Human-centered, augmented machine translation: analysing user experience, quality and productivity in interactive post-editing vs traditional post-editing

Vicent Briva-Iglesias

Vicent Briva-Iglesias
Dublin City University. SALIS, CTTS, ADAPT Centre;
vicent.brivaiglesias@dcu.ie;
ORCID: 0000-0001-8525-2677

Abstract

Recent language technology developments have disrupted the translation profession. Traditionally, the research and industry focus has been on using more computational power and training larger language models, often neglecting the users of such technology. To date, the goal of technology development has been the creation of an intelligent agent that emulates human behaviour to increase automation. As a response, a novel technology design framework has gained a foothold recently: human-centered artificial intelligence (HCAI), where, instead of human replacement, the aim is to align the development of tools with user values, preferences, and needs, at the same time as enhancing human capabilities and performance. If applied to machine translation (MT), we can talk about human-centered, augmented MT (HCAMT). This shift, moving from emulation to empowerment, places humans at the center of AI/language technology. This paper considers the analysis of machine translation user experience (MTUX) to be a way to foster HCAMT. To demonstrate this, we conduct a longitudinal user study with 11 professional translators in the English-Spanish language combination to analyse the effects of traditional post-editing (TPE) and interactive post-editing (IPE) on MTUX, translation quality, and productivity. MTUX results suggest that translators prefer IPE to TPE because they feel more in control in this newer form of translator-computer interaction and more empowered in their work with MT. Productivity results suggest that translators working with IPE achieve statistically significantly higher productivity than when working with TPE. Quality results indicate that translators offer more fluent translations in IPE, and equally adequate translations in both post-editing modalities. All these results allow for reflection on the potential adoption of IPE as a more human-centred MT modality for post-editing, one that empowers users, who have been increasingly reluctant to interact with MT post-editing in industry workflows. The paper also lays the basis for exploring HCAMT with MT users beyond professional translators, opening the door to more inclusive and diverse user-centred MT research.

Keywords: human-centred AI, user experience, translation technology, human-computer interaction, post-editing, augmentation.

Rebuda: 28 de febrer de 2025| Acceptació: 9 de octubre de 2025 | Publicació: 15 de desembre de 2025

Cita recomanada: BRIVA-IGLESIAS, VICENT (2025). «Human-centered machine translation: analysing user experience, quality and productivity in interactive post-editing vs traditional post-editing». Revista Tradumàtica. Tecnologies de la Traducció, 23, 350-382. <https://doi.org/10.5565/rev/tradumatica.504>

http://revistes.uab.cat/tradumatica
ISSN 1578-7559
Universitat Autònoma de Barcelona

Resumen

Los recientes avances en las tecnologías del lenguaje han alterado la profesión de la traducción. Tradicionalmente, tanto la investigación como la industria se han centrado en el uso de una mayor potencia computacional y en el entrenamiento de modelos lingüísticos más grandes, a menudo dejando de lado a los usuarios de estas tecnologías. Hasta la fecha, el objetivo del desarrollo tecnológico ha sido la creación de un agente inteligente que emule el comportamiento humano para aumentar la automatización. Como respuesta, recientemente ha ganado presencia un nuevo marco de diseño tecnológico: la inteligencia artificial centrada en las personas (HCAI), en la que, en lugar de sustituir a los seres humanos, el objetivo es alinear el desarrollo de las herramientas con los valores, las preferencias y las necesidades de los usuarios y, al mismo tiempo, aumentar sus capacidades y mejorar su rendimiento. Si se aplica a la traducción automática (TA), podemos hablar de traducción automática aumentada y centrada en las personas (HCAMT). Este cambio, que pasa de la emulación al empoderamiento, sitúa a las personas en el centro de la IA y de las tecnologías del lenguaje. Este artículo plantea el análisis de la experiencia de usuario en traducción automática (MTUX) como una vía para fomentar la HCAMT. Para demostrarlo, llevamos a cabo un estudio longitudinal con 11 traductores profesionales en la combinación lingüística inglés–español para analizar los efectos de la posedición tradicional (TPE) y la posedición interactiva (IPE) sobre la MTUX, la calidad de la traducción y la productividad. Los resultados de MTUX sugieren que los traductores prefieren la IPE a la TPE porque sienten que tienen más control sobre la interacción en esta forma más reciente de interacción traductor–ordenador y se sienten más empoderados en su relación con la TA. Los resultados de productividad también indican que los traductores que trabajan con IPE informan de una productividad significativamente mayor, a efectos estadísticos, que cuando trabajan con TPE. En cuanto a la calidad, los resultados indican igualmente que los traductores ofrecen traducciones más fluidas con IPE y traducciones igual de adecuadas en ambas modalidades de posedición. Todos estos resultados permiten reflexionar sobre la posible adopción de la IPE como una modalidad de TA para la posedición más centrada en las personas, que empodera a los usuarios, cada vez más reticentes a interactuar con la posedición de TA en los flujos de trabajo de la industria. El artículo también sienta las bases para explorar la HCAMT entre usuarios de TA más allá de los traductores profesionales, abriendo la puerta a una investigación en TA centrada en el usuario más inclusiva y diversa.

Palabras clave: IA centrada en humanos, experiencia de usuario, tecnología de la traducción, interacción humano-ordenador, traducción automática, posedición, potenciación.

Resum

Els avenços recents en tecnologies del llenguatge han alterat la professió de la traducció. Tradicionalment, tant la recerca com la indústria s'han centrat en l'ús de més potència computacional i en l'entrenament de models lingüístics més grans, sovint passant per alt els usuaris d'aquestes tecnologies. Fins ara, l'objectiu del desenvolupament tecnològic ha estat la creació d'un agent intel·ligent que emuli el comportament humà per augmentar l'automatització. Com a resposta, recentment ha guanyat presència un nou marc de disseny tecnològic: la intel·ligència artificial centrada en les persones (HCAI), en què, en lloc de substituir els humans, l'objectiu és alinear el desenvolupament de les eines amb els valors, les preferències i les necessitats dels usuaris i, alhora, augmentar-ne les capacitats i millorar-ne el rendiment. Si s'aplica a la traducció automàtica (TA), podem parlar de traducció automàtica augmentada i centrada en les persones (HCAMT). Aquest canvi, que passa de l'emulació a l'empoderament, situa les persones al centre de la IA i de les tecnologies del llenguatge. Aquest article considera l'anàlisi de l'experiència d'usuari en traducció automàtica (MTUX) com una via per fomentar la HCAMT. Per demostrar-ho, duem a terme un estudi longitudinal amb 11 traductors professionals en la combinació lingüística anglès–espanyol per analitzar els efectes de la postedició tradicional (TPE) i de la postedició interactiva (IPE) sobre l'MTUX, la qualitat de la traducció i la productivitat. Els resultats de MTUX suggereixen que els traductors prefereixen la IPE a la TPE perquè senten que tenen més control sobre la interacció en aquesta forma més recent d'interacció traductor–ordinador i se senten més empoderats en la seva relació amb la TA. Els resultats de productivitat també indiquen que els traductors que treballen amb IPE informen d'una productivitat significativament més alta, a efectes estadístics, que quan treballen amb TPE. Pel que fa a la qualitat, els resultats indiquen igualment que els traductors ofereixen traduccions més fluides amb IPE i traduccions igualment adequades en ambdues modalitats de postedició. Tots aquests resultats permeten reflexionar sobre la possible adopció de la IPE com una modalitat de TA per a la postedició més centrada en les persones, que empodera els usuaris, cada vegada més reticents a interactuar amb la postedició de TA en els fluxos de treball de la indústria. L'article també estableix les bases per explorar la HCAMT entre usuaris de TA més enllà dels traductors professionals, obrint la porta a una recerca en TA centrada en l'usuari més inclusiva i diversa.

Paraules clau: IA centrada en humans, experiència d'usuari, tecnologia de la traducció, interacció humà-ordinador, traducció automàtica, postedición, potenciació.

1. Introduction

In an increasingly globalised and digitalised world, machine translation (MT) has emerged as a fundamental tool for facilitating communication across linguistic and cultural boundaries (Vieira et al., 2021). The last decade has witnessed rapid MT advances, largely driven by the availability of vast amounts of data, increased computational power, and the development of sophisticated algorithms, particularly in the field of neural machine translation (NMT) and large language models (Forcada, 2017; Brown et al., 2020). These advances have yielded remarkable improvements in MT output quality, especially for high-resource language pairs, and have solidified the role of MT in both professional and everyday communication contexts (ELIS Research, 2025; Bowker, 2023).

Despite these achievements, the development of MT technologies has been predominantly technical in nature. The prevailing focus has been on enhancing translation accuracy, fluency, and overall system performance, often measured through automatic metrics or isolated quality evaluations (Rossi & Carré, 2022; Vanroy et al., 2023). However, less attention has been paid to the human and social dimensions in the process of MT development for its adoption and use (Briva-Iglesias et al., 2023; Carpuat et al., 2025). In other words, MT has been primarily conceived and developed as an autonomous technical system, rather than as a technology embedded within complex sociotechnical systems involving human users with specific needs, expectations, and experiences.

This technical emphasis reflects a broader trend within artificial intelligence (AI) research, which historically prioritised automation and the replication of human intelligence (Shneiderman, 2022a). According to the recently approved European Union Artificial Intelligence Act, "AI systems" are defined as machine-based systems operating with varying levels of autonomy, capable of generating outputs that influence their environment (European Union AI Act, 2024). We are already witnessing cases where MT is deployed as an AI system without human supervision when assimilation, rather than dissemination, is the goal, both in industry (Schmidtke & Groves, 2019) and society (Briva-Iglesias & Peñuelas Gil, 2025).

This paper argues that, for MT to achieve sustainable, ethical, and inclusive integration into professional and societal workflows, a shift from a purely technical to a sociotechnical development model is urgently needed. Following the emerging framework of human-centered artificial intelligence (HCAI) (Shneiderman, 2022a), we propose an approach that places users, and not just the outputs, at the center of MT research and development. We introduce the notion of human-centered, augmented machine translation (HCAMT), which prioritises the augmentation of human capabilities, user empowerment, and positive interaction experiences with MT systems (O'Brien, 2023).

Within this perspective, we position what was proposed as MT user experience (MTUX) (Briva-Iglesias & O'Brien, 2023) as a critical factor that can influence not only user satisfaction and acceptance but also translation productivity and quality. Previous studies have shown that negative experiences with post-editing can lead to translator resistance (Cadwell et al., 2018; Fırat, 2021; Briva-Iglesias & O'Brien, 2024), highlighting the urgent need for more human-centered approaches.

Although the range of MT users is vast, in this paper we limit our study to professional translators. Consequently, we conducted a two-week longitudinal study with 11 professional translators working from English into Spanish to investigate how different post-editing modalities affect MTUX and translation performance. Participants engaged with both traditional post-editing (TPE) and interactive post-editing (IPE) tasks (consult Section 2 for a detailed description of TPE and IPE), allowing us to systematically compare MTUX, translation productivity, and translation quality across modalities. The study addresses the following overarching research question (RQ):

- **Overarching RQ:** Is IPE a better alternative to TPE in terms of MTUX, translation productivity, and translation quality?
- This overarching RQ is explored through five specific sub-questions:
- **RQ1**: Is MTUX statistically significantly impacted by MT post-editing modality (TPE or IPE), and does this vary with increased experience?
- **RQ2**: Is translation productivity statistically significantly impacted by MT post-editing modality (TPE or IPE), and does this vary with increased experience?
- **RQ3**: Is fluency statistically significantly impacted by MT post-editing modality (TPE or IPE), and does this vary with increased experience?
- **RQ4**: Is adequacy statistically significantly impacted by MT post-editing modality (TPE or IPE), and does this vary with increased experience?
- **RQ5**: Does MTUX correlate with fluency, adequacy, or productivity?

By answering these questions, this study contributes to the broader shift towards human-centered, sociotechnical innovation in MT research, laying the foundations for more sustainable, empowering, and user-aligned MT technologies and workflows.

2. Traditional post-editing (TPE) vs interactive post-editing (IPE)

The integration of MT into professional workflows has traditionally focused on TPE, whereby translators correct static MT output (O'Brien, 2022). TPE quickly became the dominant post-editing modality in the language services industry, often driven by a focus on increasing productivity and reducing costs (ELIS Research, 2023). However, the UX implications of TPE, particularly translator fatigue, frustration, and loss of agency, have been increasingly highlighted in the literature (Cadwell et al., 2018; Läubli & Green, 2019; Jiménez-Crespo & Rodríguez, 2025).

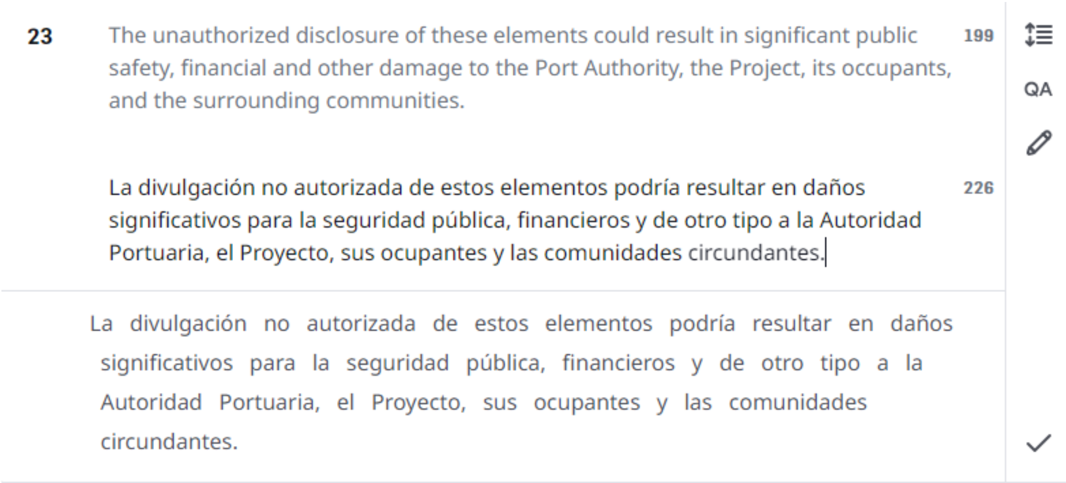

Figure 1. Graphic user interface of Lilt in the TPE modality

Figure 1 shows the graphic user interface of Lilt in the TPE modality with one of the sample texts used. In the screenshot, the translator was editing segment 23, and the MT completion proposal was already fully propagated in the target segment. Therefore, the translator had to conduct a TPE task by amending the static, adaptive MT output.

As a response to the limitations of TPE, IPE was proposed as an alternative modality. In IPE, the MT system provides adaptive translation suggestions in real time, updating predictions based on the user's edits instead of offering a static full-text output. This interaction model theoretically repositions the translator in a supervisory role, offering greater control, reduced cognitive friction, and potentially more positive UX (Alabau et al., 2016; Sánchez Torrón, 2017).

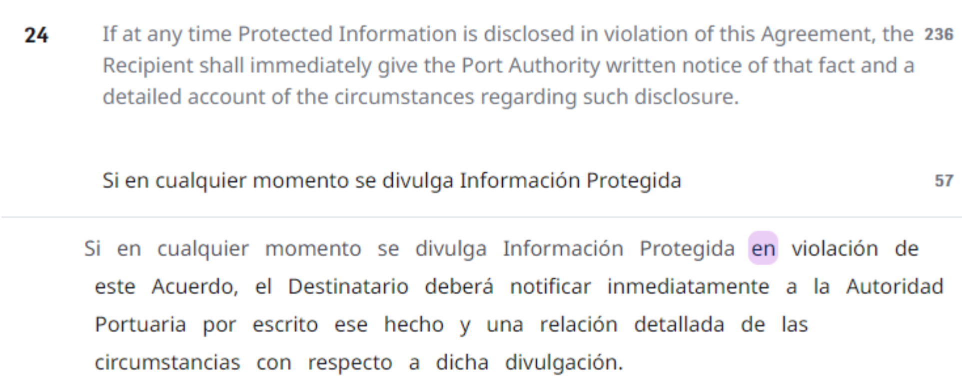

Figure 2. Graphic user interface of Lilt in the IPE modality

Figure 2 shows the graphic user interface of Lilt in the IPE modality with a different sample text. In this screenshot, the translator was editing segment 24 and could see the MT completion proposal, which they could accept partially or completely with specific hotkeys. The word "en" highlighted in purple was the next completion proposal that the system was suggesting at a word-level, which the translator could accept with a hotkey. Should the translator think that this word was not appropriate, and start writing an alternative, the MT completion proposal would change in real time based on the addition made by the translator. Thus, the translator had to conduct an IPE task by amending the interactive, adaptive MT output.

Initial explorations into IPE began with the TransType 1 project (Foster et al., 1997; Langlais et al., 2000b), which introduced word and multi-word completions based on statistical MT (SMT). Although early simulations suggested up to 66% keystroke savings with IPE, real-world user studies revealed that translators' productivity actually declined by up to 35% compared to unaided typing. Translators perceived that they worked faster with IPE, highlighting a divergence between subjective perceptions and objective productivity measures. Further advancements occurred with TransType 2 (Macklovitch, 2006), which involved a series of evaluations. Productivity gains ranging from 12.5% to 20% were observed compared to unassisted translation. Yet, it is worth stressing that IPE was only compared with unassisted translation here and not with TPE.

The CAITRA project (Koehn, 2009) marked a turning point as the first study to directly compare TPE and IPE. Although TPE was found to be 39% faster, participants rated IPE more positively in terms of user satisfaction. The CASMACAT project (Alabau et al., 2013; 2016; Sanchis-Trilles et al., 2014) further expanded this line of inquiry. Early reports emphasised user satisfaction rather than productivity. Subsequent field trials confirmed that while TPE offered higher productivity, IPE reduced keystroke activity and produced comparable or slightly superior translation quality in terms of edit distance. A longitudinal study within CASMACAT demonstrated that translators' productivity with IPE improved over time, with projections indicating that it could surpass TPE productivity between weeks 9 and 10 of continuous use. However, these projections were only hypothetical and there was no empirical data that supported them, with TPE productivity proving superior in the studies conducted.

The emergence of adaptive MT systems fundamentally changed the potential of IPE. Lilt (Green, 2016; Kovacs, 2020) exemplified this shift. Internal studies demonstrated that although IPE was initially slower than TPE by approximately 18-22%, translators' productivity improved during sessions, but no additional data on IPE benefits was shared. Complementary studies focused not only on productivity but also on cognitive and emotional dimensions. Alves et al. (2016) found that while total post-editing time did not differ significantly between TPE and IPE, IPE involved more frequent but shorter eye fixations, suggesting lower cognitive load. Similarly, Daems & Macken (2019) demonstrated that although adaptive NMT systems required fewer edits than adaptive SMT systems, product-based effort metrics (such as TER and CharacTER) did not always align with real cognitive or technical effort, a finding that cautioned against simplistic interpretations of automatic evaluation scores.

In a significant contribution, Sánchez Torrón (2017) conducted a longitudinal study comparing TPE and IPE using an NMT engine within CASMACAT. Over several weeks, IPE achieved an average 2% productivity improvement over TPE, alongside lower technical effort indicators. While quality was generally comparable, IPE exhibited slightly more fluency errors, likely attributable to differences in workflow dynamics. Additionally, translators rated IPE more positively and indicated a growing preference for interactive assistance over static post-editing.

The reviewed studies consistently indicate that while IPE holds significant potential for enhancing translation workflows, its advantages typically become evident only after an initial period of user familiarisation, and no clear empirical demonstration of IPE surpassing TPE has been produced. Many earlier experiments were constrained by technological limitations, such as high system latency, lower prediction accuracy, and suboptimal MT output quality due to SMT, if compared with the capabilities of current NMT systems (Pérez-Ortiz et al., 2022). Recent technological advancements suggest that IPE could now represent a more viable and efficient alternative to TPE than was previously possible. However, this proposition has yet to be empirically validated. It is therefore essential to systematically evaluate whether, under current technological conditions, IPE can surpass TPE in terms of MTUX, translation productivity, and translation quality — an objective that this study seeks to address.

3. Human-centered, augmented machine translation (HCAMT)

In this context of MT progress, it is important to stress that human-computer interaction (HCI) research has long emphasised the critical importance of placing users at the center of technological development. Rather than treating users as passive recipients of technological outputs, HCI advocates for the design of systems that enhance human capabilities, offer intuitive control, and foster safe, trustworthy, and empowering experiences (Shneiderman, 2022a; 2022b). Central principles of HCI — such as usability, UX, and supervisory control — have been developed to ensure that technologies serve human goals rather than replace human agency. Usability has traditionally been evaluated through metrics such as efficiency, effectiveness, and satisfaction (ISO, 2018), whereas UX has expanded the focus to encompass users' emotional, aesthetic, and value-laden interactions with systems (Hassenzahl, 2008).

In Translation Studies, however, the application of HCI concepts has been relatively limited. Early research into translator-computer interaction focused predominantly on usability, typically operationalised through productivity measures (e.g. speed of translation) and error rates (O'Brien, 2006; Carl et al., 2016). While important, this narrow framing left aside richer dimensions of UX, such as empowerment, control, satisfaction, and emotional responses to technology use. Only recently have some scholars started to explore more HCI concepts more explicitly in the translation context, such as usability (Torres-Hostench et al., 2017; Doherty & O'Brien, 2014; Rossetti, 2019; Wang et al., 2021) or acceptability (Castilho, 2016), but this body of research remains scarce and often lacks a systematic integration of validated UX frameworks developed within HCI. Translation Studies has briefly addressed UX, but research has been limited (e.g. Guerberof Arenas et al., 2021; Koponen et al., 2020; Karakanta et al., 2022), with very few studies embedding robust, longitudinal UX evaluations into translation technology assessments.

Today's language services industry is inseparable from HCI: digital tools are now embedded in commercial translation production workflows, and translators are expected to use these tools to remain competitive (ELIS Research, 2025). However, much MT research and development has prioritised productivity and output quality in support of increasing automation, paying comparatively less attention to how users experience and manage these systems. This has had serious repercussions for adoption: studies report growing rejection of TPE (Torres-Hostench et al., 2016; Macías, 2020), experiences of dehumanisation and the commodification or "uberisation" of translation work (Fırat, 2021), and non-adoption of post-editing in some settings (Cadwell et al., 2018), alongside increased fears about MT and AI (ELIS Research, 2025) and the use of opaque algorithmic mechanisms to manage translators' participation (Moorkens, 2023). These patterns can be understood as manifestations of "technological adaptation" (Briva-Iglesias, 2024). By technological adaptation, we refer to the process whereby new technologies are designed, developed, and deployed without systematically considering the needs, experiences, and agency of their future users. Instead of developing technologies that are tailored to user requirements, users are subsequently forced to adapt themselves to pre-established technological systems. This adaptation is often reactive, individualised, and lacking in systemic support, leading to experiences of frustration, alienation, and resistance among technology users (Winner, 2007; Vallor, 2024). In the translation context, technological adaptation has contributed to translators' limited acceptance of post-editing workflows and to wider concerns about the erosion of professional autonomy and well-being.

It is at this stage that we raise the question of whether the process should be the opposite: first, we should understand what users need from technology, and then we should develop new technologies to meet these needs and augment users.

Hence, we consider that the translation technology and MT communities should re-visit their technology design, development, and adoption focus, and should shift their attention from the translation productivity- and quality-first approach towards a human-centered, augmented machine translation (HCAMT) approach. This HCAMT approach is built on the basis of the HCAI framework (Shneiderman, 2022a; 2022b). It is important to note that the HCAMT approach and the focus on translation quality and productivity do not have to be mutually exclusive; rather, they should be applied together in a complementary manner. HCAMT technologies should be developed considering the needs of their users as a central point, but also their productivity and quality if applied to different translation production workflows. Given the substantial variation among MT users (Nurminen, 2021), it is clear that covering all types of users is beyond the scope of this paper. Therefore, the overarching aim of this research is to lay the basis of a new methodology to foster the development and adoption of HCAMT tools, systems, and workflows, using today's language services industry as a specific use case. By narrowing the scope of the work to this particular application, the present research explores whether IPE may be a better alternative to TPE by analysing MTUX, translation productivity, and translation quality.

In this paper, we deliberately use the term HCAMT rather than the broader label "human-centered machine translation" (HCMT) recently proposed in the MT literature (Carpuat et al., 2025). While HCMT helpfully signals a move away from purely system-centric optimisation, "human-centered" alone does not guarantee that augmentation remains foregrounded as a design and evaluation principle (Shneiderman, 2022a; 2022b; O'Brien, 2023). Given the long-standing dominance of automation-first narratives in MT, and the fact that not all readers will be familiar with HCAI's emphasis on augmentation, we make this dimension explicit in HCAMT as a conceptual anchor: MT systems should be conceived and assessed in terms of the extent to which they expand humans' capacity to understand, decide, and act in multilingual settings, rather than simply maximising output volume or minimising costs. Crucially, it is not the MT system that is "augmented", but the human users — whether professional translators in TPE/IPE workflows or non-professional users relying on MT for assimilation and communication in domains such as healthcare (Briva-Iglesias & Peñuelas Gil, 2025), academia (Bowker, 2023) or legal settings (Nurminen, 2021), to name just a few of the many potential MT users.

Building upon this gap, the concept of MTUX has been introduced to specifically capture how users perceive and interact with MT systems (Briva-Iglesias & O'Brien, 2023). MTUX refers to the perceptions and responses (both cognitive and affective) that arise from users' interactions with MT, whether anticipated or actual. Crucially, MTUX considers not only the quality of the translation output but also the overall quality of the interaction process, including the degree of user control, the system's responsiveness, the intuitiveness of the interface, and the emotional engagement of the user. Evaluating MTUX thus requires moving beyond traditional quality metrics and incorporating validated UX assessment tools such as the User Experience Questionnaire (UEQ) (Schrepp et al., 2014).

Drawing on the foundations of HCAI, the HCAMT paradigm thus redefines the role of MT systems: instead of aiming to replace professional translators, systems should be designed to augment translators' capabilities, enhancing their productivity, creativity, satisfaction, and sense of agency (O'Brien, 2023). In an HCAMT framework, the human remains in supervisory control at all times, while the machine dynamically adapts to support human intentions and goals. Operationalising HCAMT requires prioritising MTUX as a central success metric alongside traditional measures of translation quality and productivity. Furthermore, robust evaluation of HCAMT technologies must embed UX research methods — such as longitudinal designs, pre- and post-task perception surveys, and fine-grained behavioural analyses — into the assessment of MT workflows.

Designing for HCAMT use and adoption should become the guiding principle for future innovation in MT, regardless of the type of MT users involved, be they professional translators or lay users. By focusing on augmenting human capabilities rather than merely increasing automation, MT developers and researchers should be able to foster technologies that are not only more effective but also more ethical, sustainable, and widely accepted across diverse user groups. A shift towards HCAMT is critical if MT is to continue evolving as a supportive, empowering tool that enhances human communication rather than dehumanising professional translation practice.

4. Methodology

This paper employed a controlled, two-week longitudinal design with 11 professional translators to compare TPE and IPE in terms of MTUX, translation productivity, and translation quality. Translators were instructed to perform a professional translation, ready for dissemination, and they worked remotely from their usual workplaces, connecting to a Dublin City University computer via AnyDesk, therefore simulating their normal working conditions to get the most valid and real-life data. The longitudinal nature of the study allowed us to observe not only immediate user responses but also the evolution of behaviours, perceptions, and performance over time (Caruana et al., 2015; Diggle et al., 2002). Such a design is particularly suited to HCI research, where user adaptation and learning effects are expected.

4.1. Participants

Eleven professional translators were recruited for the study. Participants were selected based on the following criteria: they had between one and five years of full-time professional translation experience, Spanish as their native language, and professional expertise in the legal domain, which was the subject matter of the translation tasks. Recruitment was conducted through professional platforms (ProZ and other translator networks) following a first-come, first-served procedure, provided that the recruitment criteria were met.

Participants' professional translation experience ranged from 12 to 48 months (M = 29 months, SD = 12). None of the participants had prior experience with IPE systems, but all of them had experience with TPE, ranging from one to 24 months (M = 10 months, SD = 8). All translators were compensated at professional rates. In addition, three senior reviewers with more than five years of professional experience were hired to conduct translation quality assessments. To ensure consistency, evaluation guidelines were developed, and high inter-annotator agreement was achieved before the main evaluation was carried out by a single expert reviewer.

4.2. IPE workbench

The IPE workbench selected for the study was Lilt, a proprietary tool that supports both TPE and IPE tasks. Although open-source alternatives were considered, Lilt offered several advantages that justified its selection. To ensure valid, industry-aligned results, it was essential to work with a system that provided high-quality MT output for the English-Spanish language pair. In preliminary evaluations, Lilt's raw MT output achieved an average adequacy score of 3.4/4 and a fluency score of 3.65/4, as assessed by three reviewers. Moreover, Lilt allowed the interactive translation completion feature to be toggled on or off, enabling controlled implementation of both TPE and IPE modalities within the same environment.

Collaboration with Lilt provided access to an academic licence, allowing the study to proceed without needing to train a custom MT system, thus reducing costs and environmental impact (Zhong et al., 2023). To maintain strict experimental control, a new custom MT model ("Custom model") and a new translation memory ("Data source", according to the platform) were created within Lilt for the English-Spanish combination for each of the translation tasks for each of the translators. This ensured that all participants received identical MT proposals across both modalities and that translation tasks were consistent throughout the study.

4.3. Texts

Complex English legal contracts were the texts chosen for our controlled study. Each translator worked with 13 different texts, under different conditions (10 in TPE and three in IPE), and we randomly divided the assignments, ensuring that the combination of text and modality was counterbalanced across the experiment. All texts were controlled for length and complexity with the Flesch-Kincaid index and type-token ratio (TTR). TTR is the total number of unique words (types) divided by the total number of words (tokens) in a given segment of language. This indicates text complexity. The higher the number of unique words, the higher the complexity of the text, and the lower the TTR. The contracts used here were highly complex texts from the legal domain, with an average TTR of 0.29. Table 1 summarises information about the texts used in the longitudinal study.

	No. of Words	Flesch-Kincaid	TTR
text 1	1189	21.7	0.29
text 2	1102	26.5	0.28
text 3	1149	25.1	0.28
text 4	1032	26.9	0.33
text 5	1195	24.3	0.32
text 6	1022	27.9	0.3
text 7	1043	24.3	0.26
text 8	1098	26.3	0.3
text 9	1117	20.1	0.29
text 10	1103	27.2	0.29
text 11	1071	33.1	0.29
text 12	1094	27.4	0.3
text 13	1160	26.1	0.3
avg	1106	26	0.29
total words per translator (total no. of words)	14,375 (158,125)		

Table 1. Characteristics of the texts used in the longitudinal study

The final word count, combining the 13 texts, was 14,375 words per translator (giving a combined total of 158,125 words from the 11 translators). Nevertheless, some translators did not finish translating the complete text during the 45-minute period granted, and we therefore ended up with 120,102 translated words. Every translator worked for approximately 9.75 hours (107.25 hours in total).

If compared with all the previous IPE studies reviewed in Section 2, our study involves one of the biggest samples of TPE and IPE research in terms of translating time over a longitudinal period, the number of words translated, and the number of professional translators hired. The TransType 1 study only worked with 10 translators (four professionals and six students) for 20 minutes (Langlais et al., 2000a). TransType 2 hired six professional translators, who worked for 10 days and translated around 20,000 words (Macklovitch, 2006). The evaluation of CAITRA involved 10 students who worked on 5,000 words, but some of them were non-native speakers of the languages they were working with (Koehn, 2009). The longitudinal study of CASMACAT involved five professional translators who produced around 24,000 words over two weeks (Alabau et al., 2016). CASCAMAT's third field trial hired seven translators who worked only two days and produced around 9,000 words (Ibid.). Alves et al. (2016) worked with 16 professional translators who worked with only 36 segments. Daems & Macken (2019) hired four translators who worked with 20 segments. To facilitate replicability, or encourage further analysis of this large dataset, the source and translated texts can be found at https://zenodo.org/records/10696772.

4.4. Study design

The 11 translators participated in a two-week longitudinal study, working from Monday to Friday over 10 consecutive days, excluding weekends (see Figure 3 for an overview). Before beginning translation tasks, participants completed a five-minute pre-task questionnaire to gather information about their previous experiences with and perceptions of MT and MT post-editing (https://tally.so/r/mRX0Kw). Subsequently, they connected remotely to an encrypted computer at Dublin City University via AnyDesk and engaged in a 25-minute warm-up session, during which they read instructions for using Lilt, familiarised themselves with hotkeys for the IPE modality, and practised with a sample project.

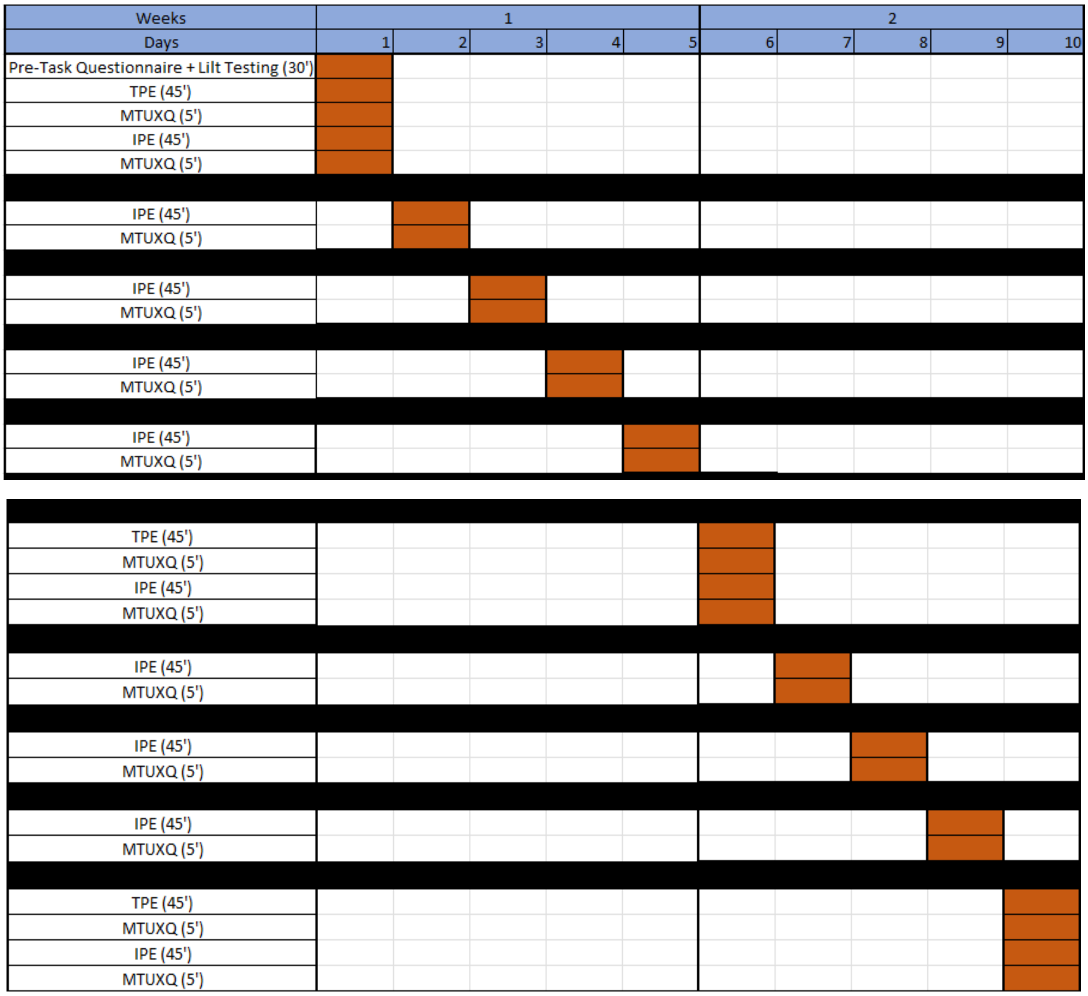

Figure 3. Design of the controlled longitudinal study

The structure of the study was as follows: on the first day of each week (Monday), translators completed a 45-minute TPE/IPE session followed by a five-minute MTUX questionnaire, then a 45-minute IPE/TPE session followed by another MTUX questionnaire. The starting modality was randomised to control for order effects. From Tuesday to Friday each week, translators completed one 45-minute IPE session per day, each followed by a five-minute MTUX questionnaire. The final day of the study (Friday, week 2) replicated the double-session structure of the first day, again with randomised starting modality.

Each translator therefore completed 10 IPE sessions and three TPE sessions. Sessions involving both TPE and IPE (days 1, 6, and 10) are referred to as "evaluation sessions" and enabled direct comparison between modalities. The 10 IPE sessions are referred to as "learning sessions" and allowed for the analysis of longitudinal changes in MTUX, translation quality, and productivity as translators gained experience with IPE.

Although the number of TPE and IPE sessions would ideally have been balanced, budgetary constraints made this impractical. Given that participants had less prior experience with IPE, a greater number of IPE interactions were prioritised to compensate for this difference in expertise, while TPE sessions were strategically distributed across the beginning, middle, and end of the study to serve as comparative baselines.

4.5. Measures

4.5.1. Machine translation user experience (MTUX)

Translators completed a self-report UX questionnaire after performing each post-editing task, resulting in 10 measures of MTUX for IPE and three measures of MTUX for TPE. As per the work reported in Briva-Iglesias & O'Brien (2023), we used the English version of the UEQ (Laugwitz et al., 2008). The UEQ is a validated questionnaire that measures UX and is commonly used in the field of HCI (Schrepp et al., 2014; 2017). It is a 26-item semantic differential scale in which each adjective pair (e.g. Annoying–Enjoyable, Impractical–Practical, Slow–Fast) is scored on a seven-point scale. The order in which the 26 items were displayed was randomised, with positive and negative poles for each item alternated to avoid any confounding order effects or response acquiescence (see Appendix 1).

4.5.2. Translation productivity

Translation productivity was measured within Lilt, while translators were performing the MT post-editing (MTPE) tasks, by recording the number of words translated per hour (WPH).

4.5.3. Translation quality

Translation quality was assessed via a four-point Likert scale human evaluation of adequacy and fluency. Although the literature recommends using multiple evaluators to minimise subjectivity when assessing translation quality (Guerberof Arenas, 2008; Rossi & Carré, 2022), it was not feasible to hire several reviewers to evaluate the full dataset (120,102 words). Instead, we adopted a methodology common in computer science to mitigate individual bias: an initial homogenisation of evaluation criteria across multiple annotators, followed by escalation to a single expert reviewer once high inter-annotator agreement (IAA) was achieved (Artstein & Poesio, 2008). Three professional reviewers were first hired to annotate 50 segments using the initial guidelines. This produced an IAA of 0.83 (Artstein, 2017). After a Zoom meeting to resolve inconsistencies and update the guidelines, a second annotation round was conducted, achieving an IAA of 0.95. Following this, one expert reviewer, selected based on this process, evaluated all translations (120,102 words) and raw MT outputs (14,734 words), taking full document context into account (Castilho, 2021). To validate consistency, 250 random segments were re-annotated by the two other reviewers midway through the evaluation, yielding an IAA of 0.88. This confirmed the robustness of the quality evaluation process based on the final annotation guidelines. The annotated data can be found in Zenodo: https://zenodo.org/records/10696772.

5. Results

This section describes the analyses performed with the data collected during the longitudinal study. It is divided into subsections that correspond to the different RQs of the paper.

5.1. RQ1. Is MTUX statistically significantly impacted by MTPE modality (TPE or IPE), and does this vary with increased experience?

The longitudinal study design allowed us to collect 10 MTUX measures for IPE and three MTUX measures for TPE. MTUX scores were collected through a seven-point Likert scale, but results were normalised to scores ranging from −3 (very bad experience) to +3 (very good experience). Although the MTUX scores could range from -3 to +3, we are only visualising the 0 to +3 range because there were no negative values. We performed different 2x3 repeated-measures ANOVAs to analyse the effect of MTPE modality (levels: TPE and IPE) and interaction session (levels: Interaction 1, Interaction 6, and Interaction 10) on MTUX scores by considering the data from the evaluation sessions.

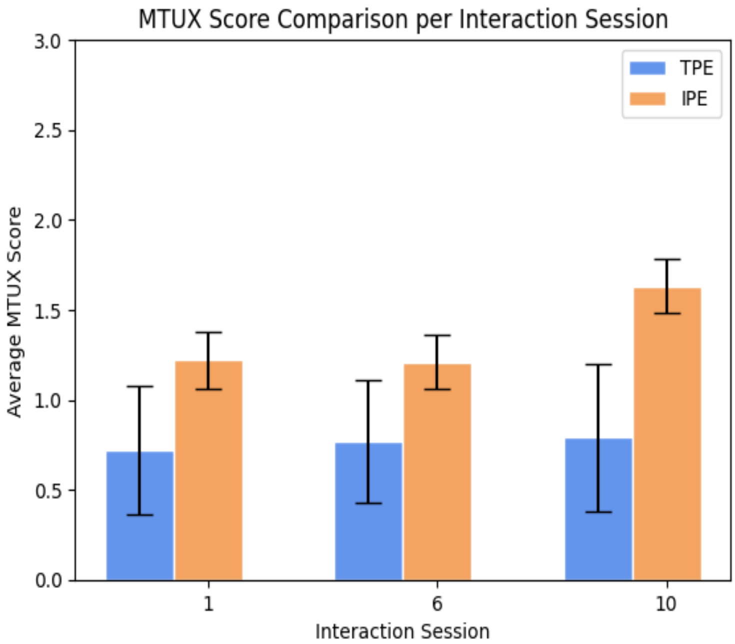

Figure 4. MTUX score comparison in the evaluation sessions (with SD bars)

Figure 4 visualises the average MTUX scores in the different MTPE modalities across the different evaluation sessions analysed. The repeated-measures ANOVA results indicated that there was a statistically significant main effect of MTPE modality on MTUX (F (1, 10) = 9.91, p = 0.01), with a statistically significantly higher MTUX score when post-editing using an IPE workflow (M = 1.24, SD = 0.16) compared to when using a TPE workflow (M = 0.72, SD = 0.36). This means that translators perceived that their interactions with MT through the IPE modality improved their UX substantially relative to TPE.

The repeated-measures ANOVA also suggests that there was a statistically significant difference in MTUX scores between the different interaction sessions across the MTPE conditions ($F\ (2,\ 20) = 4.29$, $p = 0.04$). This suggests that translators reported higher MTUX scores with increased experience with the tool (Interaction 1: TPE M = 0.72, SD = 0.36; IPE M = 1.22, SD = 0.16/ Interaction 6: TPE M = 0.77, SD = 0.34; IPE M = 1.21, SD = 0.15/ Interaction 10: TPE M = 0.79, SD = 0.15; IPE M = 1.63, SD = 0.15). However, the ANOVA results show that there was no statistically significant interaction effect between interaction session and MTPE modality ($F12,\ 120 = 2.78$, $p = 0.09$). This means that the effect of interaction session does not vary across the MTPE modalities.

5.2. RQ2. Is translation productivity statistically significantly impacted by MTPE modality (TPE or IPE), and does this vary with increased experience?

With each TPE or IPE interaction, we tracked translation time with Lilt, and we measured the number of words translated per hour (WPH) in each of the translation sessions, both for TPE and IPE. Again, a 2x3 repeated-measures ANOVA was conducted to assess whether there was a statistically significant effect of MTPE modality (levels: TPE and IPE) or interaction session (levels: Interaction 1, Interaction 6, and Interaction 10) on average translation productivity by considering the data from the evaluation sessions.

We found that the main effect of interaction session on translation productivity was not statistically significant ($F\ (2,\ 20) = 3.38$, $p = 0.05$). This suggests that when we only look at the number of interaction sessions, this factor does not seem to make a real difference in terms of productivity. This makes sense, since participants already have professional experience in TPE and, therefore, may have already attained their maximum productivity speed in this MTPE modality (this is further supported by Figure 5, where we can see that productivity in TPE is flat).

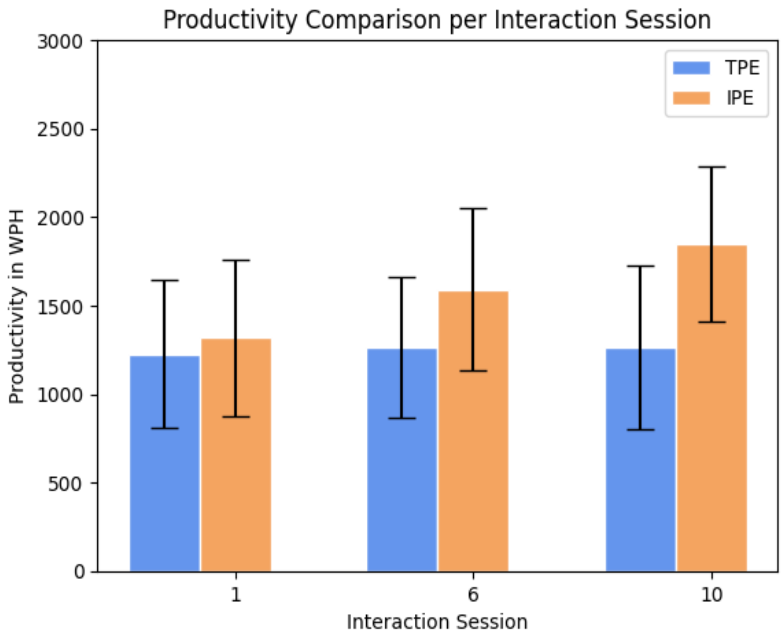

Figure 5. Productivity comparison in the evaluation sessions (with SD bars)

We found a statistically significant effect of MTPE modality on productivity (F (1, 10) = 19.63, p = 0.001). This means that the MTPE modality used has a clear impact on productivity. By looking at Figure 5, we can see that participants worked statistically significantly faster in the IPE modality (M = 1578.85; SD = 935) than in the TPE modality (M = 1359.28; SD = 1004.74).

Lastly, ANOVA results also suggest that there was a statistically significant interaction effect between the two independent variables (F (2, 20) = 16.56, p = 0.0001). This interaction effect suggests that the effect of MTPE modality varies with experience, with improvements in productivity within the IPE condition over time and TPE performance staying relatively stable across the sessions.

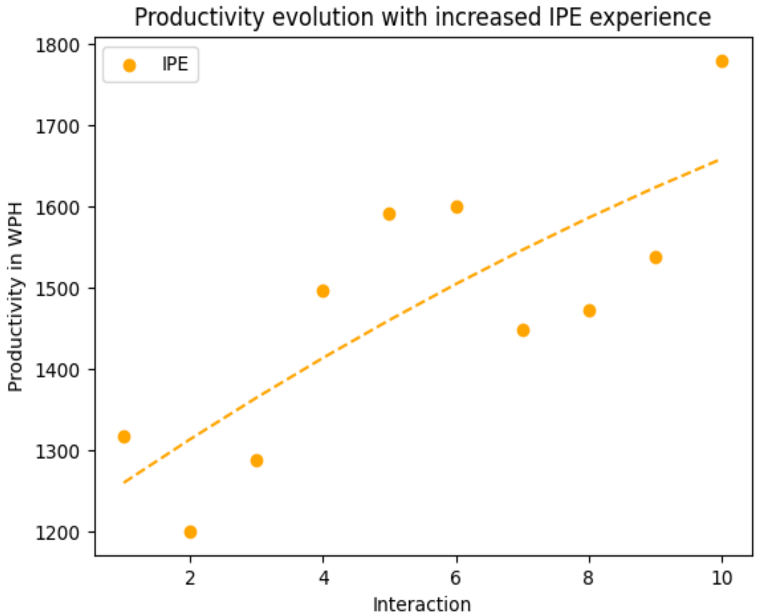

Figure 6. Productivity evolution over the learning sessions

Observing the learning sessions (see Figure 6), over a span of 10 interactions, the participants demonstrated a consistent increase in productivity, as measured in WPH. The graph illustrates this ascent from an initial average productivity of 1,317 WPH to nearly 1,800 WPH. This pattern underscores a significant enhancement in translation productivity, correlating positively with the participants' growing familiarity with the IPE functionalities. These results highlight the effectiveness of IPE tools in optimising the translation workflow, indicating that such technologies can substantially elevate productivity levels when users are given the opportunity to adapt to this new MTPE modality.

5.3. RQ3. Is fluency statistically significantly impacted by MTPE modality (TPE or IPE), and does this vary with increased experience?

Again, a two-way repeated-measures ANOVA was conducted to assess the impact of interaction session and MTPE modality on the fluency scores of participants. The results indicated a non-statistically significant effect of interaction session on fluency scores (F

(2, 20) = 1.79, p = 0.20). This suggests that there was no substantial variance in fluency scores that could be attributed to the interaction session. That is, when participants became more acquainted with the tool, their fluency score did not improve (see Figure 7).

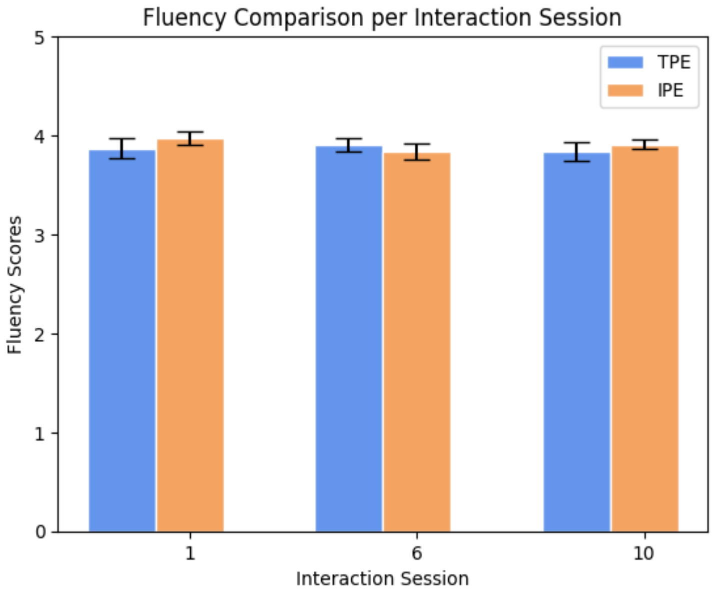

Figure 7. Fluency comparison in the evaluation sessions (with SD bars)

By contrast, the main effect of MTPE modality on fluency was statistically significant (F (1, 10) = 9.80, p = 0.01), indicating that MTPE modality alone did statistically significantly influence fluency outcomes. By observing the averages in both MTPE modalities, we can see that fluency scores in IPE (M = 3.89; SD = 0.09) were statistically significantly higher than fluency scores in TPE (M = 3.87; SD = 0.09).

The interaction effect between interaction session and MTPE modality was also statistically significant (F (2, 20) = 10.19, p = 0.001), which implies that the effect of interaction session on fluency scores is modulated by MTPE modality. In essence, while increased experience alone does not change fluency scores, the MTPE modality has a statistically significant effect on fluency scores. In addition, the two factors together (interaction session and MTPE modality) have a significant impact on fluency. That is, the effects of experience vary for each MTPE modality.

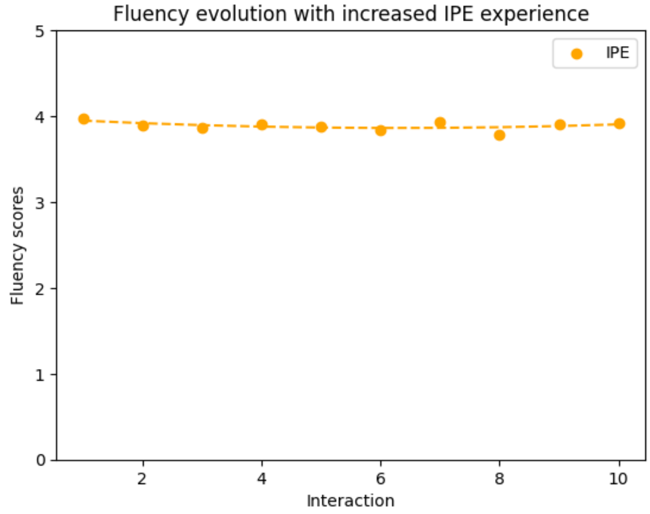

Figure 8. Fluency evolution over the learning sessions

Figure 8 illustrates that fluency scores, anchored firmly around the score of 4, remained remarkably consistent throughout the 10 learning sessions. This sustained level of fluency suggests that fluency scores did not change with increased experience with IPE. It is also worth stressing that fluency scores were close to the maximum of 4, affirming the viability of IPE for professional translation tasks, where fluent and natural-reading translations are paramount.

5.4. RQ4. Is adequacy statistically significantly impacted by MTPE modality (TPE or IPE), and does this vary with increased experience?

The statistical analysis of the influence of the interaction session and the MTPE modality on the adequacy scores of the translations produced by the different participants was also assessed through a 2x3 repeated-measures ANOVA (Figure 9).

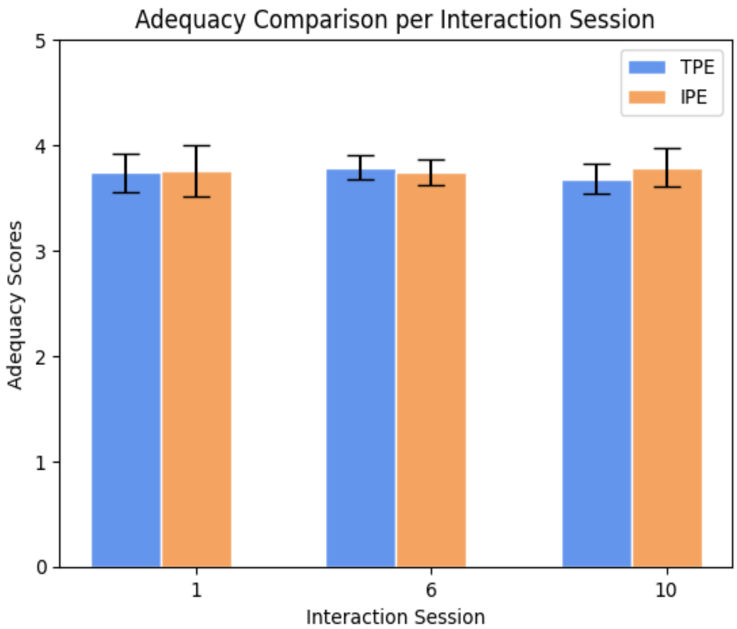

Figure 9. Adequacy comparison in the evaluation sessions

The analysis did not indicate a statistically significant main effect of interaction session on adequacy scores (F (2, 20) = 0.31, p = 0.71), suggesting that increasing interaction session alone did not substantially enhance the adequacy of translations. Similarly, the main effect of MTPE modality on translation adequacy was not statistically significant (F (1, 10) = 0.77, p = 0.40), indicating that the choice of MTPE modality by itself did not statistically significantly influence the adequacy of the translations. Although the adequacy scores of the IPE modality (M = 3.76; SD = 0.19) were higher than the adequacy scores of the TPE modality (M = 3.74; SD = 0.15), no statistically significant difference could be observed, as reflected in Figure 9.

The interaction between interaction session and MTPE modality was also not statistically significant (F (2, 20) = 1.85, p = 0.19). This implies that the potential effect of interaction session on translation adequacy did not vary with the different MTPE modalities. Collectively, these results suggest that neither increased experience with a system nor the MTPE modality used significantly impacts the adequacy of translations.

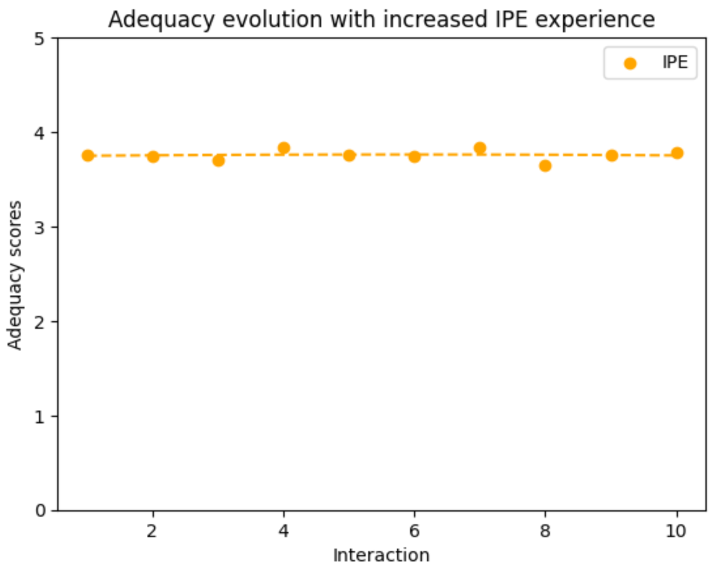

Figure 10. Adequacy evolution over the learning sessions

Figure 10 reflects the adequacy evolution across the 10 learning sessions. Adequacy scores consistently hovered around the maximum (4), indicating a stable performance in maintaining the integrity of the source content throughout the interactions. It is significant to note, however, that these scores were modestly lower than those obtained for fluency in the section above. This applies both to TPE and IPE: although IPE scores in terms of adequacy and fluency are generally higher than TPE scores, both TPE and IPE adequacy scores report slightly lower global scores than fluency. This discrepancy emphasises a critical nuance in translator-MT interactions: while fluency may be more readily achieved with the aid of state-of-the-art NMT systems, ensuring the translation's adequacy — a measure of how well the source message is preserved — may pose a greater challenge, and may depend more on the translator.

5.5. RQ5. Does MTUX correlate with fluency, adequacy, or productivity?

We also wanted to analyse whether MTUX scores had any statistically significant relationship with translation quality or productivity. Thus, different variables were checked for correlation, namely translators' overall MTUX scores in the TPE and IPE conditions with measures of translation adequacy, fluency, and productivity. First, we checked data normality with Shapiro-Wilk's test. MTUX scores, both in the TPE and the IPE modalities, followed a normal distribution (p-value > .05). Adequacy scores in IPE were also normally distributed. The other variables (i.e. productivity scores in TPE and IPE, fluency scores in TPE and IPE, and adequacy scores in TPE) did not follow a normal distribution (i.e. they violated the assumption for normality (p-value < .05)). We therefore performed a Kendall's T correlation test for every non-parametric variable, except when we correlated the normally distributed variables, where we conducted a Pearson's correlation test. In conclusion, we found there were no statistically significant correlations between MTUX scores and final translation quality. MTUX scores in the TPE condition [adequacy: r (32) = -0.04, p = 0.83; fluency: r (32) = 0.05, p = 0.67] and in the IPE condition [adequacy: r (109) = 0.11, p = 0.08; fluency: r (109) = 0.08, p = 0.19] did not significantly correlate with final translation quality. Regarding participants' productivity, MTUX scores in the TPE condition [r (32) = 0.08, p = 0.295] and in the IPE condition [r (109) = −0.09, p = 0.14], did not significantly correlate with measures of translator productivity.

6. Discussion

This study set out to investigate whether IPE could constitute a better alternative to TPE in terms of MTUX, translation productivity, and translation quality. By conducting a two-week controlled longitudinal study with professional translators, we were able to assess not only immediate interaction effects but also changes over time as participants gained experience with an IPE system. The results offer strong evidence that a shift towards interactive, HCAMT workflows can yield meaningful benefits, particularly when measured through the lens of MTUX and productivity, without compromising translation quality.

First, regarding MTUX, the results demonstrated that translators consistently reported higher UX scores when working with IPE compared to TPE in our use case. Importantly, MTUX scores increased with experience in the IPE modality, highlighting the positive impact of familiarity and system acclimatisation. These findings are particularly significant because they validate a central claim of the HCAMT approach: that enhancing the user's sense of control, responsiveness, and satisfaction in translation workflows is not only possible but measurable over time. In contrast, MTUX scores in TPE remained comparatively stagnant, suggesting that the traditional model fails to adapt to user needs in ways that foster positive engagement.

Second, the productivity findings reinforce the potential of IPE systems. While no significant differences were observed between modalities at the outset, participants' translation speed in the IPE condition improved substantially across the study, surpassing their productivity in the TPE condition by the final evaluation sessions. These gains were achieved after a relatively short learning curve, aligning with previous observations in HCI that experience and familiarity are crucial factors in maximising the benefits of interactive technologies (Albarracín, 2021; Albarracín & Wyer, 2000). Moreover, the productivity improvements in IPE were not achieved at the expense of translation quality, supporting the notion that IPE systems can enhance human performance without compromising output quality.

Third, in terms of translation quality, fluency scores were statistically significantly higher for translations produced under IPE conditions compared to TPE. Adequacy scores remained high and comparable between both modalities, with no statistically significant differences detected. These results suggest that the interactive features of IPE systems, which allow users to incrementally guide MT output, may contribute to smoother, more natural translations. In this respect, the IPE modality effectively supports the dual goals of MT quality and user empowerment.

Interestingly, no significant correlations were found between MTUX scores and measures of translation quality or productivity. This finding deserves further reflection. It suggests that while better UX is associated with greater subjective satisfaction and improved workflow, it does not necessarily predict objective gains in quality or speed at the individual interaction level. However, previous research on the topic demonstrates that MTUX, translation quality, and productivity are related but distinct dimensions of translator-computer interaction that must be independently considered in the design and evaluation of MT systems (Briva-Iglesias & O'Brien, 2024).

These results have important implications for the future of MT research and practice. They demonstrate that the traditional productivity- and quality-first paradigm, dominant in the development of TPE workflows, is insufficient for building sustainable, empowering translator-computer interactions. The findings strongly support the HCAMT approach proposed in this paper. Rather than positioning users as passive correctors of machine output, HCAMT envisions them as active agents who interact dynamically with adaptive MT systems designed around their needs, preferences, and values.

Moreover, the study provides empirical evidence of the risks of technological adaptation. In the case of TPE, translators have been asked to adapt to workflows optimised for automation and cost reduction, rather than for human agency or experience. The consistent user dissatisfaction and the stagnation in TPE workflows observed in this and prior studies (Cadwell et al., 2018; Fırat, 2021) illustrate the limits of this approach. In contrast, the positive learning curve and engagement observed in IPE interactions underscore the benefits of designing technologies that actively adapt to users, embodying the principles of HCAMT.

7. Conclusions, limitations, and future work

This study has provided empirical support for the HCAMT paradigm by demonstrating that IPE offers clear advantages over TPE in terms of MTUX and translation productivity, without compromising translation quality, at least in our specific use case: legal translation and the English-Spanish language combination. Through a controlled two-week longitudinal study with professional translators, we observed that MTUX scores achieved with IPE consistently outperformed those achieved with TPE, and that productivity gains emerged after a relatively short familiarisation period. These findings suggest that centering human needs, experiences, and agency in MT design is not only conceptually sound but also operationally beneficial. Accordingly, HCAMT emerges as a rigorous, scalable paradigm for human-aligned MT development and deployment.

Furthermore, the study critically highlighted the risks associated with technological adaptation, a process through which users are compelled to adjust to technologies developed without their active participation. The stagnation observed in TPE workflows, coupled with the growing translator dissatisfaction and resistance reported in the literature, illustrates the limitations of automation-centric development models. By contrast, HCAMT offers a more sustainable and empowering path forward, calling for technologies that augment rather than replace human translators (O'Brien, 2023), encouraging more co-creation and engaged research practices in translation and MT studies. When development follows a rigorous understanding of MT users first — and only then builds the tools — workflows are more efficient, interactions more empowering, and satisfaction demonstrably higher. This user-first inversion also makes impact testable: success metrics flow from user needs, enabling longitudinal evaluation and iterative refinement. Crucially, this principle is not confined to professional translators; it can be transferred to all MT and language-technology users wherever language mediates work or access. In short, MT designed from the human outward functions as a tool for augmentation, making it safer, more inclusive, and more accountable — and, consequently, easier to adopt at scale.

Despite these contributions, several limitations must be acknowledged. The study focused exclusively on professional translators with experience in the legal domain and in the English-Spanish language pair. As such, the findings cannot be automatically generalised to other user profiles, language combinations, or translation domains. Additionally, while the longitudinal design allowed for the observation of learning effects, the two-week duration may not fully capture longer-term adaptation and retention patterns. The study also relied on human evaluation for quality assessment, which, although carefully controlled, inherently introduces subjective elements.

As briefly suggested above, future research should widen HCAMT beyond professional translation to the much larger population of MT users. Professional translators are a small minority; estimates suggest that over 99% of users are non-professionals (Nurminen, 2021), ranging from corporate staff and clinicians to migrants, public-service providers, educators, and everyday communicators. Yet, these communities remain underrepresented in MT studies and HCAMT evidence.

Priority tasks include mapping diverse tasks, risk profiles, and success criteria; designing adaptive, context-aware MT that responds to situational constraints; and developing evaluation protocols that capture comprehension, error tolerance, trust calibration, and downstream outcomes. Advancing this agenda requires inclusive, interdisciplinary methods that combine Translation Studies, HCI, sociotechnical systems research, and digital ethics (Schmager et al., 2025). Put simply: if MT is for everyone, we must study — and design with — everyone. Design *with* and *for* the 99%, or MT will stay optimised only for the 1%.

In sum, the results of this study advocate for a fundamental reorientation of MT development priorities. By adopting HCAMT principles, future technologies can better align with human values, foster more positive user experiences, and ensure that the expansion of MT contributes not to the erosion of human agency but to its enhancement across diverse societal contexts.

References

Alabau, Vicent; Bonk, Ragnar; Buck, Christian; *et al.* (2013). CASMACAT: An open source workbench for advanced computer aided translation. *Prague Bulletin of Mathematical Linguistics*, v. 100, n. 1, pp. 101-112. DOI <10.2478/pralin-2013-0016>. [Accessed: 20251219].

Alabau, Vicent; Carl, Michael; García-Martínez, Mercedes; González-Rubio, Jesús. (2016). Learning Advanced Post-Editing. In: Carl, Michael; Bangalore, Srinivas; Schaeffer, Moritz. (eds.). *New Directions in Empirical Translation Process Research*. Cham: Springer, pp 95–110. <https://doi.org/10.1007/978-3-319-20358-4_5>. [Accessed: 20251219].

Albarracín, Dolores (2021). The Impact of Past Experience and Past Behavior on Attitudes and Behavior. In: *Action and Inaction in a Social World: Predicting and Changing Attitudes and Behavior*. Cambridge: Cambrigde University Press. <https://doi.org/10.1017/9781108878357.006>. [Accessed: 20251219].

Albarracín, Dolores; Wyer, Robert S. (2000). The Cognitive Impact of Past Behavior: Influences on Beliefs, Attitudes, and Future Behavioral Decisions. *Journal of Personality and Social Psychology*, v. 79, n. 1, pp. 5-22. <https://doi.org/10.1037//0022-3514.79.1.5>. [Accessed: 20251219].

Alves, Fabio; Sarto Szpak, Karina; Gonçalves; José Luiz; *et al.* (2016). Investigating Cognitive Effort in Post-Editing: A Relevance-Theoretical Approach. In: Hansen-Schirra, Silvia; Grucza, Sambor (eds.). *Eyetracking and Applied Linguistics*. Berlin: Language Science Press, pp. 109-141. <http://langsci-press.org/catalog/book/108>. [Accessed: 20251219].

Artstein, Ron. (2017). Inter-Annotator Agreement. In: Ide, Nancy; Pustejovsky, James (eds.). *Handbook of Linguistic Annotation.* Dordrecht: Springer Netherlands, pp. 297-313. <https://doi.org/10.1007/978-94-024-0881-2_11>. [Accessed: 20251219].

Artstein, Ron; Poesio, Massimo (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, v. 34, n. 4, pp. 555-96. <https://doi.org/10.1162/coli.07-034-R2>. [Accessed: 20251219].

Bowker, Lynne. (2023). *De-Mystifying Translation: Introducing Translation to Non-Translators.* London; New York: Rouledge. <https://doi.org/10.4324/9781003217718>. [Accessed: 20251219].

Briva-Iglesias, Vicent. (2024). *Fostering Human-Centered, Augmented Machine Translation: Analysing Interactive Post-Editing.* [PhD Thesis], Dublin City University. Dublin. <https://doras.dcu.ie/30182/>. [Accessed: 20251219].

Briva-Iglesias, Vicent; O'Brien, Sharon (2023). Measuring Machine Translation User Experience (MTUX): A Comparison between AttrakDiff and User Experience Questionnaire. In: Nurminen, Mary; Brenner, Judith; Koponen, Maarit; *et a*l. (eds.). *Proceedings of the 24th Annual Conference of the European Association for Machine Translation.* European Association for Machine Translation, pp. 335–344. <https://aclanthology.org/2023.eamt-1.33/>. [Accessed: 20251219].

Briva-Iglesias, Vicent; O'Brien, Sharon. (2024). Pre-Task Perceptions of MT Influence Quality and Productivity: The Importance of Better Translator-Computer Interactions and Implications for Training. In: Scarton, Carolina; Prescott, Charlotte; Bayliss, Chris; *et al.* (eds.). *Proceedings of the 25th Annual Conference of the European Association for Machine Translation*. European Association for Machine Translation, pp. 444–454. <https://aclanthology.org/2024.eamt-1.37/>. [Accessed: 20251219].

Briva-Iglesias, Vicent; O'Brien, Sharon; Cowan, Benjamin R. (2023). The Impact of Traditional and Interactive Post-Editing on Machine Translation User Experience, Quality, and Productivity. *Translation, Cognition & Behavior*, v. 6, n. 1, pp. 60-86. <https://doi.org/10.1075/tcb.00077.bri>. [Accessed: 20251219].

Briva-Iglesias, Vicent; Peñuelas Gil, Isabel (2025). Simplifying Healthcare Communication: Evaluating AI-Driven Plain Language Editing of Informed Consent Forms. In: Rivas Ginel, María Isabel; Cadwell, Patrick; Canavese, Paolo; Hansen-Schirra, Silvia; Kappus, Martin; Matamala, Anna; Noonan, Will (eds.). *Proceedings of the 1st Workshop on Artificial Intelligence and Easy and Plain Language in Institutional Contexts (AI & EL/PL).* European Association for Machine Translation, pp. 55–65. <https://aclanthology.org/2025.aielpl-1.6/>. [Accessed: 20251219].

Brown, Tom B.; Mann, Benjamin; Ryder, Nick; *et al.* (2020). *Language Models Are Few-Shot Learners.* arXiv:2005.14165. Preprint. <https://doi.org/10.48550/arXiv.2005.14165>. [Accessed: 20251219].

Cadwell, Patrick; O'Brien, Sharon; Teixeira, Carlos S.C. (2018). Resistance and Accommodation: Factors for the (Non-) Adoption of Machine Translation among Professional Translators. *Perspectives*, v. 26, n. 3, pp. 301-321. <https://doi.org/10.1080/0907676X.2017.1337210>. [Accessed: 20251219].

Carl, Michael; Bangalore, Srinivas; Schaeffer, Moritz (eds.) (2016). *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-20358-4>. [Accessed: 20251219].

Carpuat, Marine; Asscher, Omri; Bali, Kalika; *et al.* (2025). *An Interdisciplinary Approach to Human-Centered Machine Translation*. arXiv:2506.13468. Preprint. <https://doi.org/10.48550/arXiv.2506.13468>. [Accessed: 20251219].

Caruana, Edward Joseph; Roman, Marius; Hernández-Sánchez, Jules; Solli, Piergiorgio (2015). Longitudinal Studies. *Journal of Thoracic Disease*, v. 7, n. 11, pp. E537-40. <https://doi.org/10.3978/j.issn.2072-1439.2015.10.63>. [Accessed: 20251219].

Castilho, Sheila (2016). *Measuring Acceptability of Machine Translated Enterprise* Content [PhD Thesis]. Dublin City University. Dublin. <http://doras.dcu.ie/21342/>. [Accessed: 20251219].

Castilho, Sheila. (2021). Towards Document-Level Human MT Evaluation: On the Issues of Annotator Agreement, Effort and Misevaluation. In: Castilho, Sheila; *et al.* *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval),* pp. 34-45. <https://aclanthology.org/2021.humeval-1.4>. [Accessed: 20251219].

Daems, Joke; Macken, Lieve (2019). Interactive Adaptive SMT versus Interactive Adaptive NMT: A User Experience Evaluation. *Machine Translation*, v. 33, n. 1, pp. 1. <https://doi.org/10.1007/s10590-019-09230-z>. [Accessed: 20251219].

Diggle, Peter J.; *et al.* (2002). *Analysis of Longitudinal Data*. 2nd ed. Oxford; New York: Oxford University Press.

Doherty, Stephen; O'Brien, Sharon (2014). Assessing the Usability of Raw Machine Translated Output: A User-Centered Study Using Eye Tracking. *International Journal of Human-Computer Interaction*, v. 30, n. 1, pp. 40-51. <https://doi.org/10.1080/10447318.2013.802199>. [Accessed: 20251219].

ELIS Research. (2023). *European Language Industry Survey 2023: Trends, expectations and concerns of the European language industry.* European Association of Translation Companies (EUATC). <https://elis-survey.org/wp-content/uploads/2023/03/ELIS-2023-report.pdf>. [Accessed: 20251219].

ELIS Research. (2025). *European Language Industry Survey 2025.* European Association of Translation Companies (EUATC). <https://elis-survey.org/wp-content/uploads/2025/03/ELIS-2025_Report.pdf>. [Accessed: 20251219].

European Union (2024). *Artificial Intelligence Act (AI Act).* <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>. [Accessed: 20251219].

Fırat, Gökhan. (2021). Uberization of Translation: Impacts on Working Conditions. *The Journal of Internationalization and Localization*, v. 8, n. 1, pp. 48-75. <https://doi.org/10.1075/jial.20006.fir>. [Accessed: 20251219].

Forcada, Mikel L. (2017). Making Sense of Neural Machine Translation. *Translation Spaces*, v. 6, n. 2, pp. 2. <https://doi.org/10.1075/ts.6.2.06for>. [Accessed: 20251219].

Foster, George; Isabelle, Pierre; Plamondon, Pierre. (1997). Target-Text Mediated Interactive Machine Translation. *Machine Translation,* v. 12, pp. 175–194. <https://doi.org/10.1023/A:1007999327580>. [Accessed: 20251219].

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

Guerberof Arenas, Ana. (2008). Productivity and Quality in the Post-Editing of Outputs from Translation Memories and Machine Translation. *The International Journal of Localisation*, v. 7, n. 1, pp. 11-21.

Guerberof Arenas, Ana; Moorkens, Joss; O'Brien, Sharon (2021). The Impact of Translation Modality on User Experience: An Eye-Tracking Study of the Microsoft Word User Interface. *Machine Translation*, v. 35, n. 2, pp. 205-37. <https://doi.org/10.1007/s10590-021-09267-z>. [Accessed: 20251219].

Hassenzahl, Marc. (2008). User Experience (UX): Towards an Experiential Perspective on Product Quality. In: *Proceedings of the 20th Conference on l'Interaction Homme-Machine (New York, NY, USA), IHM '08, September 2, 11-15.* <https://doi.org/10.1145/1512714.1512717>. [Accessed: 20251219].

ISO. (2018). *ISO 9241-11:2018(En), Ergonomics of Human-System Interaction: Part 11: Usability: Definitions and Concepts.* <https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en>. [Accessed: 20251219].

Jiménez-Crespo, Miguel A.; Rodríguez, Stephanie A. (2025). Is It AI, MT or PE That Worry Professionals: Results from a Human-Centered AI Survey. In: Bouillon, Pierrette; Gerlach, Johanna; Girletti, Sabrina; *et al.* (eds.). *Proceedings of Machine Translation Summit XX Volume 1.* European Association for Machine Translation, pp. 407-419.<https://aclanthology.org/2025.mtsummit-1.32/>. [Accessed: 20251219].

Karakanta, Alina; Bentivogli, Luisa; Cettolo, Mauro; Negri, Matteo; Turchi, Marco (2022). Post-Editing in Automatic Subtitling: A Subtitlers' Perspective. In: Monizz, Helena; et al. (eds.). *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation, June,* 261-70. European Association for Machine Translation. <https://aclanthology.org/2022.eamt-1.29>. [Accessed: 20251219].

Koehn, Philipp. (2009). A Process Study of Computer-Aided Translation. *Machine Translation*, v. 23, n. 4, pp. 4. <https://doi.org/10.1007/s10590-010-9076-3>. [Accessed: 20251219].

Koponen, Maarit; Sulubacak, Umut; Vitikainen, Kaisa; Tiedemann, Jörg (2020). MT for Subtitling: Investigating Professional Translators' User Experience and Feedback. In: . Ortega, John E.; Federico, Marcello; Orasan, Constantin; Popovic, Maja (eds.). *Proceedings of 1st Workshop on Post-Editing in Modern-Day Translation.* Association for Machine Translation in the Americas. <https://aclanthology.org/2020.amta-pemdt.6>. [Accessed: 20251219].

Kovacs, Geza. (2020). Predictive Translation Memory in the Wild: A Study of Interactive Machine Translation Use on Lilt. In: O'Brien, Sharon; Simard, Michael (eds.). *Workshop on the Impact of Machine Translation (iMpacT 2020).* Association for Machine Translation in the Americas, pp. 152–216. <https://aclanthology.org/2020.amta-impact.7/>. [Accessed: 20251219].

Langlais, Philippe, Foster, George; Lapalme, Guy (2000a). TransType: A Computer-Aided Translation Typing System. *ANLP-NAACL 2000 Workshop: Embedded Machine Translation Systems*. <https://www.aclweb.org/anthology/W00-0507>. [Accessed: 20251219].

Langlais, Philippe; Foster, George; Lapalme, Guy (2000b). Unit Completion for a Computer-Aided Translation Typing System. In: *Sixth Applied Natural Language Processing Conference*. <https://doi.org/10.3115/974147.974166>. [Accessed: 20251219].

Läubli, Samuel; Green, Spence (2019). Translation Technology Research and Human-Computer Interaction. In: O'Hagan, Minako (eds.). *The Routledge Handbook of Translation and Technology.* London: Routledge. <https://doi.org/10.4324/9781315311258>. [Accessed: 20251219].

Laugwitz, Bettina; Held, Theo; Schrepp, Martin. (2008). Construction and Evaluation of a User Experience Questionnaire. International. *Journal of Interactive Multimedia and Artificial Intelligence*, v. 4, n. 4, pp. 76. <https://doi.org/10.1007/978-3-540-89350-9_6>. [Accessed: 20251219].

Macías, Lorena Pérez. (2020). What Do Translators Think About Post-Editing? : A Mixed-Methods Study of Translators' Fears, Worries and Preferences on Machine Translation Post-Editing. *Revista Tradumàtica: tecnologies de la traducció,* n. 18, pp. 11-32. <https://doi.org/10.5565/rev/tradumatica.227>. [Accessed: 20251219].

Macklovitch, Elliott. (2006). TransType2: The Last Word. In: Calzolari, Nicoletta; Choukri, Khalid; Gangemi, Aldo; Maegaard, Bente; Mariani, Joseph; Odijk, Jan; Tapias, Daniel (eds.)*. Proceedings of the 5th Edition of the International Conference on Language Resources and Evaluation.* European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2006/pdf/14_pdf.pdf>. [Accessed: 20251219].

Moorkens, Joss. (2020). 'A Tiny Cog in a Large Machine': Digital Taylorism in the Translation Industry. *Translation Spaces*, v. 9, n. 1, pp. 12-34. <https://doi.org/10.1075/ts.00019.moo>. [Accessed: 20251219].

Moorkens, Joss. (2023). 'I Am Not a Number': On Quantification and Algorithmic Norms in Translation. *Perspectives,* v. 32, n. 3, pp. 477-492. <https://doi.org/10.1080/0907676X.2023.2278536>. [Accessed: 20251219].

Nurminen, Mary. (2021). Investigating the Influence of Context in the Use and Reception of Raw Machine Translation [Doctoral thesis]. Tampere University. Tampere. <https://researchportal.tuni.fi/en/publications/investigating-the-influence-of-context-in-the-use-and-reception-o>. [Accessed: 20251219].

O'Brien, Sharon (2006). Pauses as Indicators of Cognitive Effort in Post-Editing Machine Translation Output. *Across Languages and Cultures*, v. 7, n. 1, pp. 1. <https://doi.org/10.1556/Acr.7.2006.1.1>. [Accessed: 20251219].

O'Brien, Sharon (2022). How to Deal with Errors in Machine Translation: Post-Editing. Machine Translation for Everyone. In: Dorothy Kenny (ed.). *Machine translation for everyone: Empowering users in the age of artificial intelligence.* Berlin: Language Science Press, pp. 105-120. <https://zenodo.org/record/6759982/files/342-Kenny-2022-6.pdf?download=1>. [Accessed: 20251219].

O'Brien, Sharon. (2023). Human-Centered Augmented Translation: Against Antagonistic Dualisms. *Perspectives*, v. 32, n. 3, pp. 391-406. <https://doi.org/10.1080/0907676X.2023.2247423>. [Accessed: 20251219].

Olohan, Maeve. (2017). Knowing in Translation Practice: A Practice-Theoretical Perspective. *Translation Spaces*, v. 6, n. 1, pp. 159-80. <https://doi.org/10.1075/ts.6.1.08olo>. [Accessed: 20251219].

Pérez-Ortiz, Juan Antonio; Forcada, Mikel L.; Sánchez-Martínez, Felipe (2022). How Neural Machine Translation Works. In: Kenny, Dorothy (ed.). *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Berlin: Language Science Press, pp. 141-164. <https://zenodo.org/record/6760020/files/342-Kenny-2022-8.pdf?download=1>. [Accessed: 20251219].

Rossetti, Alessandra (2019). *Simplifying, Reading, and Machine Translating Health Content: An Empirical Investigation of Usability* [PhD Thesis]. Dublin City University. Dublin. <http://doras.dcu.ie/23124/>. [Accessed: 20251219].

Rossi, Caroline; Carré, Alice (2022). How to Choose a Suitable NMT Solution?: Evaluation of MT Quality. In: Kenny, Dorothy (eds.). *Machine translation for everyone: Empowering users in the age of artificial intelligence.* Dublin: Language Science Press, pp. 51-79. <https://zenodo.org/record/6759978/files/342-Kenny-2022-4.pdf?download=1>. [Accessed: 20251219].

Sánchez Torrón, Marina (2017). *Productivity in Post-Editing and in Neural Interactive Translation Prediction: A Study of English-to-Spanish Professional Translators* [PhD Thesis]. University of Auckland. Auckland

Sanchis-Trilles, Germán; Alabau, Vicent; Buck, Christian; *et al.* (2014). Interactive Translation Prediction versus Conventional Post-Editing in Practice: A Study with the CasMaCat Workbench. *Machine Translation*, v. 28, n. 3, pp. 217-35. <https://doi.org/10.1007/s10590-014-9157-9>. [Accessed: 20251219].

Schmager, Stefan; Pappas, Ilias O.; Vassilakopoulou, Polyxeni (2025). Understanding Human-Centred AI: A Review of Its Defining Elements and a Research Agenda. *Behaviour & Information Technology*, v. 44, n. 15, pp. 3771-3810. <https://doi.org/10.1080/0144929X.2024.2448719>. [Accessed: 20251219].

Schmidtke, Dag; Groves, Declan. (2019). Automatic Translation for Software with Safe Velocity. In: Forcada, Mikel; Way, Andy; Tinsley, John; Shterionov, Dimitar; Rico, Celia; Gaspari, Federico (eds.). P*roceedings of Machine Translation Summit XVII: Translator, Project and User Tracks.* European Association for Machine Translation, pp. 159–166. <https://aclanthology.org/W19-6729>. [Accessed: 20251219].

Schrepp, Martin; Hinderks, Andreas; Thomaschewski, Jörg (2014). Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In: Marcus, Aaron (ed.). *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience: Thrid International Conference DUXU 2014*. Cham: Springer, pp. 383-392. <https://doi.org/10.1007/978-3-319-07668-3_37>. [Accessed: 20251219].

Schrepp, Martin; Thomaschewski, Jörg; Hinderks, Andreas (2017). Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI),* v. 4, n. 4. <https://doi.org/10.9781/ijimai.2017.445>. [Accessed: 20251219].

Shneiderman, Ben. (2022a). Human-Centered AI. Oxford: Oxford University Press. <https://academic.oup.com/book/41126>. [Accessed: 20251219].

Shneiderman, Ben. (2022b). Human-Centered AI: Ensuring Human Control While Increasing Automation. In: *Proceedings of the 5th Workshop on Human Factors in Hypertext (HUMAN '22).* Association for Computing Machinery. <https://doi.org/10.1145/3538882.3542790>. [Accessed: 20251219].

Torres-Hostench, Olga; Moorkens, Joss; O'Brien, Sharon; Vreeke, Joris (2017). Testing Interaction with a Mobile MT Post-Editing App. Translation & Interpreting: *The International Journal of Translation and Interpreting Research*, v. 9, n. 2, pp. 138-50. <https://www.trans-int.org/index.php/transint/article/view/645>. [Accessed: 20251219].

Torres-Hostench, Olga; Presas, Marisa; Cid-Leal, Pilar (coords.). (2016). *L'ús de Traducció Automàtica i Postedició a les Empreses de Serveis Lingüístics de l'Estat Espanyol: Informe de recerca Projecta 2015.* <https://ddd.uab.cat/record/166753>. [Accessed: 20251219].

Vallor, Shannon. (2024). Defining Human-Centered AI: An Interview with Shannon Vallor. In: Régis, Catherine; Denis, Jean-Louis; Axente, María Luciana; Kishimoto, Atsuo. *Human-Centered AI: A multidisciplinary Perspective for Policy-Makers, Auditors and users*. 1st ed. Boca Raton: Chapman and Hall/CRC, pp. 13-20 <https://doi.org/10.1201/9781003320791-3>. [Accessed: 20251219].

Vanroy, Bram; Tezcan, Arda; Macken, Lieve (2023). MATEO: MAchine Translation Evaluation Online. In: Nurminen, Mary; Brenner, Judith; Koponen, Maarit; *et al.* (eds.). *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pp. 499-500. <https://aclanthology.org/2023.eamt-1.52/>. [Accessed: 20251219].

Vieira, Lucas Nunes; O'Hagan, Minako; O'Sullivan, Carol (2021). Understanding the Societal Impacts of Machine Translation: A Critical Review of the Literature on Medical and Legal Use Cases. *Information, Communication & Society*, v. 24, n. 11, pp. 1515-32. <https://doi.org/10.1080/1369118X.2020.1776370>. [Accessed: 20251219].

Wang, Xiangling; Wang, Tingting; Martín, Ricardo Muñoz and Yanfang Jia (2021). Investigating Usability in Postediting Neural Machine Translation: Evidence from Translation Trainees' Self-Perception and Performance. *Across Languages and Cultures*, v. 22, n. 1, pp. 100-123. <https://doi.org/10.1556/084.2021.00006>. [Accessed: 20251219].

Winner, Langdon (2007). Do Artifacts Have Politics? In: Weckert, John (ed.). *Computer Ethics*. 1st ed. Routledge. <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315259697-21/artifacts-politics-langdon-winner>. [Accessed: 20251219].

Zhong, Junhao; Zhong, Yilin; Han, Minghui; Yang, Tianjian; Zhang, Qinghua (2023). The Impact of AI on Carbon Emissions: Evidence from 66 Countries. *Applied Economics*, v. 56, n. 25, pp. 2975-2989. <https://doi.org/10.1080/00036846.2023.2203461>. [Accessed: 20251219].

Appendix 1. MTUX Questionnaire

This research aims to study machine translation user experience (MTUX), a new concept in Translation Studies, which we consider to be the user experience of translators interacting with machine translation (MT).

The initial hypothesis of this work is that a positive MTUX may influence the translation process, resulting in increased productivity in the long term and a more enjoyable task for MT users in the short term, as well as in a final product (translation) of better quality.

You will now complete one MTUX questionnaire (User Experience Questionnaire), so we can analyse your UX when interacting with MT. In the questionnaire, you will find word pairs that are intended to aid you in assessing the product that you have just become acquainted with. The word pairs represent extreme opposites, with seven graduations possible between them. An example:

Please mark the box to acknowledge that you have read the questionnaire's instructions and that you commit to complete it in its entirety.

Adjective pairs:

Annoying-Enjoyable

Not understandable-Understandable

Creative-Dull

Easy to learn-Difficult to learn

Valuable-Inferior

Boring-Exciting

Not interesting-Interesting

Unpredictable-Predictable

Fast-Slow

Inventive-Conventional

Obstructive-Supportive

Good-Bad

Complicated-Easy

Unlikable-Pleasing

Usual-Leading edge

Unpleasant-Pleasant

Secure-Not secure

Motivating-Demotivating

Meets expectations-Does not meet expectations

Inefficient-Efficient

Clear-Confusing

Impractical-Practical

Organised-Cluttered

Attractive-Unattractive

Friendly-Unfriendly

Conservative-Innovative