

Methodological considerations for emotion research in MTPE: A multimethod design using eye-tracking and self-report questionnaires

María Inmaculada Vicente López

1. Introduction

The integration of emotional dimensions into research on machine translation post-editing (MTPE) constitutes an important step toward a more comprehensive understanding of the cognitive-affective processes underlying technologically mediated translation tasks. However, several methodological challenges persist, particularly in the operationalization of emotional constructs, which are difficult to define and measure with precision in human-computer interaction contexts. This complexity is further compounded by participant variability, as emotional responses differ widely across individuals and tasks. Moreover, efforts to preserve ecological validity by studying MTPE in realistic settings introduce additional constraints, limiting experimental control and making it more difficult to isolate emotion-specific responses due to overlapping physiological signals.

To address these challenges, recent studies in CTIS have increasingly adopted multimethod research designs that combine psychometric, physiological, and behavioral data to investigate emotional responses during translation (Cifuentes Férez et al. 2024; Rojo López et al. 2021a; Weng et

al. 2022) and interpreting (Chiang 2010; Korpál 2017; Moser-Mercer 2005; Pinazo Calatayud & Jiménez Ivars 2002; Rojo López et al. 2021b).

Within this multimethod landscape, eye-tracking has emerged as a well-established, non-invasive tool in CTIS, particularly for investigating cognitive effort and visual attention in MTPE (Daems 2016; Koglin 2015; Nitzke 2019; Moorkens et al. 2015; Teixeira 2014). Yet despite its widespread use, its application to emotion research remains largely underexplored in translation studies, except for Weng et al. (2022).

In the following sections, we reflect on the methodological challenges, decisions, and insights derived from a study investigating anxiety, emotional arousal, attitude, visual attention, and cognitive effort in MTPE using a multimethod research design, combining eye-tracking, psychometric instruments, and behavioral data. While grounded in a specific study (Vicente López 2024), the discussion aims to provide practical considerations for researchers seeking to implement ecologically valid, multimethod approaches to investigate affective dimensions of MTPE. These reflections contribute to ongoing methodological debates in CTIS, with particular emphasis on the integration of physiological and psychometric data in experimental research.

2. Methodological challenges of emotion research in MTPE using eye-tracking

Emotion research in MTPE entails a range of conceptual and methodological challenges that are also pertinent to the broader field of CTIS, where emotions—despite gaining increasing recognition—remain insufficiently integrated into experimental designs (Rojo López & Cifuentes Férez 2021).

A central conceptual challenge in researching emotions in CTIS is the tension between the assumed universality of emotions and their variability across individuals and contexts (Rojo López & Cifuentes Férez 2021). In the case of MTPE, emotional responses are shaped by a complex interplay of factors, including machine translation (MT) output quality, domain familiarity, translation tool usability, post-editing pricing, client expectations, professional experience, and the translator's attitude toward technology. Several of these variables are often linked to “cognitive friction” (Cooper 2004), an

additional load not inherent to the task but caused by translation tool features that distract or hinder task performance (O'Brien & Ehrensberger-Dow 2020). Beyond tool-specific issues, negative emotional responses to MTPE often stem from broader concerns about automation and industry practices, including socio-economic pressures and the dehumanizing effects of its implementation (Nunes Vieira 2018).

Individual variability further complicates emotion research in MTPE. Personality traits such as self-esteem or trait anxiety can further shape how translators react emotionally to the task (Rojo López et al. 2021a; Cifuentes Férrez et al. 2024). Extant research in CTIS suggests that professional experience may buffer stress and support more effective emotional regulation under pressure (Hvelplund 2016; Rojo López & Ramos Caro 2018). However, the mechanisms underlying these coping strategies—and their regulatory function—remain insufficiently understood. Even under tightly controlled experimental conditions, translators approach tasks with differing mental models, strategies, and emotional dispositions. Instead of being viewed as a methodological flaw, such variability should be recognized as a defining feature of translation as a human activity (Jääskeläinen 2010; Pym 2013) and as central to understanding translation as a situated and person-dependent process (Bolaños Medina 2014). Accounting for individual differences can help minimize the impact of uncontrolled variables, improve internal validity, and reduce the risk of overgeneralizing from small or homogeneous samples (O'Brien 2009).

Emotions are not directly observable, and their investigation in the context of MTPE requires a multimethod methodological approach capable of capturing both subjective perceptions and externally observable manifestations of those emotions (Rojo López & Cifuentes Férrez 2021). As Muñoz and Rojo (2025) argue, combining behavioral and physiological techniques (e.g., eye-tracking, skin conductance) with self-report methods (e.g., cued retrospection, self-report scales) enhances methodological validity by providing converging lines of evidence about emotional experience and its cognitive impact.

Eye-tracking provides real-time, high-resolution data on eye movements and pupil dilation, which can serve as indicators of cognitive effort, visual attention, and behavioral patterns (Hvelplund 2014), as well as emo-

tional arousal (Holmqvist et al. 2011). In MTPE research, it has traditionally been used to assess cognitive effort through metrics such as fixation duration, fixation count, gaze time, and pupil dilation (Daems 2016; Koglin 2015; Nitzke 2019; Moorkens et al. 2015). However, its potential to investigate emotional dimensions in MTPE remains largely underexplored. In particular, pupil dilation—a well-established marker of emotional arousal—has received limited attention in translation research (Weng & Zheng 2020; Weng et al. 2022). Similarly, patterns of visual attention, including fixations on the source text, MT output, or online search, may reflect not only cognitive processing but also affective responses (Lehr & Hvelplund 2020).

The use of eye-tracking technology to investigate translators' emotions in MTPE—understood as a form of human-computer interaction—presents several technical and methodological challenges.

Among the technical challenges, researchers must manage the high volume of data generated, deal with participant recruitment and post-hoc exclusion, and address threats to ecological validity (O'Brien 2009). A persistent difficulty in MTPE research lies in balancing ecological validity with experimental control. Although screen-based eye-trackers are non-invasive, they require stable posture, minimal head movement, and controlled lighting conditions (Hvelplund 2014), which may constrain translators' natural behavior or induce low-level stress that interferes with the emotions under investigation (Rojo López & Cifuentes Férez 2021). As O'Brien (2009) points out, removing translators from their familiar working environments—where factors such as monitor size, software configuration, or keyboard layout may influence performance—further undermines ecological validity. Conversely, naturalistic studies offer greater ecological validity but do not allow for the establishment of cause-and-effect relationships (Rojo López & Muñoz Martín 2022). In response to these challenges, CTIS scholars increasingly advocate for hybrid designs that embed experimental tasks within more realistic settings, such as MTPE carried out in commercial computer-assisted translation (CAT) environments (Ragni & Nunes Vieira 2022) or simulating an office working environment (O'Brien 2009).

Among the methodological challenges is the frequent misalignment between physiological indicators and participants' self-reports of emotional experience. For example, eye-tracking may fail to reveal signs of emotional

arousal when post-editing low-quality MT output or may show no distinctive patterns of visual attention distribution, even though participants report heightened levels of anxiety. Weng et al. (2022) observed a similar dissociation in their study on time pressure in translation: while self-reported stress and anxiety levels rose significantly under stringent time conditions, some physiological indicators—such as galvanic skin response and heart rate variability—followed a different, non-linear pattern or failed to reach statistical significance. As Rojo López and Cifuentes Férez (2021, 10) note, “whereas statistically significant effects are reported for self-report or subjective measures, physiological data often fail to reach significance or display unexpected patterns.”

Finally, the interpretation of eye-tracking metrics concerning emotional processes in translation remains both complex and underexplored. Some eye-tracking metrics—such as pupil dilation, fixation duration, and blink rate—are considered indicators of emotional arousal (Holmqvist et al. 2011; Gieshoff et al. 2021), but they remain underexplored in CTIS, and their role in capturing affective responses is yet to be systematically investigated. Moreover, some of these metrics, particularly pupil dilation, have been interpreted both as an indicator of cognitive effort and emotional arousal. For instance, while some scholars associate larger pupil sizes with heightened stress or emotional arousal (Weng et al. 2022), others interpret the same physiological response as an indicator of cognitive effort during translation tasks (Hvelplund 2011; 2016; 2017).

This conceptual ambiguity complicates interpretation and raises questions about the construct validity of physiological indicators, which may reflect overlapping cognitive and emotional processes that are difficult to disentangle. As such, eye-tracking data should not be treated as direct or exclusive evidence of emotional arousal. These limitations underscore the need for multimethod research designs that triangulate eye-tracking metrics with complementary data sources, including physiological measures (e.g., heart rate, skin conductance), introspective instruments (e.g., psychometric scales, cued retrospection), and behavioral indicators (e.g., visual attention patterns, task time, or keystroke activity), to more reliably infer emotional engagement in MTPE (Rojo López & Ramos Caro 2022).

3. The study

This section outlines the methodological framework of our study, which employed a multimethod research design combining psychometric, physiological, and behavioral data. The primary objective was to examine differences in translators' anxiety, emotional arousal, attitude, and cognitive effort during an English-to-Spanish MTPE task, using two types of MT technologies: statistical machine translation (SMT) and neural machine translation (NMT). The study also aimed to explore the potential of pupil dilation as a physiological indicator of emotional arousal and to assess the influence of MT technology, anxiety, and attitude on visual attention patterns among student and professional translators.

Emotional arousal was measured via pupil dilation, while anxiety was assessed using the State-Trait Anxiety Inventory (STAI) (Seisdedos, 1998; Spielberger et al. 1983), which included trait anxiety (STAI-T) and state anxiety measured before (PRE-STAI-S) and after the task (POST-STAI-S). Cognitive effort was measured through fixation durations, and visual attention distribution was analyzed through fixation counts on the source text, MT output, and Internet browser. These aims were grounded in the assumption that the lower quality of SMT output may increase cognitive effort and elicit more negative emotional responses, potentially resulting in greater pupil dilation, higher post-task anxiety, and distinct visual attention patterns.

3.1 Participants

The participant pool comprised 38 native Spanish speakers divided into two groups: 18 professional translators and 20 final-year undergraduate students of Translation and Interpreting at the University of Murcia. This sample was obtained after excluding participants whose eye-tracking data failed to meet quality criteria (see Section 3.4).

Professional translators (5 men, 13 women) ranged in age from 27 to 62 (median = 41) and had between 4 and 38 years of professional experience (median = 16.5). All had experience with MTPE, ranging from 1 to 10 years (median = 7). Recruitment was carried out through Spanish professional associations (ASETRAD, TREMÉDICA, ACE TRADUCTORES) and professional networking platforms (LinkedIn, ProZ). Participation was compensated with a digital voucher.

The student group (4 men, 16 women) was more homogeneous in age (20–21 years; median = 21) and had no prior MTPE experience. However, all students completed an MTPE workshop integrated into a translation technologies course, which provided them with foundational theoretical and practical knowledge prior to eye-tracking sessions.

A significant challenge in MTPE research lies in the heterogeneity of the sample population, as studies often recruit translators with varying levels of expertise and diverse educational and professional backgrounds (Nunes Vieira 2015). In some cases, information about participants' prior experience with MTPE is either omitted or derived from general self-reports (Carl & Buch-Kromann 2010; Nitzke 2019), making it difficult to assess the role of expertise with any degree of precision.

It is important to acknowledge that MTPE is rarely the sole activity undertaken by professional translators; rather, it is typically integrated into broader workflows that include tasks such as translation, revision, localization, subtitling, or project management. This diversity makes it difficult to isolate MTPE-specific experience with precision, as its extent and frequency can vary considerably from one translator to another. While more accurate indicators, such as the number of years or frequency of engagement with MTPE tasks, would be desirable, such data are often unavailable or difficult to elicit from participants. These challenges reflect long-standing concerns about how to define and operationalize professional expertise in translation process research (O'Brien 2009).

In our study, establishing a reliable measure of MTPE experience to recruit professional translators with comparable backgrounds proved challenging, especially due to the diversity of their professional trajectories.

Initially, we considered detailed measures—such as the total volume of MTPE work (e.g., word counts) or the proportion of MTPE within the overall workload. However, participants, especially those with extensive experience, found it difficult to provide accurate estimates. These figures vary greatly depending on the client, project type, and area of specialization, and the regularity of MTPE assignments is not necessarily consistent over time. Given the rapid evolution of language technology and the rise of MTPE within the translation industry (Nimdzi 2025), it is also reasonable to assume that the frequency of MTPE tasks has increased in recent years. This

variability further complicates efforts to establish a stable and meaningful measure of translation experience.

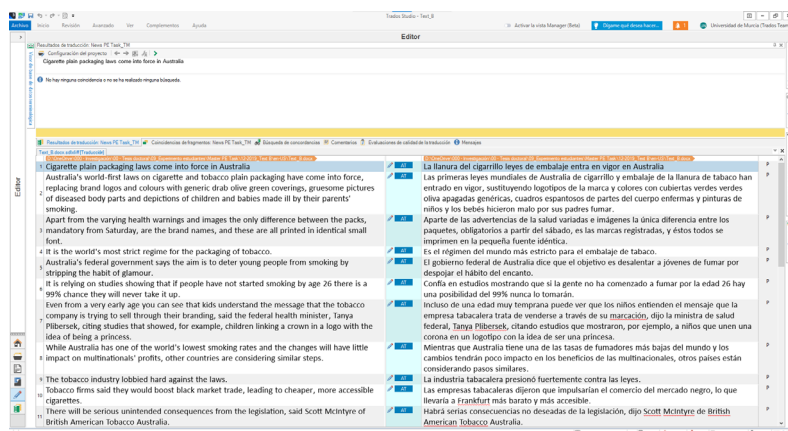
In line with previous research (Moorkens et al. 2015; Stasimioti & Sosoni 2021), we opted to quantify MTPE experience in terms of years. Although it would have been ideal to select participants with similar levels of MTPE experience, doing so would have significantly reduced the sample size, given the well-documented difficulties of recruiting professional translators for empirical research (O'Brien 2009).

3.2 Materials and tools

The experimental task consisted of a full MTPE from English to Spanish of a newspaper article carried out in SDL Trados Studio 2017¹. This CAT tool was chosen for its widespread use in the translation industry, offering an ecologically valid approximation of professional working conditions. Its use aligns with recent calls for research on cognitive effort during MTPE in commercial translation environments, where empirical studies remain scarce (Ragni & Nunes Vieira 2022).

The MTPE task was configured as a dedicated project within Trados Studio, using a document pre-translated with MT output and an empty translation memory (Figure 1).

Figure 1. MT output preloaded in Trados Studio at the start of the task



1 <https://www.trados.com/>

3.2.1. Source text and MT engines

The experimental task involved the full MTPE of a 191-word newspaper article originally published in *The Guardian*, reporting on legislation concerning plain tobacco packaging in Australia. The source text was extracted from the WMT13 corpus, used in the Eighth Workshop on Statistical Machine Translation—an MT evaluation workshop held in 2013 as part of the Annual Meeting of the Association for Computational Linguistics (ACL)².

The decision to use a journalistic text was based on two main considerations. First, journalistic content typically covers general-interest topics, which helps minimize potential bias—particularly among professional translators with specialized domain expertise. As Saldanha and O’Brien (2013) note, recruiting professional translators for process-oriented research can be challenging. Selecting a genre with broad appeal thus increased the likelihood of attracting a more diverse and representative participant sample. Second, the use of journalistic texts aligns with previous research in MTPE (Daems 2016; Koponen et al. 2012; Nunes Vieira 2015), supporting comparability across studies and contributing to methodological continuity.

To examine the impact of MT system type on emotional arousal, anxiety, cognitive effort, and visual attention distribution, all participants—students and professionals—were randomly assigned to one of four groups. One student group (10 participants) and one professional group (9 participants) post-edited a 220-word output generated by DeepL³, an NMT system. The remaining two groups (10 students and 9 professionals) post-edited a 214-word output produced by PROMT⁴, which operated as an SMT system at the time of the study (see Appendices A and B). Both MT systems were generic engines, trained on large multilingual corpora spanning a range of domains, in line with (Stasimioti et al. 2020).

Although an in-depth manual evaluation of MT quality was beyond the scope of this study, we conducted two automated MT quality assessments to confirm the anticipated quality differences between both MT sys-

² <https://www.statmt.org/wmt13>

³ <https://www.deepl.com/translator>

⁴ <https://www.online-translator.com>

tems. The first evaluation employed METEOR metric (Lavie & Denkowski 2009), which evaluates the similarity between MT output and human reference translations. Scores range from 0 (no similarity) to 1 (perfect match). Unlike other metrics such as BLEU, METEOR accounts for morphological variation, synonyms, and paraphrasing, and includes penalties for grammatical inconsistencies—features that make it more strongly correlated with human judgment (Nunes Vieira 2015). As shown in Table 1, DeepL achieved higher METEOR scores than PROMT.

Table 1. **METEOR scores**

System	METEOR	
	M	SD
DeepL	0,71	0,16
PROMT	0,63	0,13

To complement this analysis, we employed a neural evaluation metric—BLEURT—via the MATEO⁵ platform (MACHINE Translation Evaluation Online), developed by Ghent University (Vanroy et al. 2023). This web-based tool supports evaluation across various metrics, including both word-based and character-based approaches. Table 2 shows the corpus-level BLEURT scores and confidence intervals obtained through bootstrap resampling (n = 1000). As anticipated, DeepL outperformed PROMT by a statistically significant margin.

Table 2. **BLEURT scores**

System	bleurt ($\mu \pm 95\% \text{ CI}$)
Reference: DeepL	73,1 (73,0 $\pm$ 7,0)
PROMT	52,7 (52,7 $\pm$ 11,3) ($p = 0,0010$)*

Statistical significance set at $p < 0.05$

These results, however, should be interpreted with caution, as MT system performance is not static but depends on several factors, including the

5 <https://mateo.ivdnt.org/>

syntactic complexity of the source text, sentence length, and the presence of prepositional structures. Although this study did not explicitly control for these variables, prior research has shown that such linguistic features can substantially affect MT output quality (Aziz et al. 2012; Koponen et al. 2011; Nunes Vieira 2015; Tatsumi & Roturier 2010). These considerations are crucial when using automated metrics to compare MT systems, as variations in source text structure can cause quality fluctuations unrelated to the system itself. Uncontrolled variation in source text features can compromise the validity of system comparisons and obscure findings related to translators' MTPE behavior. This underscores the importance of transparent and systematic source text selection, particularly when automatic metrics are used as proxies for human judgment.

3.2.2. Eye-tracker

Participants' eye movements and pupil dilation were recorded during the MTPE task using a Tobii T120 eye-tracker (Tobii© Technology 2010), a high-precision, remote device integrated into a 17-inch monitor. With a sampling rate of 120 Hz—equivalent to 120 data points per second per eye, or one data point every 8.3 milliseconds—the system provided high temporal resolution, enabling the generation of a dense and detailed dataset throughout the task. This setup allowed for the collection of fixation duration, pupil size, and fixation counts, facilitating in-depth analysis across participant groups and MT technologies.

Although the device offers high spatial accuracy (0.5 degrees), additional measures were taken to optimize data quality. To minimize tracking errors and enable more precise fixation analysis, the font size in the Trados Studio editor was increased to 18 pt. This adjustment, following the recommendations of O'Brien (2009), supports the reliable interpretation of gaze data at both the word and sentence level—although such fine-grained analysis was beyond the scope of the present study. Gaze data were collected, processed, and exported using Tobii Studio v.3.3.2, a software platform specifically developed for Tobii eye-trackers, which provides a robust environment for managing and analyzing eye movement recordings.

Data collection with professional translators was conducted at a language service provider (LSP), while student sessions took place in a meet-

ing room at the University of Murcia (see Section 3.4). In both settings, eye-tracking data were collected under standardized conditions, including consistent artificial lighting and careful calibration of the device's distance and height.

The main challenge arose from the COVID-19 pandemic, which restricted in-person contact and limited access to the LSP's premises, resulting in participant attrition and the interruption of data collection. To resume the study, the experimental setting was relocated to the home offices of participating professional translators. Special effort was made to preserve the original experimental conditions: key components—such as the lamp, monitor, keyboard, mouse, and computer—were transported to each location to closely replicate the initial setup. These adjustments enabled data collection to continue with minimal disruption while maintaining consistency with the intended research design.

3.2.3. Self-report questionnaires

Two self-report questionnaires were administered to assess participants' anxiety and their attitudes toward MTPE. Anxiety was measured using the STAI, a widely recognized psychometric tool designed to evaluate both state anxiety—a temporary emotional condition—and trait anxiety, a stable predisposition to experience anxiety. The STAI consists of 40 items, divided equally between the two dimensions, each rated on a 4-point Likert scale.

Participants' attitudes toward MTPE were assessed using a tailor-made questionnaire, developed to align with the specific aims of the study in the absence of a validated instrument suited to this context. The design drew on previous research examining attitudes towards translation technology (e.g., Bundgaard 2017; Daems 2016; De Almeida 2013; Dillon & Fraser 2006; Guerberof Arenas 2013; Nunes Vieira 2015; Rossi & Chevrot 2019). The questionnaire comprised six statements addressing perceptions of MT as a professional tool, with a focus on perceived productivity, output quality, satisfaction, and its anticipated impact on participants' professional futures (see Appendix C). Each item was rated on a continuous scale from 0 to 100, where 0 indicated an extremely negative perception and 100 an extremely positive one. For hypothesis testing, responses were grouped into three

attitudinal categories: Negative (NEG, 0–44), Neutral (NEU, 45–55), and Positive (POS, 56–100).

One limitation of this study is that participants were randomly assigned to post-edit either NMT or SMT output, while their anxiety levels and attitudes toward MT were assessed post hoc. As a result, the groups were unevenly distributed in terms of baseline anxiety between the NMT and SMT groups—one group reported uniformly low levels, while the other exhibited consistently high levels—potentially reducing statistical power and contributing to the absence of significant differences or correlations in the analysis. Despite this limitation, the design allowed for an initial exploration of how individual affective and attitudinal differences may influence MTPE behavior.

A second limitation concerns the inherent vulnerability of self-report measures to bias. As Jankowiak and Korpál (2018) note, although such measures are widely used in psychological research, they are less robust than physiological indicators. To strengthen the reliability of emotional assessments, future studies could incorporate additional physiological measures alongside pupillometry. Techniques such as facial expression analysis, heart rate monitoring, and galvanic skin response have proven effective in detecting stress-related responses (Weng et al. 2022).

3.3. Experimental Procedure

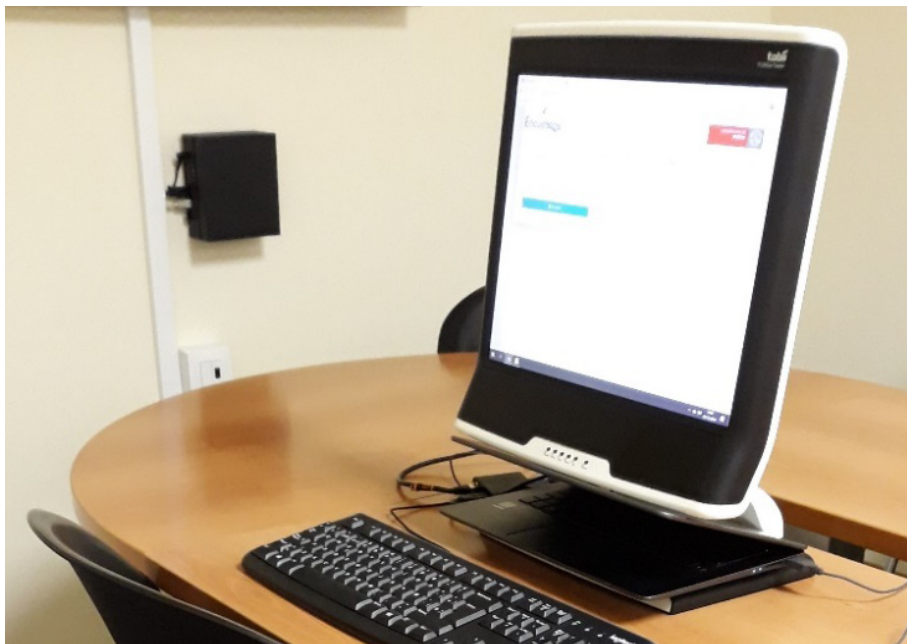
As previously noted, eye-tracking sessions were conducted in two separate locations: data collection for professional translators took place at the offices of *Equus Traducciones*, an LSP based in Granada, Spain, while sessions for translation students were held at the University of Murcia.

Conducting eye-tracking studies outside traditional laboratory settings presents certain challenges—such as ensuring consistent lighting conditions and maintaining a fixed distance between participants and eye-tracking sensors (Holmqvist et al. 2011; Hvelplund 2011; O’Brien 2009; Rösener 2016). However, collecting data from professionals in a real work environment offered a valuable opportunity to capture cognitive and emotional behavior in a realistic context. To mitigate potential disruptions, professional eye-tracking sessions were conducted individually, outside of business hours, in a quiet office equipped with blackout blinds and a stable artificial

light source provided by a floor lamp. These precautions helped ensure high-quality eye-tracking data while maintaining ecological validity.

Student eye-tracking sessions were conducted in a windowless meeting room at the University of Murcia, using the same equipment and set-up as in the professional setting. Participants worked on a laptop running Trados Studio 2017 and Tobii Studio, connected to a Tobii T120 eye-tracker integrated into a 17-inch monitor. The screen, laptop, and eye-tracker were placed on a stable table free of other devices or cables and connected to an external keyboard and mouse to simulate a standard office workstation. Participants sat in non-rolling chairs to minimize movement, and chair height was individually adjusted to ensure optimal positioning relative to the eye-tracker, thereby improving calibration accuracy (Hvelplund 2011).

Figure 2. **Experimental setup featuring the Tobii T120 eye-tracker**



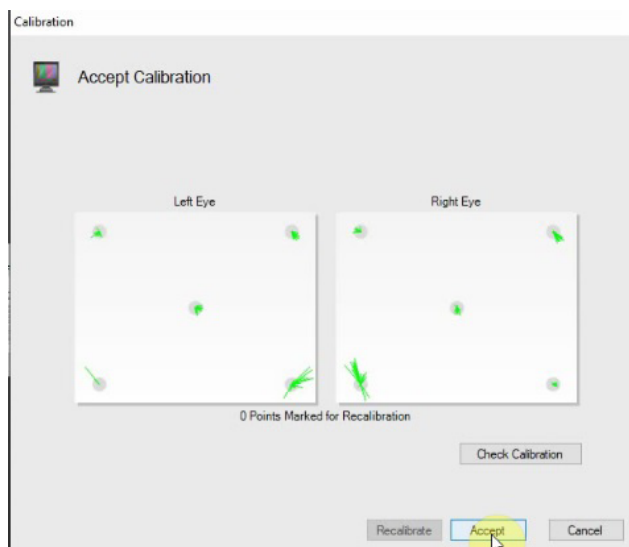
Participants were received individually and first asked to read and sign a consent form outlining the conditions of their participation. Before beginning the main task, the experimental protocol was explained, and

the eye-tracking technology was briefly introduced. Participants were then seated and asked to complete two pre-task questionnaires: the State Anxiety Scale (PRE-STAI-S) and the Trait Anxiety Scale (STAI-T).

A five-point calibration procedure was performed using Tobii Studio to ensure accurate gaze tracking. Prior to their appointment, participants were instructed to avoid wearing mascara, contact lenses, or bifocal glasses, and to tie their hair back to ensure optimal accuracy of the eye-tracking sensors (Holmqvist et al. 2011). When these guidelines were not fully followed, minor onsite adjustments—such as removing eye makeup or adjusting hairstyles—were made to optimize calibration and ensure data quality.

Participants were then shown an on-screen MTPE brief instructing them to perform a full MTPE following TAUS guidelines (Massardo et al. 2016). To provide contextual grounding, a one-minute display of the original newspaper article in PDF format—preserving the visual layout and photographs from its publication in *The Guardian*—was presented before the task began. Participants were not informed of the origin of the MT output (SMT or NMT), were not subject to time constraints, and were allowed to use online resources freely.

Figure 3. Calibration result in Tobii Studio



3.4. Eye-tracking data preparation and quality assurance

Prior to data extraction for statistical analysis, eye-tracking recordings were segmented in Tobii Studio into two task phases: MTPE and online searches (see Figure 4). To analyze visual attention distribution, specific areas of interest (AOIs) were defined: SOURCE TEXT, NMT (DeepL), SMT (PROMT), and INTERNET, which was activated when participants switched to Google Chrome to conduct searches (Figures 5–7).

Figure 4. Task segmentation in a recorded MTPE session

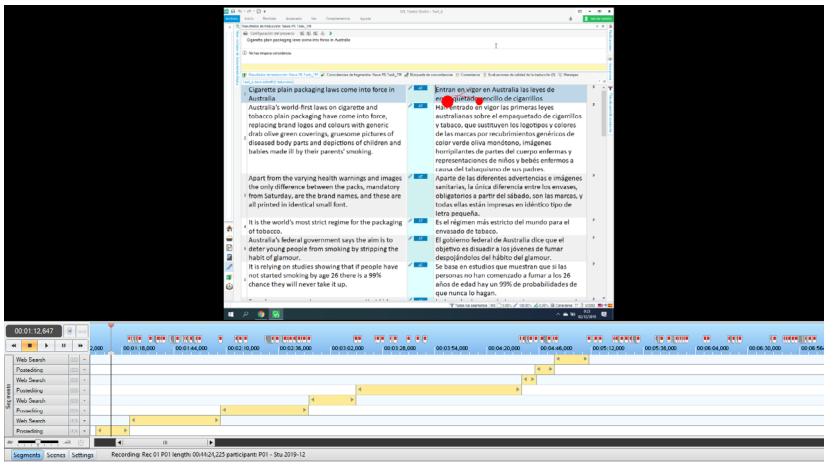


Figure 5. Delimited AOIs for participants who post-edited NMT

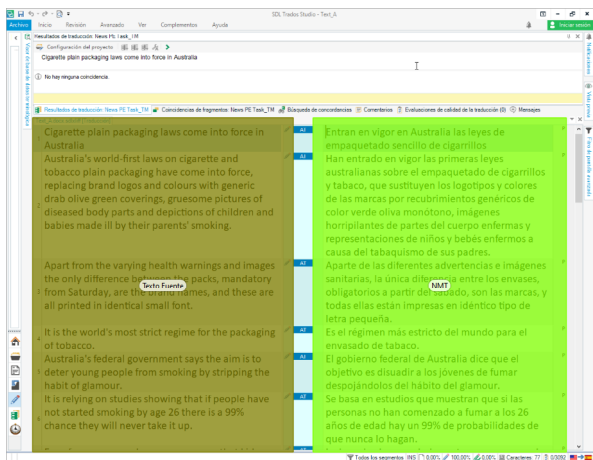


Figure 6. **Defined AOIs for participants who post-edited SMT**

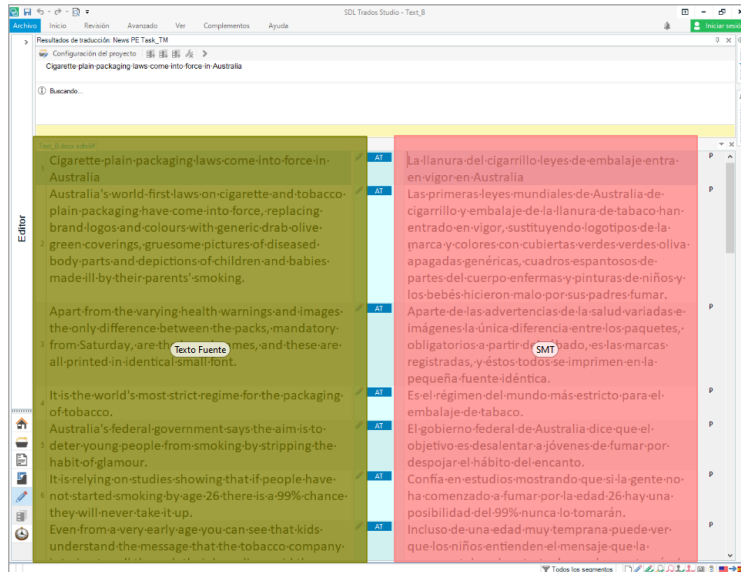
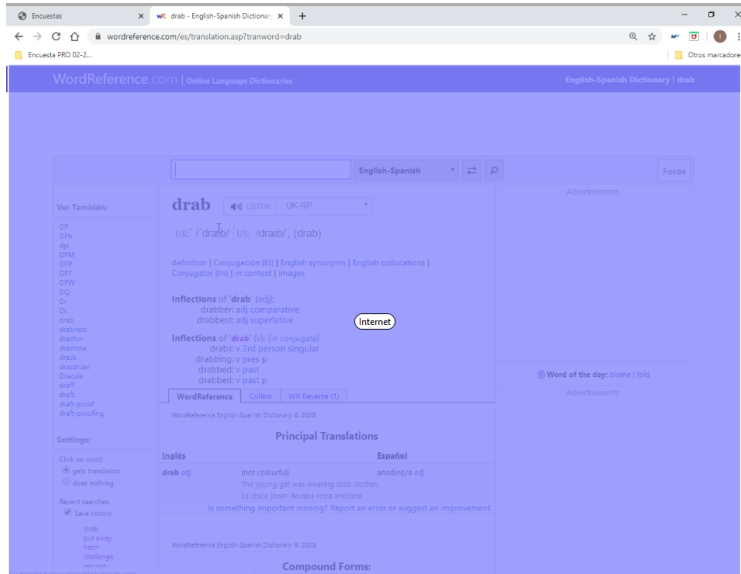


Figure 7. **Internet AOI activated during Google Chrome searches**



Once the recordings were segmented and AOIs delimited, the eye-tracking data were exported to Excel. Several quality assurance measures were applied prior to statistical analysis to ensure the reliability of the dataset. Tobii Studio's confidence metric was used to assess the accuracy with which each eye was detected, ranging from 0 (highest confidence) to 4 (eye not detected). Pupil size measurements from both eyes were averaged, and only those classified with a 'high confidence' score were retained for analysis.

In addition, the quality of the eye-tracking data was evaluated using two criteria: fixation duration and Gaze Time on Screen (GTS) (Hvelplund 2014). These measures were used to identify and exclude unreliable data. The first filter applied was fixation duration, which allowed for the removal of fixations too brief to indicate meaningful visual attention. A threshold of 150 milliseconds was adopted, consistent with previous studies in translation process research, where thresholds typically range from 100 ms (Nunes Vieira 2015) to 180 ms (Sjørup 2013) and 200 ms (Hvelplund 2011).

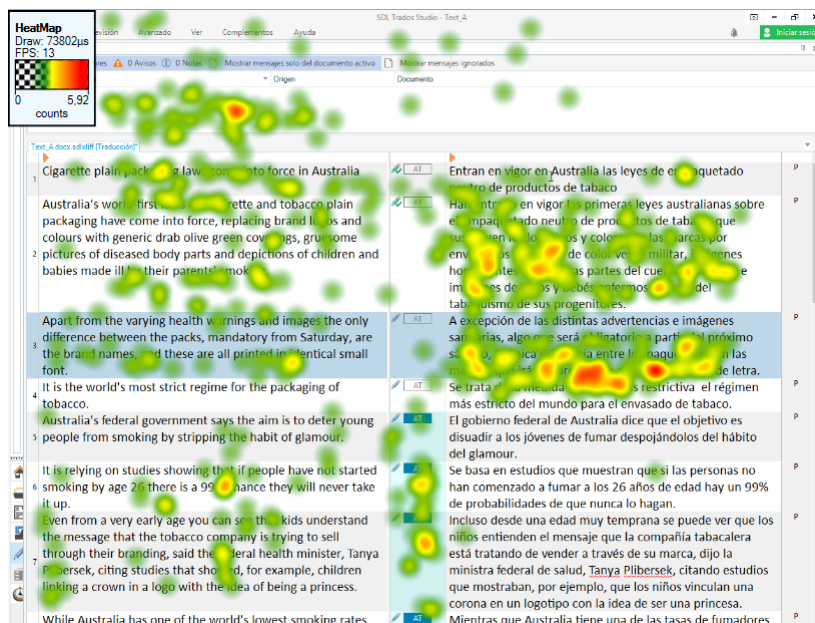
The second criterion, GTS, refers to the proportion of time a participant remained visually engaged with the computer screen during the task (Hvelplund 2014). This metric, provided by Tobii Studio, is calculated by dividing the total fixation duration by the total task time and expressing the result as a percentage. A high GTS score may reflect sustained screen engagement or effective eye-tracking performance, while a low score may indicate limited screen interaction or technical issues. As Hvelplund (2011) notes, low GTS values can compromise the interpretability of cognitive effort measures, underscoring the need to confirm consistent visual engagement throughout the task.

Following Hvelplund's method—setting the exclusion threshold at one standard deviation below the group mean—participants with GTS scores below 68% (professionals) and 61% (students) were excluded. Based on this criterion, three professional translators (49%, 54%, and 67%) and two students (25% and 47%) were removed from the final dataset. These quality control measures ensured that only data reflecting stable and continuous visual attention were retained for analysis.

In addition to the quality criteria outlined above, all recordings were visually inspected using the replay feature in Tobii Studio to identify anomalies in gaze behavior. Two professionals were excluded due to gaze patterns

that deviated from the defined AOIs, two others due to calibration issues caused by bifocal glasses, and one participant due to a technical issue with Trados Studio. Additionally, five student participants were excluded for not adhering to task instructions, having used external MT engines, and pasted the output into Trados Studio instead of completing the task within Trados Studio interface.

Figure 8. Heatmap of eye-tracking recordings for visual inspection



These procedures enhanced the reliability and interpretability of the eye-tracking data. However, the rigorous quality control process reduced the number of professional participants from 25 to 18 and student participants from 27 to 20, resulting in only 73% of the initial sample being included in the final analysis. Such attrition is common in eye-tracking studies, where strict calibration requirements and data quality thresholds frequently result in the exclusion of a substantial portion of participant recordings (Hvelplund 2014).

While the resulting sample size may appear modest and could limit the generalizability of the findings to the broader translator population, it aligns with established practice in eye-tracking studies on translation processes. Such studies typically involve small samples, often ranging from as few as four participants (O'Brien 2007; Moorkens 2018) to no more than twenty-four (Hvelplund 2011). In addition to participant attrition, this is largely due to the methodological complexity and labor-intensive nature of eye-tracking research in translation and MTPE (O'Brien 2009), which involves managing two texts (the source and the MT output) and accounting for the integration of external resource consultation.

While any empirical study may benefit from a larger sample, the 38 participants included in this study were deemed sufficient to support meaningful analysis and allow for cautious generalization of the findings. In addition to a relatively robust sample size, the study was strengthened by the substantial volume of eye-tracking and questionnaire data collected. Specifically, fixation duration analyses were based on 24,641, 17,112, 71,664, and 71,187 data points; pupil dilation analyses included 152,434, 150,544, 146,280, and 156,698 data points; and the attitude questionnaire yielded 14,244, 15,112, and 11,346 data points across different conditions. This high data density enhanced the statistical power of the linear mixed-effects regression (LMER) models, increasing the reliability of the results and the robustness of the identified effects.

3.5. Statistical methods

The data collected in this study were analysed using both descriptive and inferential statistical methods. Descriptive statistics were used to summarise the results across all variables, offering an overall view of the patterns observed. Inferential statistics were then applied to assess whether the differences between student and professional translators, as well as between SMT and NMT conditions, were statistically significant. The following sections detail the inferential methods used in the analysis.

3.5.1. Linear mixed-effects regression analysis

Inferential statistics were conducted using LMER to analyze pupil dilation, fixation duration, and attitude measures. LMER was chosen for its

advantages over traditional factorial ANOVA designs, particularly its ability to account for both fixed and random effects (Baayen 2009; Balling 2008). By incorporating random effects, LMER models estimate individual variation and the dependency between data points, enabling a more accurate assessment of whether observed group differences are statistically significant beyond intra-participant variability (Balling 2008). In this study, each data point represents a recorded pupil size (in millimeters) and a fixation duration (in milliseconds), with the Tobii T120 eye-tracker capturing 120 data points per second per eye. Therefore, LMER is particularly well suited for analyzing repeated measures data in experimental research (Balling & Hvelplund 2015).

In this study, several LMER models were implemented in R (R Core Team 2023) to examine the effects of three independent variables—*Group*, *Attitude*, and *MTOutput*—on the dependent variables *Pupil* and *FixationDuration*. Each model included *Participant* as a random effect to account for individual variability, while one of the independent variables (*Group*, *Attitude*, or *MTOutput*) was included as a fixed effect, depending on the specific research question addressed by the model.

Before fitting the LMER models, the Shapiro-Wilk test confirmed non-normality in pupil dilation and fixation duration data across participant groups. Therefore, a logarithmic transformation was applied to *Pupil* and *Fixationduration* prior to LMER analysis (Balling & Hvelplund 2015). LMER models were analyzed using the *lme4* (Bates et al. 2015) and *lmerTest* (Kuznetsova et al. 2017) packages.

3.5.2. Mann-Whitney U-Test

The Mann-Whitney U test (Mann & Whitney, 1947) was used to analyze anxiety measures, as the Shapiro-Wilk test confirmed non-normality in both student and professional translator datasets. This non-parametric alternative to Student's *t*-test is recommended for small samples and non-normally distributed data (Mellinger & Hanson 2017; Saldanha & O'Brien 2014).

3.5.3. Kendall rank correlation coefficient analysis

Kendall's τ (Kendall, 1938) was used to examine the relationships between emotional variables (e.g., attitude, anxiety, and pupil dilation) and

variables reflecting duration and cognitive effort (e.g., fixation count and fixation duration). This test was selected for its suitability for non-parametric data and its robustness in handling outliers and non-linear relationships (Mellinger & Hanson 2017).

4. Conclusions

This paper examined the methodological challenges of implementing a multimethod research design that integrated physiological, behavioral, and self-report measures in an eye-tracking study on MTPE. It contributed to the growing body of research in CTIS by combining eye-tracking with a psychometric scale (the STAI) and an ad hoc questionnaire. This design enabled the investigation of the interplay between cognitive effort, emotional arousal, and individual factors—specifically anxiety and attitude toward MT—and their influence on visual attention within a commercial CAT tool environment.

A key strength of this eye-tracking study is its commitment to ecological validity. By collecting data from both student and professional translators in controlled yet realistic settings—including an LSP—and by conducting the MTPE task within Trados Studio while allowing the use of Internet resources, the study addresses a well-documented gap in eye-tracking research on MTPE, which has largely lacked naturalistic experiments and real-life case studies (Ragni & Nunes Vieira 2022).

Nevertheless, the study also revealed several methodological challenges that are common across CTIS research. One of the most prominent is the difficulty of operationalizing and measuring complex constructs such as ‘professional experience,’ ‘attitude,’ and ‘emotional arousal’ in ways that are both precise and scalable. While pupillometry provided valuable insights into participants’ physiological responses, interpreting these signals remains problematic due to their association with overlapping cognitive processes. Measuring professional experience was equally challenging, as translators’ engagement with MTPE varies widely across projects and time, making it difficult to capture through standardized metrics such as years of experience alone. Similarly, due to the lack of validated instruments for assessing translators’ attitudes toward MTPE, we relied on an ad hoc questionnaire, which

limits the ability to ensure conceptual alignment between the construct and its measurement, potentially affecting the validity and replicability of the findings (Mellinger & Hanson 2020).

Another limitation of the study arose from the post hoc assessment of anxiety, conducted after participants were randomly assigned to groups. This led to an unintended imbalance in baseline anxiety levels between the NMT and SMT groups, potentially affecting the results and highlighting the importance of accounting for individual affective differences in experimental design.

One of the key methodological challenges in emotion research is accounting for individual variability—a challenge this study addressed through the use of LMER, a statistical method well suited to modeling repeated measures data (e.g., eye-tracking) and capturing both fixed and random effects. By including *Participant* as a random effect, LMER allowed us to control for intra-individual differences, enhancing the reliability of findings despite the heterogeneous sample.

Future MTPE research should more systematically address individual variability through detailed participant profiling prior to group assignment. Measures such as anxiety, personality traits, MTPE experience, CAT tool proficiency, and attitudes toward technology can support group balancing or serve as covariates in statistical models. Comprehensive profiling strengthens research validity and replicability while enabling clearer insight into sample characteristics and better control of individual-level variables that may influence experimental outcomes (Muñoz & Rojo 2025). In addition, within-subject designs offer a powerful means of controlling for individual differences by allowing each participant to serve as their own control. This approach reduces the influence of confounding variables and extraneous factors, thereby enhancing statistical power and enabling the detection of smaller effects within modest sample sizes—a common scenario in CTIS research (Muñoz & Rojo 2025). Complementing this, multimethod and triangulated designs—integrating physiological, behavioral, and self-report data with qualitative insights—would further support a nuanced understanding of the interplay between emotion and cognition.

The rapidly evolving technological landscape introduces further methodological complexities for CTIS. While the present study focused on

the contrast between SMT and NMT systems, the proliferation of machine translation tools—such as domain-specific MT engines and generative AI (e.g., large language models)—demands closer attention in future research. These emerging tools are likely to reshape MTPE workflows and introduce novel cognitive demands, with implications for both cognitive effort and emotional responses. However, research methodologies have not consistently kept pace with these developments, creating a gap between professional practice and the frameworks used to study it. This complicates the operationalization of constructs such as cognitive effort and emotional arousal and highlights the need for experimental designs that more accurately reflect translators' evolving interactions with technology.

Research funding

This work was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación, and FEDER/UE funds (grant number PID2021-123650NB-I00). This funding source played no role in the collection, analysis, and interpretation of data, or the decision to submit the article for publication.

References

- Aziz, Wilker, Sheila Castilho, and Lucia Specia. 2012. "{PET}: A Tool for Post-Editing and Assessing Machine Translation." In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, 3982–87. Istanbul: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/985_Paper.pdf.
- Baayen, Rolf Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511801686>.
- Balling, Laura Winther. 2008. "A Brief Introduction to Regression Designs and Mixed-Effects Modelling by a Recent Convert." *Copenhagen Studies in Language* 36: 175–92. <https://doi.org/10.1075/ts.21013.sil>.

- Balling, Laura Winther, and Kristian Tangsgaard Hvelplund. 2015. "Design and Statistics in Quantitative Translation (Process) Research." *Translation Spaces* 4 (1): 170–187. <https://doi.org/10.1075/ts.4.1.08bal>.
- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Bolaños Medina, Alicia. 2014. "La investigación de las diferencias individuales en traductología cognitiva" *TRANS. Revista de Traductología* 18: 199–214. <https://doi.org/10.24310/TRANS.2014.v0i18.3253>.
- Bundgaard, Kristine. 2017. "Translator Attitudes towards Translator–Computer Interaction: Findings from a Workplace Study." *Hermes* 56: 125–44. <https://doi.org/10.7146/hjlc.v0i56.97228>.
- Carl, Michael, and Matthias Buch-Kromann. 2010. "Correlating Translation Product and Translation Process Data of Professional and Student Translators." In *EAMT 2010—14th Annual Conference of the European Association for Machine Translation*. Saint Raphaël: European Association for Machine Translation.
- Chiang, Yung Nan. 2010. "Foreign Language Anxiety and Student Interpreters' Learning Outcomes: Implications for the Theory and Measurement of Interpretation Learning Anxiety." *Meta* 55 (3): 589–601. <https://doi.org/10.7202/045079AR>.
- Cifuentes-Férez, Paula, Ana María Rojo López, and Laura Espín López. 2024. "Self-Efficacy as a Protective Factor When Translating under Time Pressure." *Vigo International Journal of Applied Linguistics*, no. 21. <https://doi.org/10.35869/vial.v0i21.4215>.
- Cooper, Alan. 2004. *The Inmates Are Running the Asylum: Why Hi-Tech Products Drive Us Crazy and How to Restore the Sanity*. Indianapolis: Sams Publishing.
- Daems, Joke. 2016. *A Translation Robot for Each Translator? A Comparative Study of Manual Translation and Post-Editing of Machine Translations: Process, Quality and Translator Attitude*. PhD Thesis, Ghent University.
- De Almeida, Giselle. 2013. *Translating the Post-Editor: An Investigation of Post-Editing Changes and Correlations with Professional Experience across Two Romance Languages*. PhD Thesis, Dublin City University.
- Dillon, Sarah, and Janet Fraser. 2006. "Translators and TM: An Investigation of Translators' Perceptions of Translation Memory Adoption." *Machine Translation* 20 (2): 67–79. <https://doi.org/10.1007/s10590-006-9004-8>.

- Gieshoff, Anne Catherine, Caroline Lehr, and Andrea Hunziker Heeb. 2021. "Stress, Cognitive, Emotional and Ergonomic Demands in Interpreting and Translation." *Cognitive Linguistic Studies* 8 (2): 404–39. <https://doi.org/10.1075/cogls.00084.gie>.
- Guerberof Arenas, Ana. 2013. "What Do Professional Translators Think about Post-Editing?" *JoSTrans: The Journal of Specialised Translation* 19: 75–95.
- Holmqvist, Kenneth, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost van de Weijer. 2011. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press.
- Hvelplund, Kristian Tangsgaard. 2017. "Translators' Use of Digital Resources during Translation." *Hermes* 56: 71–87. <https://doi.org/10.7146/hjlc.v0i56.97205>.
- . 2016. "Cognitive Efficiency in Translation." In *Reembedding Translation Process Research*, edited by Ricardo Muñoz Martín, 149–70. Amsterdam and Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/btl.128.08hve>.
- . 2014. "Eye Tracking and the Translation Process: Reflections on the Analysis and Interpretation of Eye-Tracking Data." In *MonTI. Monografías de Traducción e Interpretación*, edited by Ricardo Muñoz Martín, 201–23. Alicante: Publicaciones de la Universidad de Alicante. <https://doi.org/10.6035/monti.2014.ne1.6>.
- . 2011. *Allocation of Cognitive Resources in Translation: An Eye-Tracking and Key-Logging Study*. PhD Thesis, Copenhagen Business School.
- Jääskeläinen, Riitta. 2010. "Are All Professionals Experts?" In *Translation and Cognition*, edited by Gregory M. Shreve and Erik Angelone, 213–27. Amsterdam and Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/ata.xv.12jaa>.
- Jankowiak, Katarzyna, and Paweł Korpala. 2018. "On Modality Effects in Bilingual Emotional Language Processing: Evidence from Galvanic Skin Response." *Journal of Psycholinguistic Research* 47 (3): 663–77. <https://doi.org/10.1007/s10936-017-9552-5>.
- Kendall, Maurice G. 1938. "A New Measure of Rank Correlation." *Biometrika* 30 (1–2): 81–93.
- Koglin, Arlene. 2015. "An Empirical Investigation of Cognitive Effort Required to Post-Edit Machine Translated Metaphors Compared to the Translation of Metaphors." *Translation & Interpreting* 7 (1): 126–141.
- Koponen, Maarit, Wilker Aziz, Luciana Ramos, and Lucia Specia. 2012. "Post-Editing Time as a Measure of Cognitive Effort." In *Proceedings of the AMTA 2012 Workshop on Post-Editing Technology and Practice*, 11–20. http://amta2012.amtaweb.org/AMTA-2012Files/html/13/13_paper.pdf

- Korpál, Paweł. 2017. *Linguistic and Psychological Indicators of Stress in Simultaneous Interpreting*. Poznań: Adam Mickiewicz University Press.
- Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82 (13): 1–26. <https://doi.org/10.18637/jss.v082.i13>.
- Lavie, Alon, and Michael J. Denkowski. 2009. "The Meteor Metric for Automatic Evaluation of Machine Translation." *Machine Translation* 23 (2): 105–15. <https://doi.org/10.1007/s10590-009-9059-4>.
- Lehr, Caroline, and Kristian Tangsgaard Hvelplund. 2020. "Emotional Experts: Influences of Emotion on the Allocation of Cognitive Resources during Translation." In *Multilingual Mediated Communication and Cognition*, edited by Ricardo Muñoz Martín and Sandra L. Halverson, 44–68. Abingdon: Routledge.
- Mann, H. B., and D. R. Whitney. 1947. "On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other." *The Annals of Mathematical Statistics* 18 (1): 50–60. <https://doi.org/10.1214/aoms/1177730491>.
- Massardo, Isabella, Jaap van der Meer, Sharon O'Brien, Fred Hollowood, Nora Aranberri, and Katrin Drescher. 2016. *MT Postediting Guidelines*. TAUS Signature Editions.
- Mellinger, Christopher D., and Thomas A. Hanson. 2020. "Methodological Considerations for Survey Research: Validity, Reliability, and Quantitative Analysis." *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 19: 172–90. <https://doi.org/10.52034/lanstts.v19i0.549>.
- Mellinger, Christopher, and Thomas Hanson. 2017. *Quantitative Research Methods in Translation and Interpreting Studies*. London and New York: Routledge. <https://doi.org/10.4324/9781315647845>.
- Moorkens, Joss. 2018. "Eye Tracking as a Measure of Cognitive Effort for Post-Editing of Machine Translation." In *Eye Tracking and Multidisciplinary Studies on Translation*, edited by Callum Walker and Federico M. Federici, 55–70. Amsterdam and Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/btl.143>.
- Moorkens, Joss, Sharon O'Brien, Igor da Silva, Norma Fonseca, and Fabio Alves. 2015. "Correlations of Perceived Post-Editing Effort with Measurements of Actual Effort." *Machine Translation* 29 (3): 267–84. <https://doi.org/10.1007/s10590-015-9175-2>.
- Moser-Mercer, Barbara. 2005. "Remote Interpreting: Issues of Multi-Sensory Integration in a Multilingual Task." *Meta* 50 (2): 727–38. <https://doi.org/10.7202/011014ar>.

- Muñoz Martín, Ricardo, and Ana María Rojo López. 2025. "Introduction." In *Research Methods in Cognitive Translation and Interpreting Studies*, 1–20. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/rmal.10.intro>.
- Nimdzi. 2025. *The 2025 Nimdzi 100: The Size and State of the Language Services Industry*. MultiLingual Media LLC.
- Nitzke, Jean. 2019. *Problem Solving Activities in Post-Editing and Translation from Scratch: A Multi-Method Study*. New Empirical Perspectives on Translation and Interpreting. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.2546446>.
- Nunes Vieira, Lucas. 2018. "Automation Anxiety and Translators." *Translation Studies* (13) 1: 1–21. <https://doi.org/10.1080/14781700.2018.1543613>.
- . 2015. *Cognitive Effort in Post-Editing of Machine Translation: Evidence from Eye Movements, Subjective Ratings, and Think-Aloud Protocols*. PhD Thesis, Newcastle University.
- O'Brien, Sharon. 2009. "Eye Tracking in Translation-Process Research: Methodological Challenges and Solutions." In *Methodology, Technology and Innovation in Translation Process Research: A Tribute to Arnt Lykke Jakobsen*, edited by Inger M. Mees, Susanne Göpferich, and Fabio Alves, 15–36. Frederiksberg: Samfundslitteratur. <https://doi.org/10.1075/target.24.1.13tir>.
- . 2007. "Eye-tracking and Translation Memory Matches." *Perspectives: Studies in Translatology* 14 (3): 185–205. <https://doi.org/10.1080/09076760708669037>.
- O'Brien, Sharon, and Maureen Ehrensberger-Dow. 2020. "MT Literacy-A Cognitive View." *Translation, Cognition and Behavior* 3 (2): 145–64. <https://doi.org/10.1075/tcb.00038.obr>.
- Pinazo Calatayud, Daniel, and Amparo Jiménez Ivars. 2002. "Fear of public speaking, state anxiety and performance: The case of interpreter trainees." *Ansiedad y estrés* 8 (1): 37–47. <https://www.ansiedadystres.es/sites/default/files/rev/ucm/2002/anyes2002a4.pdf>.
- Pym, Anthony. 2013. "Translation Skill-Sets in a Machine-Translation Age." *Meta* 58 (3): 487–503. <https://doi.org/10.7202/1025047ar>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ragni, Valentina, and Lucas Nunes Vieira. 2022. "What Has Changed with Neural Machine Translation? A Critical Review of Human Factors." *Perspectives* 30 (1): 137–58. <https://doi.org/10.1080/0907676X.2021.1889005>.
- Rojo López, Ana María, and Paula Cifuentes Férez. 2021. "Introduction: Experimenting with Emotions: Insights into Empirical Emotion Research in Cognitive Trans-

- lation Studies.” *Onomázein Revista de Lingüística Filología y Traducción*, no. 8, 1–15. <https://doi.org/10.7764/onomazein.ne8.01>.
- Rojo López, Ana María, Paula Cifuentes Férez, and Laura Espín López. 2021a. “The Influence of Time Pressure on Translation Trainees’ Performance: Testing the Relationship between Self-Esteem, Salivary Cortisol and Subjective Stress Response.” *PLoS ONE* 16 (9): e0257727. <https://doi.org/10.1371/journal.pone.0257727>.
- Rojo López, Ana María, Ana Isabel Foulquié-Rubio, Laura Espín López, and Francisco Martínez Sánchez. 2021b. “Analysis of Speech Rhythm and Heart Rate as Indicators of Stress on Student Interpreters.” *Perspectives* 29 (4): 591–607. <https://doi.org/10.1080/0907676X.2021.1900305>.
- Rojo López, Ana María, and Ricardo Muñoz Martín. 2022. “Translation Process Research.” In *The Routledge Handbook of Translation and Methodology*, 1st ed., 356–72. London: Routledge. <https://doi.org/10.4324/9781315158945-26>.
- Rojo López, Ana María, and Marina Ramos Caro. 2022. “Introduction: Probing the Process: Towards More Integrative Research Methods.” *The International Journal of Translation and Interpreting Research* 14 (2): 1–7. <https://doi.org/10.12807/ti.114202.2022.a01>.
- . 2018. “The Role of Expertise in Emotion Regulation.” In *Innovation and Expansion in Translation Process Research*, edited by Isabel Lacruz and Rita Jääskeläinen, 105–29. Amsterdam and Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/ata.18.06roj>.
- Rösener, Christoph. 2016. “Eye Tracking and Beyond: The Dos and Don’ts of Creating a Contemporary Usability Lab.” In *Eyetracking and Applied Linguistics*, edited by Silvia Hansen-Schirra and Sambor Grucza, 143–63. Berlin: Language Science Press. <https://doi.org/10.17169/langsci.b108.297>.
- Rossi, Caroline, and Jean Pierre Chevrot. 2019. “Uses and Perceptions of Machine Translation at the European Commission.” *The Journal of Specialised Translation*, no. 31: 177–200. <https://shs.hal.science/halshs-01893120v2>.
- Saldanha, Gabriela, and Sharon O’Brien. 2013. *Research Methodologies in Translation Studies*. London: Routledge.
- Seisdedos Cubero, Nicolás. 1997. *STAI: Cuestionario de Ansiedad Estado-Rasgo: Manual*. Adaptación española del cuestionario y redacción del manual realizadas por Nicolás Seisdedos. Madrid: TEA Ediciones.
- Sjørup, Annette Camille. 2013. *Cognitive Effort in Metaphor Translation: An Eye-Tracking and Key-Logging Study*. PhD Thesis, Copenhagen Business School.

- Spielberger, C. D., R. L. Gorsuch, R. Lushene, P. R. Vagg, and G. A. Jacobs. 1983. *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press. <https://psycnet.apa.org/doiLanding?doi=10.1037%2F06496-000>.
- Stasimioti, Maria, and Vilelmini Sosoni. 2021. "Investigating Post-Editing: A Mixed-Methods Study with Experienced and Novice Translators in the English-Greek Language Pair." In *Translation, Interpreting, Cognition: The Way out of the Box*, edited by Tra&Co Group, 79–104. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.4545037>.
- Stasimioti, Maria, Vilelmini Sosoni, Despoina Mouratidis, and Katia Kermanidis. 2020. "Machine Translation Quality: A Comparative Evaluation of SMT, NMT and Tailored-NMT Outputs." In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT 2020)*, Lisboa, Portugal, 441–50.
- Tatsumi, Midori, and Johann Roturier. 2010. "Source Text Characteristics and Technical and Temporal Post-Editing Effort: What Is Their Relationship?" In *Proceedings of the Second Joint EM+/CNGL Workshop: Bringing MT to the User – Research on Integrating MT in the Translation Industry (JEC 2010)*, 43–51. <http://www.mt-archive.info/JEC-2010-Tatsumi.pdf>.
- Teixeira, Carlos S. C. 2014. *The Impact of Metadata on Translator Performance: How Translators Work with Translation Memories and Machine Translation*. PhD Thesis, Universitat Rovira i Virgili. <https://doi.org/10.13140/RG.2.1.2190.2887>.
- Tobii Technology. 2010. *Product Description for Tobii T & X Series Eye Trackers*. Stockholm: Tobii Technology. <https://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-tx-product-description.pdf>.
- Vanroy, Bram, Arda Tezcan, and Lieve Macken. 2023. "MATEO: Machine Translation Evaluation Online." In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, 499–500, <https://aclanthology.org/2023.eamt-1.52>.
- Vicente López, María Inmaculada. 2024. *Aspectos cognitivos y emocionales de la posesición de traducción automática: Un estudio de movimientos oculares*. PhD Thesis, Universidad de Murcia.
- Weng, Yu, and Binghan Zheng. 2020. "A Multi-Methodological Approach to Studying Time-Pressure in Written Translation: Manipulation and Measurement." *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 19:218–36. <https://doi.org/10.52034/lanstts.v19i0.548>.

Weng, Yu, Binghan Zheng, and Yanping Dong. 2022. “Time Pressure in Translation.” *Target: International Journal of Translation Studies* 34 (4): 601–26. <https://doi.org/10.1075/target.20148.wen>.

Appendix A. Comparative table of source text and NMT output

N.º	Source Text 191 words	MT Output - NMT 220 words
1	Cigarette plain packaging laws come into force in Australia	Entran en vigor en Australia las leyes de empaquetado sencillo de cigarrillos
2	Australia's world-first laws on cigarette and tobacco plain packaging have come into force, replacing brand logos and colours with generic drab olive green coverings, gruesome pictures of diseased body parts and depictions of children and babies made ill by their parents' smoking.	Han entrado en vigor las primeras leyes australianas sobre el empaquetado de cigarrillos y tabaco, que sustituyen los logotipos y colores de las marcas por recubrimientos genéricos de color verde oliva monótono, imágenes horripilantes de partes del cuerpo enfermas y representaciones de niños y bebés enfermos a causa del tabaquismo de sus padres.
3	Apart from the varying health warnings and images the only difference between the packs, mandatory from Saturday, are the brand names, and these are all printed in identical small font.	Aparte de las diferentes advertencias e imágenes sanitarias, la única diferencia entre los envases, obligatorios a partir del sábado, son las marcas, y todas ellas están impresas en idéntico tipo de letra pequeña.
4	It is the world's most strict regime for the packaging of tobacco.	Es el régimen más estricto del mundo para el envasado de tabaco.
5	Australia's federal government says the aim is to deter young people from smoking by stripping the habit of glamour.	El gobierno federal de Australia dice que el objetivo es disuadir a los jóvenes de fumar despojándolos del hábito del glamour.
6	It is relying on studies showing that if people have not started smoking by age 26 there is a 99% chance they will never take it up.	Se basa en estudios que muestran que si las personas no han comenzado a fumar a los 26 años de edad hay un 99% de probabilidades de que nunca lo hagan.
7	Even from a very early age you can see that kids understand the message that the tobacco company is trying to sell through their branding, said the federal health minister, Tanya Plibersek, citing studies that showed, for example, children linking a crown in a logo with the idea of being a princess.	Incluso desde una edad muy temprana se puede ver que los niños entienden el mensaje que la compañía tabacalera está tratando de vender a través de su marca, dijo la ministra federal de salud, Tanya Plibersek, citando estudios que mostraban, por ejemplo, que los niños vinculan una corona en un logotipo con la idea de ser una princesa.

Appendix B. Comparative table of source text and SMT output

N.º	Source Text 191 words	MT Output - NMT 220 words
1	Cigarette plain packaging laws come into force in Australia	La llanura del cigarrillo leyes de embalaje entra en vigor en Australia
2	Australia's world-first laws on cigarette and tobacco plain packaging have come into force, replacing brand logos and colours with generic drab olive green coverings, gruesome pictures of diseased body parts and depictions of children and babies made ill by their parents' smoking.	Las primeras leyes mundiales de Australia de cigarrillo y embalaje de la llanura de tabaco han entrado en vigor, sustituyendo logotipos de la marca y colores con cubiertas verdes verdes oliva apagadas genéricas, cuadros espantosos de partes del cuerpo enfermas y pinturas de niños y los bebés hicieron malo por sus padres fumar.
3	Apart from the varying health warnings and images the only difference between the packs, mandatory from Saturday, are the brand names, and these are all printed in identical small font.	Aparte de las advertencias de la salud variadas e imágenes la única diferencia entre los paquetes, obligatorios a partir del sábado, es las marcas registradas, y éstos todos se imprimen en la pequeña fuente idéntica.
4	It is the world's most strict regime for the packaging of tobacco.	Es el régimen del mundo más estricto para el embalaje de tabaco.
5	Australia's federal government says the aim is to deter young people from smoking by stripping the habit of glamour.	El gobierno federal de Australia dice que el objetivo es desalentar a jóvenes de fumar por despojar el hábito del encanto.
6	It is relying on studies showing that if people have not started smoking by age 26 there is a 99% chance they will never take it up.	Confía en estudios mostrando que si la gente no ha comenzado a fumar por la edad 26 hay una posibilidad del 99% nunca lo tomarán.
7	Even from a very early age you can see that kids understand the message that the tobacco company is trying to sell through their branding, said the federal health minister, Tanya Plibersek, citing studies that showed, for example, children linking a crown in a logo with the idea of being a princess.	Incluso de una edad muy temprana puede ver que los niños entienden el mensaje que la empresa tabacalera trata de venderse a través de su marcación, dijo la ministra de salud federal, Tanya Plibersek, citando estudios que mostraron, por ejemplo, a niños que unen una corona en un logotipo con la idea de ser una princesa.

Appendix C. Ad hoc questionnaire on translators' attitudes toward machine translation

<p>Las siguientes preguntas están relacionadas con tu opinión personal sobre la tecnología de traducción automática.</p> <p>Por favor, contesta con sinceridad indicando un valor en una escala de 0 a 100, en la que 0 = Totalmente en desacuerdo y 100 = Totalmente de acuerdo.</p>
Me gusta consultar la traducción realizada por un motor de traducción automática como referencia para elaborar mi propia traducción.
Utilizar un motor de traducción automática para elaborar mi traducción me permite trabajar más rápido.
Al posteditar traducción automática cometo menos errores de traducción.
Me preocupa que la tecnología de traducción automática sustituya a los traductores humanos en el futuro y haga desaparecer esta profesión.
Me resulta frustrante la calidad de las traducciones que proponen los motores de traducción automática porque pienso que yo podría hacerlo mejor.
Las traducciones que proponen los motores de traducción automática son muy fiables.
¿Te has sentido cómodo trabajando con este traductor automático? ¿Por qué?
¿Cuales son las principales ventajas e inconvenientes que has notado?
¿Estas familiarizado o tienes experiencia en la traducción de textos periodísticos?
¿Estas satisfecho/a con el resultado final de tu postedición? Explica por qué.
¿Crees que las traducciones propuestas por este motor han influido en la calidad final del texto? Explica si la influencia ha sido positiva o negativa.

Methodological considerations for emotion research in MTPE: A multimethod design using eye-tracking and self-report questionnaires

Abstract

This paper explores the methodological challenges of researching cognitive and emotional dimensions in machine translation post-editing (MTPE) through a multimethod design. The study combined psychometric, physiological, and behavioral data to examine differences in translators' anxiety, emotional arousal, cognitive effort, and visual attention during an English-to-Spanish MTPE task involving statistical machine translation (SMT) and neural machine translation (NMT). Emotional arousal was measured via pupil dilation, anxiety through the State-Trait Anxiety Inventory, and cognitive effort via fixation durations. Visual attention was analyzed through fixation counts on the source text, machine translation (MT) output, and Internet browser. The influence of MT technology, anxiety, and attitude on visual attention distribution was explored using data from both professional and student translators working in ecologically valid conditions. This study reflects critically on the methodological challenges encountered, including issues related to construct operationalization, participant variability, and ecological validity. At the same time, it demonstrates the value of a multimethod research design—combining eye-tracking and self-report measures—for investigating the interplay between cognition and emotion in MTPE. Finally, it argues that the rise of new MT technologies, including generative AI, poses an emerging methodological challenge that future Cognitive Translation and Interpreting Studies (CTIS) should address.

Keywords

Eye-tracking, machine translation post-editing, cognitive effort, emotional arousal, anxiety, attitude toward MT, visual attention

Author

María Inmaculada Vicente López
University of Murcia.
minmaculada.vicente@um.es
<https://orcid.org/0000-0002-6169-0805>

How to cite

Vicente López, María Inmaculada. 2025. "Methodological considerations for emotion research in MPTE: a multimethod design using eye-tracking and self-report questionnaires." *MeTI: Methods in Translation and Interpreting* 0: xxx-xxx.