Pre-task perceptions of MT influence quality and productivity: the importance of better translator-computer interactions and implications for training

Vicent Briva-Iglesias
SALIS, D-REAL
Dublin City University
vicent.brivaiglesias2@mail.dcu.ie

Sharon O'Brien
SALIS
Dublin City University
sharon.obrien@dcu.ie

Abstract

This paper presents a user study with 11 professional English-Spanish translators in the legal domain. We analysed whether negative or positive translators' pre-task perceptions of machine translation (MT) being an aid or a threat had any relationship with final translation quality and productivity in a post-editing workflow. Pre-task perceptions of MT were collected in a questionnaire before translators conducted post-editing tasks and were then correlated with translation productivity and translation quality after an Adequacy-Fluency evaluation. Each participant translated 13 texts over two consecutive weeks, accounting for 120,102 words in total. Results show that translators who had higher levels of trust in MT and thought that MT was not a threat to the translation profession reported higher translation quality and productivity. These results have critical implications: improving translator-computer interactions and fostering MT literacy in translation training may be crucial to reducing negative translators' pre-task perceptions, resulting in better translation productivity and quality, especially adequacy.

1 Introduction

MT has become an undisputed element of today's workflows in the language services industry (ELIS Research, 2023). Different studies suggest that improvements in these systems over time have allowed translators to see their productivity increase without a negative impact on the quality produced and, therefore, most research in the field of MT has focused on estimating the productivity and quality of MT systems (Moorkens et al., 2018a; Rossi and Carré, 2022). However, this adoption of MT is not always accompanied by positive user feedback, as some translators have shown little satisfaction in working and interacting with MT through post-editing workflows, either because of a reduction in pay, a sense of dehumanisation of the translation process, or the commodification and uberisation of the language services industry (Moorkens, 2020; Fırat, 2021; Cadwell et al., 2018).

In the Translation Studies and the MT fields, research centered on analysing human factors in today's translator-computer interactions is still relatively limited, and the perceptions, user experiences (UX) or feelings of MT users, or even whether these feelings and experiences have any effect on their interactions, have been scarce (Koponen et al., 2020; Karakanta et al., 2022; Briva-Iglesias and O'Brien, 2023; Briva-Iglesias et al., 2023; Guerberof Arenas et al., 2021). In this context, we present the results of a study (part of a larger project) (Briva-Iglesias, 2024) that explores whether translators' pre-task perceptions of MT have any relationship with final translation quality and productivity. Below, we first present work related to our study, then we outline the methodology and, finally, results are described and discussed.

2 Related Work

In the last decades of research in natural language processing (NLP), the focus has been on making technical advancements, mainly by increasing the size of the language models and the computational power used to obtain better results (Brown et al., 2020), but often neglecting the repercus-

Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1), pages 444–454
June 24-27, 2024 ©2024 European Association for Machine Translation

sions or risks that this path has or may have on humans (Bender et al., 2021; Shneiderman, 2022). Research in translation technologies has followed a parallel path to NLP research, and most studies have focused on evaluating the quality of MT or comparing different MT paradigms (Drugan, 2013; Moorkens et al., 2018a; Rossi and Carré, 2022). Such research is necessary, but technical changes should also be accompanied by socio-technical studies and their impact on users. Olohan (2011) criticized this path by commenting that "the human and organisational aspects are not addressed at all, or only implicitly, [...] when the system is being developed".

This has meant that the study of human factors and their interaction with technology has lagged behind and received less attention in translation technology research. However, it has not been completely forgotten. For example, Gaspari et al. (2014) analysed the perceptions of 438 users of online MT systems, and the majority of participants commented that they were not happy with the results, especially with the quality offered. Moorkens et al. (2018b) studied the perception of post-editing effort in the literary field, and collected the data with questionnaires and short interviews, which they then analysed qualitatively. Through a questionnaire completed by 1850 people, O'Brien et al. (2017) investigated how translators interacted with CAT tools, and found that there were certain levels of cognitive friction and that some functionalities of CAT tools irritated them.

Not only freelance or corporate translators have received the attention of academia, but also translators in governmental organisations and international institutions. Rossi and Chevrot (2019) surveyed French translators at the European Commission to analyse the level of acceptance of MT, and suggested that fear of the technology was the element that hindered its adoption. Cadwell, O'Brien, and Teixeira (2018) conducted a similar study, comparing the level of MT uptake of in-house and institutional translators, sharing similar results.

Translation in a migration context has also received attention, as multilingual communication is key in crisis scenarios (Piller et al., 2020), and Pérez-Macías, Ramos, and Rico (2020) analysed the perceptions of MT and post-editing of translators in a migration context, which were negative in general terms.

In contrast, Koponen et al. (2020) focused on the audiovisual domain and analysed what 12 professional translators thought about MT and what was their UX after post-editing subtitles. The resulting comments ranged from negative to neutral. These results are in line with other research on audiovisual translation, post-editing and UX, where translators do not view post-editing in subtitling projects favourably (Etchegoyhen et al., 2018; Matusov et al., 2019; Karakanta et al., 2022). In a similar vein, Briva-Iglesias, O'Brien, and Cowan (2023) analysed the MTUX of translators in the legal domain to see what translators preferred from two different post-editing modalities.

However, despite having found research on the perceptions that translators have of MT, the aforementioned studies are exclusively descriptive of participant's perceptions and did not analyse whether these perceptions have any relationship with the quality of the final text or the productivity of translators. This is the gap that this article aims to fill.

In cognitive science, multiple studies show that past experiences and perceptions have a great impact and are a determinant for future beliefs, attitudes and behaviours (Albarracín, 2021; Albarracín and Wyer, 2000). In Translation Studies, de Almeida (2013) suggested that positive perceptions towards MT had an impact on post-editing effort, and Stasimioti and Sosoni (2019) reported that training in MT changed perceptions of MT and post-editing.

Hence, if translators are not happy with their past interactions with MT, and if, before starting a post-editing assignment, they already have a negative opinion about that future interaction (pre-task perceptions), what will the consequences be for the final product (that is, the translation)? Are we in a vicious circle in which translators' negative pre-task perceptions of MT affect the final quality of the translation and/or their productivity?

3 Methodology

The research question we address in this paper is: *Do translators' (positive or negative) pre-task perceptions of MT have any statistically significant relationship with the final translation quality or productivity when doing MTPE tasks?* To answer this question from a novel point of view, we conducted a human-computer interaction-informed study, where we recruited 11 professional translators in the English-Spanish legal translation com-

bination and carried out a pre-task questionnaire to examine their opinions, past experiences and attitudes towards MT. This questionnaire was followed by the translation of 13 texts using an interactive MT workflow. Subsequently, a professional, expert reviewer assessed the quality of the translations after ensuring consistent evaluation criteria with three professional reviewers. We examined the data obtained using different statistical analysis methods to find out whether there was any correlation between the past experiences and attitudes of translators towards MT and their resulting translation quality and productivity. Our hypothesis is that translators with negative pre-task perceptions of MT may produce translations with lower quality than their peers with positive pre-task perceptions because their predisposal to interact with MT will affect their translation processes. The following sub-sections describe the methodology used in-depth.

3.1 Participants

We recruited 11 professional translators in the English-Spanish language combination at an hourly rate of €20. To do this, we posted a job advert on ProZ (one of the most prominent job search platforms in the language services world) and X (which also has a large translator community). By posting on two different platforms and hiring participants on a first-come, first-served basis, we wanted to reach a large number of people without introducing any bias in the selection of participants. Participants were hired as long as they met the three basic conditions for participation: i) be native Spanish translators, ii) have professional experience in legal translation, and iii) have less than 5 years of professional experience. We decided to include the experience limitation because we wanted to minimise bias due to variable levels of experience. In addition, the translators were to perform post-editing, and previous studies suggested that people with more years of experience tended to have more problems interacting with technologies and were more likely to reject their daily use (Alabau et al., 2016).

Translators performed thirteen post-editing sessions of 45 minutes in Lilt (Green, 2016) over ten consecutive days (two weeks). In these sessions, three sessions were conducted through traditional post-editing, and ten sessions through interactive post-editing. The tasks were divided this way for reasons of the project in which the present study is framed, but this has no impact on the data shown here, as all translators worked with the same texts, under the same conditions and had the same amount of time to translate. Translators were instructed to "Perform a full post-editing of the text, with the goal of achieving a perfectly fluent and adequate translation for a client in the legal domain. Any mistranslation may have critical legal consequences for the client, so ensure that you offer a professional translation. There is no problem if you do not finish the whole text in the allocated time".

3.2 Translators' pre-task perceptions

In order to collect translators' pre-task perceptions of MT, we created an online questionnaire to be completed before starting the post-editing task. This included the following questions.

- Experience in MTPE tasks: How long have you engaged with MTPE tasks? Give an approximate time of use with months or years and months (e.g., 1 year and 6 months). [These experiences were then normalized to the number of months].
- Do you like MTPE?: 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 translation projects.
- Do you trust MTPE?: 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 delivery a professional translation project.
- MT as a threat: Please rate how much you agree or disagree with this statement: "Machine Translation is a threat to the sustainability of the translation profession (Score 1 is "Disagree", Score 7 is "Agree").
- Is MTPE boring?: Please rate the following statement: "When I am doing MTPE tasks, I find them [SCORE]". (Score 1 is "Boring", Score 7 is "Engaging").

The responses to the questionnaire were the translators' pre-task perceptions that we correlated with final translation quality and productivity to examine if there was any relationship between them.

3.3 Translation Quality Evaluation

We worked with legal contracts in the English-Spanish combination and controlled the difficulty and length of the texts so that all translators worked with thirteen equally complex texts. For each text, a new task was set, and no translation memory was added. Difficulty was controlled with the Flesch-Kincaid index and the type-token ratio (Graesser et al., 2004). The total number of words translated and evaluated were 120,102. After obtaining the translations, translation quality was evaluated via human evaluation by using 1-4 Adequacy and Fluency scores with a professional, expert evaluator.

Although there are many different methods for evaluating translation quality (Moorkens et al., 2018a; Drugan, 2013), to answer our research question (*Do translators' (positive or negative) pre-task perceptions of MT have any statistically significant relationship with the final translation quality or productivity when doing MTPE tasks?*), we needed to obtain a final score of the translation quality of each translator. We considered the Adequacy and Fluency assessment to be the most appropriate method for our study, as it allowed us to obtain very detailed quality scores for each translator from two different points of views and has been extensively used in MT evaluation (Kocmi et al., 2022; Barrault et al., 2020). We discarded the MQM-based assessment (Freitag et al., 2021) because it focuses on the precise types of errors, and we did not need such a granular translation quality evaluation, plus it increased substantially the translation quality evaluation costs.

Best practices in human evaluation of translation quality recommend using several evaluators to reduce any potential subjectivity (Freitag et al., 2021). As an alternative, we have decided to follow common best practices in the fields of Computer Science and Information Retrieval (Artstein, 2017), also with recognised and widely-accepted methods for reducing evaluator subjectivity, and we have implemented the evaluation only with one expert reviewer after refining the evaluation criteria with a total of three reviewers through two different iterations. The scoring guidelines were updated after each iteration. The three reviewers were recruited by following the same methodology used for recruiting the translators, which can be found in the section 3.1 above, and they had +5 years of professional exprience. The process followed for the quality evaluation was as follows:

First, we created a document explaining in detail the quality evaluation task to be carried out. Detailed scoring guidelines were also designed, in which each possible score (both for Adequacy and Fluency) was described in detail, and two examples were included for each type of score. The aim of these guidelines was to homogenise the evaluation criteria, and thus make the study and the results reproducible and reliable, trying to reduce the personal and subjective bias of each evaluator.

Once the first draft of the scoring guidelines was devised (containing two examples for every type of Adequacy and Fluency mistake), 50 translated segments were sent to the three reviewers. Texts evaluated in the iterations were fragments of English-Spanish legal contracts, similar in content and difficulty to the bulk of translations. The three reviewers annotated the translations and evaluated them according to the criteria of the scoring guidelines.

Subsequently, the Inter-Annotator Agreement (IAA) was calculated using Fleiss' Kappa. IAA can range between 0 and 1 and, generally, an IAA above 0.8 indicates that the consistency between annotations is high (Artstein, 2017). The IAA for our first round of annotations (Iteration 1) was 0.83, indicating that the scoring guidelines were clear, that the annotation consistency of the evaluators was high, but that there was still room for improvement.

Then, a Zoom meeting was held with the 3 reviewers to discuss the discrepancies of annotation in Iteration 1, and the scoring guidelines were updated with additional examples after some discussion. The main changes included re-wording and clarifying the annotation criteria, and more detailed explanations of the annotation limits, with the aim of improving the homogeneity of the annotation and increasing the consistency of the evaluations. Iteration 2 was then prepared, with 50 new segments, to be annotated by following the updated scoring guidelines in the same procedure as in Iteration 1. The IAA of Iteration 2 increased to 0.95, reflecting that the second version of the guidelines was clearer and more concise, and that a consistent evaluation could be obtained when evaluating translations if the guidelines were followed[1].

We then evaluated all the translations (120,102

[1]Link to the final scoring guidelines: https://zenodo.org/records/11091928

words; 13 translations per each of the 11 translators) with a single reviewer, who we considered as the expert reviewer after the first two iterations, the homogenisation of criteria, and the updating of the scoring guidelines. To corroborate that the expert reviewer was still maintaining the annotation criteria halfway through the evaluation of the texts, the other reviewers, who participated in the earlier iterations, performed a cross-check evaluation. For this cross-check evaluation, 250 segments were randomly selected for annotation and the level of consistency was recalculated. The resulting IAA from the cross-check was 0.88, which also indicated a high consistency in the annotation criteria according to the elaborated scoring guidelines.

The end result is an Adequacy and Fluency score for each translator at the segment level. However, after translators performed the post-editing tasks, we observed that, in the allocated time for translation, some translators finished the texts, while others did not. Thus, the Adequacy and Fluency results have been normalised by calculating the average of all the segments translated by each translator. This normalisation has been carried out independently for both Adequacy and Fluency. By doing this, we can compare the results without any bias and independently for Adequacy and Fluency. Thus, we have a global quality score for each translator, ranging from 1 to 4[2].

3.4 Translation Productivity

Translation productivity was tracked in the CAT tool through a word per hour (WPH) measurement. In other words, we collected the translation productivity of every translator in each of the texts in WPH.

3.5 Statistical Analyses

First, we plotted every variable (translators' pre-task perceptions, fluency scores, adequacy scores and productivity measurements) in histograms to see whether the variables were normally distributed, and to strengthen our methodology we also performed the Shapiro-Wilk's test. As data violated the assumptions of normal distribution (p over .05), we conducted a Kendall's T correlation test for all the variables so as to explore the relationships between the measures collected (Mellinger and Hanson, 2016). Due to the number of correlations performed increasing the likelihood of type I error, we recommend interpreting correlations at the .05 level with caution. In addition, it is worth stressing that the strength of the correlation coefficients vary according to the statistical test conducted. Therefore, by following Schober, Boer, and Schwarte (2018) advice, we interpret Kendall T's correlation coefficient strength in the following form: Weak (0.06-0.25), Moderate (0.26 to 0.49), Strong (0.50 to 0.71), and Very strong (0.71 to 1). Below, different heatmaps display the correlation coefficients of every pre-task perception variable in relationship to adequacy, fluency, and productivity. Variables containing an asterisk "*" indicate a statistically significant correlation. Also, the p-values are given for every variable in the wording.

4 Results

This section presents the correlations of translators' pre-task perceptions with fluency, adequacy and productivity.

4.1 Translators' pre-task perceptions of MT and fluency

Figure 1 shows in a heatmap the correlation coefficients resulting from the statistical analysis by considering translators' pre-task perceptions and final fluency scores for each of the texts. By looking at Figure 1, we can see that translators' feeling of boredom or engagement when performing MTPE assignments in a professional environment (r (10) = -.012, p = .85) showed no statistically significantly correlation with Fluency scores.

However, all the other pre-task perceptions variables showed a statistically significant correlation with Fluency. On the one hand, we can observe two variables that show statistically significantly weak correlations. Whether translators had more or less experience in conducting MTPE tasks had no particular relationship with fluency results (r (10) = .12, p = .04). This means that, even if translators were new to interacting with MT in a professional environment, their fluency was not different to those translators with experience in post-editing. In a similar way, translators' attitude towards liking or disliking post-editing tasks (r (10) = -.21, p = .0007) showed a weak statistically significant correlation; this means that we cannot claim a re-

[2]The dataset including the source texts, the translations, the quality scores for fluency and adequacy, as well as the productivity measures can be found in: https://zenodo.org/records/11092027.

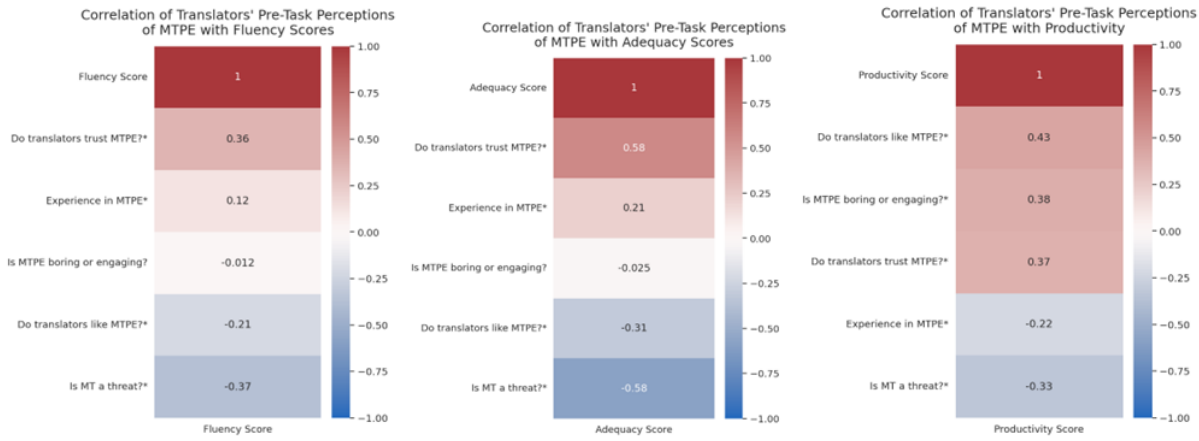

Figure 1: Correlation of translators' pre-task perceptions of MTPE with fluency, adequacy and productivity

lationship between the fact that translators liked post-editing or not with the production of more fluent translations. On the other hand, we can observe two variables that show stronger relationships. Translators' pre-task perceptions of MT being a threat to the translation profession (r (10) = -.37, p = .0001) and the level of trust they had on MTPE (r (10) = .36, p = .0001) showed statistically significant moderate correlations. In other words, this means that translators who had higher levels of trust in MTPE tasks as an aid in their professional translation projects tended to report higher fluency scores. In a similar way, those translators who thought that MT was a threat to their profession tended to produce less fluent translations.

4.2 Translators' pre-task perceptions of MT and adequacy

The correlation heatmap of Figure 1 provides a visual summary of the statistical analysis conducted on translators' pre-task perceptions of MTPE and their translation quality results, specifically focusing on translation adequacy.

One of the most notable results is that we observed a strong statistically significant positive correlation (r (10) = .58, p = 0.0001) between translators' trust in MTPE and the adequacy scores, implying that higher trust in the system is linked to higher performance levels in producing adequate translations. The other remarkable result is that translators' view of MT as a threat yielded a strong statistically significant negative correlation (r (10) = -.58, p = 0.0001), suggesting that apprehensions about the technology's impact on the profession may undermine translation adequacy.

Factors such as the enjoyment of conducting MTPE tasks (r (10) = -.31, p = 0.0001), and the overall experience in MTPE (r (10) = .21, p = 0.0004), showed less pronounced yet statistically significant correlations, indicating that these perceptions might not be as critical in influencing the adequacy of translation outcomes.

These insights contribute to the ongoing discourse on the human factors influencing contemporary translator-computer interactions, underscoring the complex interplay between subjective perceptions and objective translation performance metrics. These results indicate that translators' lack of trust in MTPE tasks and the consideration of MT as a threat to their profession may have a strong effect on translation quality, especially adequacy, even before the task has already started.

4.3 Translators' pre-task perceptions of MT and productivity

In terms of productivity, Figure 1 provides a quantitative depiction of the correlations between translators' pre-task perceptions of MTPE and their measured productivity in WPH. Here, the results from every pre-task perception variable were statistically significant.

The data indicates that positive perceptions towards MT, such as liking MTPE tasks as a professional aid (r (10) = .43, p = 0.0001), finding them engaging (r (10) = .38, p = 0.0001), or the level of trust in MT (r (10) = .37, p = 0.0001) are moderately correlated with higher productivity scores. Conversely, the negative moderate correlation with the perception of MT as a professional threat (r (10) = -.33, p = 0.0001), although displaying a weaker association, highlights potential areas of concern. These findings may reflect a complex-

ity in MTPE's perceived impact on the translation industry, which could influence translator productivity.

There is also a weak statistically significant negative correlation with MTPE experience (r (10) = -.22, p = 0.0001).

5 Discussion of the results

After analysing the results, we can see that translators' pre-task perceptions of MT have a higher correlation with adequacy scores than with fluency and productivity scores. Research by Castilho et al. (2017) suggested that NMT systems produce very fluent translations and, therefore, the effect of translators' pre-task perceptions may not be that impactful on fluency scores if the system already offers a fluent MT output. In terms of adequacy scores, however, translators' pre-task perceptions have a bigger impact. These results indicate that, in post-editing tasks with NMT systems that offer good MT quality (English-Spanish in our case), adequacy scores have a higher dependence on the translator, while fluency scores have a lower dependence on translators' translation and/or post-editing skills because the MT system already offers higher quality MT output. The expert reviewer assessed the MT output of the MT system used in this study, which obtained a global score of 3.48/4 in terms of adequacy and 3.71/4 in terms of fluency, showcasing good quality results. It may be interesting to further validate this idea by replicating this study in a different language combination, particularly in a case in which NMT systems offer lower MT quality (i.e. a low-resource language combination). In terms of productivity, in general terms, we can see that translators who had positive pre-task perceptions of MT tended to translate faster when conducting post-editing tasks than their peers with negative pre-task perceptions of MT.

Experience in MTPE tasks has no correlation with final translation quality or productivity in the data analysed. These results indicate that it is the translator who matters. Results suggest that a professional translator with good translation skills will offer good quality translations and will work equally faster when interacting with MT, regardless of their experience in providing MTPE language solutions.

It is worth stressing that the most notable correlation coefficients were observed in two specific variables: translators' level of trust in MTPE and the perception of MT being a threat to the translation profession. The level of trust translators have in MT shows a strong correlation with adequacy and a moderate correlation with fluency and productivity. This is interesting, as translators who trust MT systems to help them work in their professional, daily tasks offered higher final translation quality than those who did not trust MT systems. This is in line with previous research in cognitive science (Albarracín, 2021), which indicated that prior negative perceptions are an important and crucial determinant for future attitudes and behaviours. In our case, translators' pre-task perceptions of MTPE tasks had a strong negative correlation with the quality of the final product, that is, the translation. This may be because translators who do not trust MT do not enjoy this interaction or do not give their best when interacting with MT in their regular, professional workflows. This also applies to productivity: translators with higher levels of trust on MT translated faster, probably because they were more enthusiastic about engaging with MT. Those translators who did not trust MT were probably more reluctant to engage with MT in the best of their abilities.

This backs up the results of the second pre-task perception variable with a strong negative correlation in our study, that is, whether translators consider MT as a threat to the translation profession. The perception of MT being a threat to translators showed a strong association with final translation quality, both in terms of Adequacy and Fluency, and the results showed statistical significance. The correlation of this pre-task perception variable with productivity is weaker, but still moderate and statistically significant. What these correlations mean is that translators who, even before starting a post-editing task, think that MT is a threat and harmful for the translation profession are more likely to produce lower translation quality and to translate slower.

The study's findings on the relationship between translators' pre-task perceptions and translation quality and productivity have profound implications, offering novel insights into the dynamics of modern translator-computer interactions. It is evident that the approach translators adopt towards a task plays a critical role in determining the final outcome, with varying degrees of influence on different aspects of translation performance (in terms

of quality or productivity). These results highlight the problem of modern translator-computer interactions, and suggest that we should give higher attention to the improvement of these interactions, probably by looking at the MTUX (Koponen et al., 2020; Karakanta et al., 2022; Briva-Iglesias and O'Brien, 2023), putting the human in the centre of modern translator-computer interactions (Shneiderman, 2022), reducing recent complaints about dehumanisation (Moorkens, 2020). Also, this has great implications for the training of translators, as it highlights the importance of MT literacy from different points of view (Bowker and Ciro, 2019), which are further detailed in the conclusions.

6 Conclusions

Since the emergence of language technologies, the perceptions of those who interact with them have been studied, ranging from professional translators to gisting users (Nurminen, 2019). Particular attention has been paid to the perceptions of professional translators, who saw their traditional workflows disrupted by these new technologies. However, most studies to date have been descriptive and did not take into account the relationship between perceptions and the quality of the translation or the productivity of the translator. This article aims to fill this gap in the literature with a longitudinal study of 11 professional English-Spanish translators in the legal domain, to explore whether negative or positive translators' pre-task perceptions of MT have any relationship with the final translation quality and productivity in a post-editing workflow. In terms of limitations, it would have been better to increase the sample size of the translations or the number of translators. However, we hired 11 professional translators who produced a total of 120,102 words over 10 consecutive days. It would have also been ideal to have the three reviewers assess all the translations, but due to budget constraints, we reduced the reviewer bias through different evaluation iterations, the measurement of IAA, and the refinement of a set of quality scoring guidelines. Exploring additional domains to the legal field would have also been positive to assess whether these results are generalisable to other translation specialisations.

As a conclusion, the results suggest that translators with negative pre-task perceptions of MT tend to deliver poorer quality translations, as well as to translate slower, than their peers with positive pre-task perceptions of MT. Specially, in our study, we observed that translators who thought that MT was a threat to their profession or distrusted MT as an aid in their work obtained lower quality and productivity scores. These were the two variables with the highest correlation coefficients. By contrast, translators with positive pre-task perceptions obtained better translation quality and productivity scores. This may have happened because translators with negative pre-task perceptions of MT may have not interacted with MT adequately or with an open-minded point of view, impacting their final translation quality and productivity.

This research opens up new questions: do these results suggest that there is a direct relationship between pre-task perceptions of translation technologies and final quality and productivity results? Would translators with negative pre-task perceptions of MT obtain better quality results if they translated without MT? And what would happen if we trained these translators and taught them to see MT as an aid to augment their skills and help them in their professional tasks?

Although we now have new questions to answer, what is clear is that the level of trust in MT and the conception of MT as a threat to the translation profession have a strong correlation with final quality results, especially Adequacy, and a moderate correlation with productivity. These correlations have important implications. Translators fearing MT, or those who are more reluctant to trust their interactions with MT, may not be leveraging the advantages and benefits MT offers them. Therefore, the results call for multiple actions to be taken in order to:

- Increase translators' confidence in their interactions with MT as a tool that can be useful and support them in professional projects, always bearing in mind that translators are the ones controlling the interaction, and that MT functions as a support that can offer alternative terminology solutions or facilitate understanding of the source text, among other forms of assistance. The main goal of technologies should be to augment translators and reduce their human cognitive limitations (O'Brien, 2023; Raisamo et al., 2019; Alicea, 2018; Shneiderman, 2022), pursuing human-centered, augmented machine translation (Briva-Iglesias, 2024), not to replace and substitute them.

- Present MT as a tool that can facilitate translators' work, either to increase productivity or to open doors to new professional markets and domains, as is the case of *language engineers* (Briva-Iglesias and O'Brien, 2022). As Stasimioti and Sosoni (2019) reported, training translators on MT will change their perceptions of MT because they will learn what MT allows them to do or not. Translators' technological and MT literacy is now more important than ever in the AI age (Bowker and Ciro, 2019).

These two elements would increase the adoption and use of MT as assistance, as well as reduce the negativity of translators' pre-task perceptions of MT. However, it is vital to stress that this MT literacy must be accompanied by a broad and holistic view of MT, including its limitations, so that translators acquire a critical view of when it is appropriate and when not to use MT, as it may also involve important ethical issues (Moorkens, 2022).

Acknowledgements

This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d real) under Grant No. 18/CRT/6224.

References

Alabau, Vicent, Michael Carl, Mercedes García-Martínez, and Jesús González-Rubio. 2016. *Learning Advanced Post-Editing*.

Albarracín, Dolores and Robert S. Wyer. 2000. The Cognitive Impact of Past Behavior: Influences on Beliefs, Attitudes, and Future Behavioral Decisions. *Journal of personality and social psychology*, 79(1):5–22, July.

Albarracín, Dolores, editor. 2021. *The Impact of Past Experience and Past Behavior on Attitudes and Behavior*. Cambridge University Press.

Alicea, Bradly. 2018. An Integrative Introduction to Human Augmentation Science.

Artstein, Ron. 2017. *Inter-Annotator Agreement*. Springer Netherlands.

Barrault, Loïc, Magdalena Biesialska, Ondřej Bojar, Marta R. Costa-jussà, Christian Federmann, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Matthias Huck, Eric Joanis, Tom Kocmi, Philipp Koehn, Chi-kiu Lo, Nikola Ljubešić, Christof Monz, Makoto Morishita, Masaaki Nagata, Toshiaki Nakazawa, Santanu Pal, Matt Post, and Marcos Zampieri. 2020. Findings of the 2020 Conference on Machine Translation (WMT20). In *Proceedings of the Fifth Conference on Machine Translation*, pages 1–55. Association for Computational Linguistics.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623. Association for Computing Machinery.

Bowker, Lynne and Jairo Buitrago Ciro. 2019. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Emerald Publishing Limited.

Briva-Iglesias, Vicent and Sharon O'Brien. 2022. The Language Engineer: A Transversal, Emerging Role for the Automation Age. *Quaderns de Filologia - Estudis Lingüístics*, 27:17–48.

Briva-Iglesias, Vicent and Sharon O'Brien. 2023. Measuring Machine Translation User Experience: A Comparison between AttrakDiff and User Experience Questionnaire. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 335–344.

Briva-Iglesias, Vicent, 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).

Briva-Iglesias, Vicent. 2024. *Fostering human-centered, augmented machine translation: analysing interactive post-editing*. PhD thesis. Dublin City University.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners.

Cadwell, Patrick, Sharon O'Brien, and Carlos S. C. Teixeira. 2018. Resistance and accommodation: Factors for the (non-) adoption of machine translation among professional translators. *Perspectives*, 26(3):301–321.

Castilho, Sheila, Joss Moorkens, Federico Gaspari, Iacer Calixto, John Tinsley, and Andy Way. 2017.

Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics*, 108(1):109–120.

de Almeida, Giselle. 2013. *Translating the Post-Editor: An Investigation of Post-Editing Changes and Correlations with Professional Experience across Two Romance Languages*. Doctoral, Dublin City University.

Drugan, Joanna. 2013. *Quality in Professional Translation: Assessment and Improvement*, volume 9. A&C Black.

ELIS Research. 2023. EUROPEAN LANGUAGE INDUSTRY SURVEY 2023.

Etchegoyhen, Thierry, Anna Fernández Torné, Andoni Azpeitia, Eva Martínez Garcia, and Anna Matamala. 2018. Evaluating Domain Adaptation for Machine Translation Across Scenarios. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA).

Fırat, Gökhan. 2021. Uberization of translation: Impacts on working conditions. *The Journal of Internationalization and Localization*, 8(1):48–75.

Freitag, Markus, George Foster, David Grangier, Viresh Ratnakar, Qijun Tan, and Wolfgang Macherey. 2021. Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation. *arXiv:2104.14478 [cs]*, April.

Gaspari, Federico, Antonio Toral, Sudip Kumar Naskar, Declan Groves, and Andy Way. 2014. Perception vs reality: Measuring machine translation post-editing productivity. In *Third Workshop on Post-Editing Technology and Practice*, volume 60.

Graesser, Arthur C., Danielle S. McNamara, Max M. Louwerse, and Zhiqiang Cai. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2):193–202.

Green, Spence. 2016. Interactive Machine Translation. In *Conferences of the Association for Machine Translation in the Americas*, page 93.

Guerberof Arenas, Ana, Joss Moorkens, and Sharon O'Brien. 2021. The impact of translation modality on user experience: An eye-tracking study of the Microsoft Word user interface. *Machine Translation*, 35(2):205–237.

Karakanta, Alina, Luisa Bentivogli, Mauro Cettolo, Matteo Negri, and Marco Turchi. 2022. Post-editing in Automatic Subtitling: A Subtitlers' perspective. In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pages 261–270, Ghent, Belgium. European Association for Machine Translation.

Kocmi, Tom, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Thamme Gowda, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Rebecca Knowles, Philipp Koehn, Christof Monz, Makoto Morishita, Masaaki Nagata, Toshiaki Nakazawa, Michal Novák, Martin Popel, and Maja Popović. 2022. Findings of the 2022 Conference on Machine Translation (WMT22). In Koehn, Philipp, Loïc Barrault, Ondřej Bojar, Fethi Bougares, Rajen Chatterjee, Marta R. Costa-jussà, Christian Federmann, Mark Fishel, Alexander Fraser, Markus Freitag, Yvette Graham, Roman Grundkiewicz, Paco Guzman, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Tom Kocmi, André Martins, Makoto Morishita, Christof Monz, Masaaki Nagata, Toshiaki Nakazawa, Matteo Negri, Aurélie Névéol, Mariana Neves, Martin Popel, Marco Turchi, and Marcos Zampieri, editors, *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 1–45. Association for Computational Linguistics.

Koponen, Maarit, Umut Sulubacak, Kaisa Vitikainen, and Jörg Tiedemann. 2020. MT for Subtitling: Investigating professional translators' user experience and feedback. In Ortega, John E., Marcello Federico, Constantin Orasan, and Maja Popovic, editors, *Proceedings of 1st Workshop on Post-Editing in Modern-Day Translation*, pages 79–92. Association for Machine Translation in the Americas.

Matusov, Evgeny, Patrick Wilken, and Yota Georgakopoulou. 2019. Customizing Neural Machine Translation for Subtitling. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 82–93. Association for Computational Linguistics.

Mellinger, Christopher and Thomas Hanson. 2016. *Quantitative Research Methods in Translation and Interpreting Studies*. Routledge.

Moorkens, Joss, Sheila Castilho, Federico Gaspari, and Stephen Doherty, editors. 2018a. *Translation Quality Assessment: From Principles to Practice*, volume 1 of *Machine Translation: Technologies and Applications*. Springer International Publishing.

Moorkens, Joss, Antonio Toral, Sheila Castilho, and Andy Way. 2018b. Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, 7(2):240–262.

Moorkens, Joss. 2020. "A tiny cog in a large machine": Digital Taylorism in the translation industry. *Translation Spaces*, 9(1):12–34.

Moorkens, Joss. 2022. Ethics and machine translation. *Machine translation for everyone*, pages 121–140.

Nurminen, Mary. 2019. Decision-making, Risk, and Gist Machine Translation in the Work of Patent Professionals. In *Proceedings of the 8th Workshop on Patent and Scientific Literature Translation*, pages 32–42. European Association for Machine Translation.

O'Brien, Sharon, Maureen Ehrensberger-Dow, Marcel Hasler, and Megan Connolly. 2017. Irritating CAT tool features that matter to translators. *Hermes: Journal of Language and Communication in Business*, 56:145–162.

O'Brien, Sharon. 2023. Human-Centered augmented translation: Against antagonistic dualisms. *Perspectives*, pages 1–16.

Olohan, Maeve. 2011. Translators and translation technology: The dance of agency. *Translation Studies*, 4(3):342–357.

Pérez-Macías, Lorena, María del Mar Sánchez Ramos, and Celia Rico. 2020. Study on the Usefulness of Machine Translation in the Migratory Context: Analysis of Translators' Perceptions. *Open Linguistics*, 6(1):68–76.

Piller, Ingrid, Jie Zhang, and Jia Li. 2020. Linguistic diversity in a time of crisis: Language challenges of the COVID-19 pandemic. *Multilingua*, 39(5):503–515.

Raisamo, Roope, Ismo Rakkolainen, Päivi Majaranta, Katri Salminen, Jussi Rantala, and Ahmed Farooq. 2019. Human augmentation: Past, present and future. *International Journal of Human-Computer Studies*, 131:131–143.

Rossi, Caroline and Alice Carré. 2022. How to choose a suitable NMT solution?: Evaluation of MT quality.

Rossi, Caroline and Jean-Pierre Chevrot. 2019. Uses and perceptions of Machine Translation at the European Commission. *The Journal of specialised translation (JoSTrans)*.

Schober, Patrick, Christa Boer, and Lothar A. Schwarte. 2018. Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, 126(5):1763.

Shneiderman, Ben. 2022. *Human-Centered AI*. Oxford University Press.

Stasimioti, Maria and Vilelmini Sosoni. 2019. Undergraduate Translation Students' Performance and Attitude vis-à-vis Machine Translation and Post-editing: Does Training Play a Role. In *41st Translating and the Computer Conference*, pages 125–136.