394

Joke Daems. 2016. A translation robot for each translator? A comparative study of manual translation and post-editing of machine translations: process, quality and translator attitude. dissertation, Ghent University. Faculty of Arts and Philosophy, Ghent, Belgium.

Lucile Danilov. 2023. Gameloc Manifesto.

Max Deryagin, Miroslav Pošta, and Daniel Landes. 2021. Machine Translation Manifesto. Audiovisual Translators Europe.

Diana Díaz Montón. 2024. Video game localizers. In *Handbook of the Language Industry*, pages 225–250. De Gruyter Mouton, Berlin, Boston. https://doi.org/10.1515/9783110716047-011

En Chair et en Os. 2023. Literature, Film, Press, Video Games: Say No to Soulless Translations.

Finnish National Board on Research Integrity TENK. 2019. The Ethical Principles of Research with Human Participants and Ethical Review in the Human Sciences in Finland. https://tenk.fi/en/ethical-review/ethical-review-human-sciences

Ana Guerberof-Arenas and Antonio Toral. 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

Hanna Hagström and Jan Pedersen. 2022. Subtitles in the 2020s: The Influence of Machine Translation. *Journal of Audiovisual Translation*, 5(1):207–225. https://doi.org/10.47476/jat.v5i1.2022.195

Damien Hansen and Pierre-Yves Houlmont. 2022. A Snapshot into the Possibility of Video Game Machine Translation. In Janice Campbell, Stephen Larocca, Jay Marciano, Konstantin Savenkov, and Alex Yanishevsky, editors, *Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Volume 2: Users and Providers Track and Government Track)*, pages 257–269, Orlando, USA. Association for Machine Translation in the Americas.

Sylvia Jaki, Maren Bolz, and Sophie Röther. 2024. KI-Technologien in der Audiovisuellen Translation – Umfrageergebnisse aus der deutschen Translationsindustrie. *trans-kom*, 17(2):320–342.

Miguel A. Jimenez-Crespo. 2024. *Localization in Translation*. Routledge. https://doi.org/10.4324/9781003340904

Samuel Läubli, Chantal Amrhein, Patrick Düggelin, Beatriz Gonzalez, Alena Zwahlen, and Martin Volk. 2019. Post-editing Productivity with Neural Machine Translation: An Empirical Assessment of Speed and Quality in the Banking and Finance Domain. In Mikel Forcada, Andy Way, Barry

Haddow, and Rico Sennrich, editors, *Proceedings of Machine Translation Summit XVII: Research Track*, pages 267–272, Dublin, Ireland. European Association for Machine Translation.

Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In Andreas Holzinger, editor, *USAB 2008, Lecture Notes in Computer Science, Vol. 5298*, pages 63–76, Graz, Austria. https://doi.org/10.1007/978-3-540-89350-9_6

Lionbridge Games. 2024. Games Machine Translation Services.

Lieve Macken, Daniel Prou, and Arda Tezcan. 2020. Quantifying the Effect of Machine Translation in a High-Quality Human Translation Production Process. *Informatics*, 7(2). https://doi.org/10.3390/informatics7020012

Carme Mangiron. 2022. Audiovisual translation and multimedia and game localisation. In *The Routledge Handbook of Translation and Methodology*, pages 410–424. Routledge.

Joss Moorkens, Antonio Toral, Sheila Castilho, and Andy Way. 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

Joss Moorkens, Andy Way, and Séamus Lankford. 2024. *Automating Translation*. Routledge, London. https://doi.org/10.4324/9781003381280

Sharon O'Brien. 2023. Human-Centered augmented translation: against antagonistic dualisms. *Perspectives*:1–16. https://doi.org/10.1080/0907676X.2023.2247423

María Isabel Rivas Ginel. 2023. *The ergonomics of CAT tools for video game localisation*. PhD thesis, Université Bourgogne Franche-Comté.

María Isabel Rivas Ginel and Sarah Theroine. 2022. Machine Translation and Gender biases in video game localisation: a corpus-based analysis. *Journal of Data Mining & Digital Humanities*, Towards robotic translation? https://doi.org/10.46298/jdmdh.9065

Andrew Rothwell. 2023. Retranslating Proust Using CAT, MT, and Other Tools. In Andrew Rothwell, Andy Way, and Roy Youdale, editors, *Computer-Assisted Literary Translation*, pages 106–125. Routledge, New York. https://doi.org/10.4324/9781003357391

Paola Ruffo, Joke Daems, and Lieve Macken. 2024. Measured and perceived effort: assessing three literary translation workflows. *Revista Tradumàtica*(22):238–257. https://doi.org/10.5565/rev/tradumatica.378

Minna Ruokonen and Leena Salmi. 2024. Finnish literary translators' use of translation technology and tools: processes, profiles, and purposes. *Mikael: Finnish Journal of Translation and Interpreting Studies*, 17(1):138–154. https://doi.org/10.1080/0907676X.2019.1629468

Moritz Schaeffer, David Huepe, Silvia Hansen-Schirra, Sascha Hofmann, Edinson Muñoz, Boris Kogan, Eduar Herrera, Agustín Ibáñez, and Adolfo M. García. 2019. The Translation and Interpreting Competence Questionnaire: an online tool for research on translators and interpreters. *Perspectives*, 28(1):90–108. https://doi.org/10.1080/0907676X.2019.1629468

Martin Schrepp. 2023. User Experience Questionnaire Handbook.

Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2014. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, volume 8517, pages 383–392, Cham. Springer. https://doi.org/10.1007/978-3-319-07668-3_37

Carlos S. C. Teixeira and Sharon O'Brien. 2019. Investigating the cognitive ergonomic aspects of translation tools in a workplace setting. In Hanna Risku, Regina Rogl, and Jelena Milosevic, editors, *Translation Practice in the Field: Current research on socio-cognitive processes*, Benjamins Current Topics, pages 79–103. John Benjamins Publishing Company. https://doi.org/10.1075/bct.105.05tei

Silvia Terribile. 2024. Is post-editing really faster than human translation? *Translation Spaces*, 13(2):171–199. Version of Record published: 19 Dec 2023.

Antonio Toral, Martijn Wieling, and Andy Way. 2018. Post-editing Effort of a Novel With Statistical and Neural Machine Translation. *Frontiers in Digital Humanities*, 5(9). https://doi.org/10.3389/fdigh.2018.00009

Lucas Nunes Vieira. 2020. Post-editing of machine translation. In Minako O'Hagan, editor, *The Routledge Handbook of Translation and Technology*, Routledge Handbooks in Translation and Interpreting Studies, pages 319–336. Routledge. https://doi.org/10.4324/9781315311258-22

Xiaochun Zhang. 2022. Video game localization: Translating interactive entertainment. In *The Routledge Handbook of Translation and Media*, pages 369–383. Routledge. https://doi.org/10.4324/9781003221678-27

Amir Arsalan Zoraqi and Mohsen Kafi. 2024. Profiles, Perceptions, and Experiences of Video Game Translators. *Games and Culture*:1–22. https://doi.org/10.1177/1555412023122

A Appendix: Items of the user experience questionnaire

List of the questionnaire items as used for analysis in section 4.2, with the negative adjective on the left and the positive adjective on the right:

1. annoying / enjoyable
2. not understandable / understandable
3. dull / creative
4. difficult to learn / easy to learn
5. inferior / valuable
6. boring / exciting
7. not interesting / interesting
8. unpredictable / predictable
9. slow / fast
10. conventional / inventive
11. obstructive / supportive
12. bad / good
13. complicated / easy
14. unlikable / pleasing
15. usual / leading edge
16. unpleasant / pleasant
17. not secure / secure
18. demotivating / motivating
19. does not meet expectations / meets expectations
20. inefficient / efficient
21. confusing / clear
22. impractical / practical
23. cluttered / organized
24. unattractive / attractive
25. unfriendly / friendly
26. conservative / innovative

List of the questionnaire items as presented while answering the UEQ, where half of the items have the negative adjective on the left and the other half have the negative adjective on the right:

1. annoying / enjoyable
2. not understandable / understandable
3. creative / dull
4. easy to learn / difficult to learn
5. valuable / inferior
6. boring / exciting
7. not interesting / interesting
8. unpredictable / predictable
9. fast / slow
10. inventive / conventional
11. obstructive / supportive
12. good / bad
13. complicated / easy
14. unlikable / pleasing
15. usual / leading edge
16. unpleasant / pleasant
17. secure / not secure
18. motivating / demotivating
19. meets expectations / does not meet expectations
20. inefficient / efficient
21. clear / confusing
22. impractical / practical
23. organized / cluttered
24. attractive / unattractive
25. friendly / unfriendly
26. conservative / innovative

B Appendix: Table with user experience values after C2 – static post-editing

The values in the following table are visualized in the boxplot in Figure 3.

User experience with C2 – static post-editing

	Group	Mean	Median	SD	Min.	Max.
1 annoying / enjoyable	**A**	-1.83	-2.00	1.47	-3	1
	B	-0.50	-1.00	1.76	-2	2
2 not understandable / understandable	**A**	0.00	0.00	1.67	-3	2
	B	2.17	2.50	0.98	1	3
3 dull / creative	**A**	-2.33	-2.50	0.82	-3	-1
	B	-1.00	-2.00	2.00	-3	2
4 difficult to learn / easy to learn	**A**	0.17	0.00	0.41	0	1
	B	2.50	3.00	0.84	1	3
5 inferior / valuable	**A**	-2.00	-2.00	0.63	-3	-1
	B	-0.17	0.00	0.75	-1	1
6 boring / exciting	**A**	-2.17	-2.50	0.98	-3	-1
	B	0.17	0.50	1.94	-3	2
7 not interesting / interesting	**A**	-1.83	-2.50	1.60	-3	1
	B	-0.50	-1.50	2.07	-2	3
8 unpredictable / predictable	**A**	-0.33	-0.50	1.63	-2	2
	B	0.17	0.50	1.83	-2	2
9 slow / fast	**A**	-0.17	0.00	1.60	-2	2
	B	0.83	1.00	1.33	-1	2
10 conventional / inventive	**A**	-1.00	-1.00	1.55	-3	1
	B	-0.83	-0.50	1.47	-3	1
11 obstructive / supportive	**A**	-1.83	-2.00	1.47	-3	1
	B	-0.33	0.00	1.86	-3	2
12 bad / good	**A**	-1.83	-2.00	1.47	-3	1
	B	-0.50	-0.50	1.87	-3	2
13 complicated / easy	**A**	-0.67	-0.50	0.82	-2	0
	B	-0.17	0.00	1.33	-2	2
14 unlikable / pleasing	**A**	-1.83	-2.00	1.47	-3	1
	B	-0.83	-1.00	1.83	-3	2
15 usual / leading edge	**A**	-0.83	-0.50	0.98	-2	0
	B	-0.67	-0.50	1.21	-2	1

User experience with C2 – static post-editing

	Group	Mean	Median	SD	Min.	Max.
16 unpleasant / pleasant	**A**	-1.67	-2.00	1.37	-3	1
	B	-0.67	-1.00	1.97	-3	2
17 not secure / secure	**A**	-1.50	-2.00	1.22	-3	0
	B	-0.17	0.00	1.60	-2	2
18 demotivating / motivating	**A**	-1.50	-2.00	1.76	-3	1
	B	-0.83	-1.00	2.14	-3	2
19 does not meet expectations / meets expectations	**A**	-1.17	-1.50	0.98	-2	0
	B	0.83	1.50	1.94	-2	3
20 inefficient / efficient	**A**	-1.67	-2.00	1.37	-3	1
	B	-0.50	0.00	1.22	-2	1
21 confusing / clear	**A**	-1.33	-1.00	0.52	-2	-1
	B	0.33	0.50	1.75	-2	3
22 impractical / practical	**A**	-1.67	-2.00	1.51	-3	1
	B	-0.17	0.00	1.60	-2	2
23 cluttered / organized	**A**	-1.50	-1.50	1.38	-3	0
	B	-0.17	-0.50	1.94	-2	3
24 unattractive / attractive	**A**	-2.17	-2.50	0.98	-3	-1
	B	-1.00	-1.50	1.26	-2	1
25 unfriendly / friendly	**A**	-1.17	-2.00	1.83	-3	2
	B	0.17	0.00	1.60	-2	2
26 conservative / innovative	**A**	-0.67	0.00	1.03	-2	0
	B	-0.50	0.00	1.22	-2	1

C Appendix: Table with user experience values after C3 – flexible post-editing

The values in the following table are visualized in the boxplot in Figure 4.

User experience with C3 – flexible post-editing

	Group	Mean	Median	SD	Min.	Max.
1 annoying / enjoyable	**A**	-0.33	-1.50	2.25	-2	3
	B	0.33	1.00	1.97	-2	2
2 not understandable / understandable	**A**	0.83	1.00	2.04	-2	3
	B	2.17	3.00	1.60	-1	3
3 dull / creative	**A**	-1.00	-2.50	2.76	-3	3
	B	-0.17	-0.50	2.48	-3	3
4 difficult to learn / easy to learn	**A**	1.50	1.50	1.05	0	3
	B	2.33	2.50	0.82	1	3
5 inferior / valuable	**A**	-0.17	-0.50	2.48	-3	3
	B	0.00	0.00	1.90	-3	2
6 boring / exciting	**A**	-1.17	-2.00	2.14	-3	2
	B	0.67	1.00	1.97	-2	3
7 not interesting / interesting	**A**	-0.33	-1.00	2.34	-3	3
	B	0.50	1.00	2.17	-2	3
8 unpredictable / predictable	**A**	0.00	0.00	1.90	-3	2
	B	0.83	1.50	1.60	-2	2
9 slow / fast	**A**	0.17	0.00	2.14	-2	3
	B	1.00	1.50	1.26	-1	2
10 conventional / inventive	**A**	0.17	0.50	2.32	-3	3
	B	-1.00	-1.00	1.10	-2	0
11 obstructive / supportive	**A**	-0.17	-0.50	2.32	-3	3
	B	0.67	1.00	1.86	-2	3
12 bad / good	**A**	-0.33	-1.50	2.25	-2	3
	B	0.83	1.50	1.94	-2	3
13 complicated / easy	**A**	0.17	-0.50	1.94	-2	3
	B	1.67	2.50	1.75	-1	3
14 unlikable / pleasing	**A**	-0.17	-1.00	2.64	-3	3
	B	0.83	1.00	1.72	-1	3
15 usual / leading edge	**A**	0.17	0.50	1.83	-2	2
	B	-0.83	-0.50	1.47	-3	1

User experience with C3 – flexible post-editing

	Group	Mean	Median	SD	Min.	Max.
16 unpleasant / pleasant	**A**	0.00	-0.50	2.53	-3	3
	B	1.00	1.50	1.79	-2	3
17 not secure / secure	**A**	0.00	-0.50	1.67	-2	2
	B	-0.33	-1.00	2.42	-3	3
18 demotivating / motivating	**A**	-0.33	-0.50	2.66	-3	3
	B	0.33	1.00	1.97	-2	2
19 does not meet expectations / meets expectations	**A**	0.00	-0.50	2.19	-3	3
	B	0.17	0.50	2.48	-3	3
20 inefficient / efficient	**A**	-0.17	-1.50	2.48	-2	3
	B	0.17	0.00	2.14	-2	3
21 confusing / clear	**A**	0.50	0.50	1.87	-2	3
	B	1.50	2.00	1.76	-2	3
22 impractical / practical	**A**	-0.33	-1.50	2.66	-3	3
	B	1.00	1.00	1.41	-1	3
23 cluttered / organized	**A**	0.00	-1.00	2.37	-2	3
	B	0.83	1.50	1.94	-2	3
24 unattractive / attractive	**A**	-0.33	-1.00	2.42	-3	3
	B	0.00	0.50	1.67	-2	2
25 unfriendly / friendly	**A**	0.00	-0.50	1.79	-2	3
	B	1.83	2.00	1.17	0	3
26 conservative / innovative	**A**	-0.50	-1.00	2.17	-3	2
	B	0.00	0.00	0.63	-1	1