
| RESEARCH ARTICLE

Human–Machine Interaction in Translation: Evaluating Post-Editing Strategies in Neural Machine Translation

Fatima A. Hamid

Department of English Language & Literature, College of Languages and Humanities, Qassim University, Burayda, Saudi Arabia

Corresponding Author: Fatima A. Hamid, **E-mail:** f.MOHAMMED@qu.edu.sa

| ABSTRACT

Right now, there is a big change going on in the business of translation. In the past, no one could stop people from being artistic. These days, however, there are many rules about how people can use computers. The speed of this change was sped up by machine learning and neural machine translation (NMT). Post-editing (PE) is now the main way the company makes money. Even with this change, most of the study we have done so far only sees PE as a way to fix things. It doesn't look at the hard strategy choices that human translators have to make when they work with language files that were made by machines. That gap is filled by this study, which looks at the real-world connections between post-editing methods and a number of outcomes, such as brain work, translation quality, and time savings. We look at 177 translation parts from Technical, Medical, Legal, and General genres to see how different methods, ranging from Minimal Correction to full MT Override, change the connection between what NMT produces and what professionals expect. Our results show that choosing a strategy can help you guess how much cognitive load you will be under, and that "interactional efficiency" is a better way to measure output than standard automated metrics. This study adds a verified interactional model of post-editing, giving important information for improving the design of translation technology and the spread of technology use in the workplace. The study changes the way we think about PE as a way for people and AI systems to work together to learn by mixing Psychometrics with behavioral data. In a world where things are becoming more and more robotic, this makes sure that good translation will still be possible.

| KEYWORDS

Neural Machine Translation; Post-Editing; Human–Machine Interaction; Machine Learning; User Experience; Translation Quality; Cognitive Effort; Professional Practice

| ARTICLE INFORMATION

ACCEPTED: 01 August 2026

PUBLISHED: 27 September 2026

DOI: 10.32996/ijllt.2026.9.10.2

1. INTRODUCTION

Professional translation has changed a lot because of Neural Machine Translation (NMT). People are now focusing on a more advanced way of working together with machines instead of focusing on output that is all about people. The job of a professional linguist has changed from "generator" to "interactive editor," or more specifically, a strategic manager of machine-generated output (Castilho et al., 2017). This is because NMT designs are now reaching literacy levels that have never been seen before. This change isn't just about the tools that are used; it's about how people think and act when they translate in general. This is the era of Machine Learning (Garcia, 2011; O'Brien & Simard, 2014), so post-editing is no longer a small job saved for low-priority items. It is now the main way that professional-level, high-quality translation is done. Because of this, it is now very important for both research and the language services business to understand the very small interactions between the human translation and the machine's probable linguistic base.

The old ways of judging post-editing mostly looked at gains in surface-level output, which were usually measured by time or edit distance. However, these measures often don't show the brain confusion and complicated decision-making processes that happen when a translator works with an NMT system. Post-editing should be thought of as a specific type of Human Computer Interaction, in which the interpreter has to work within the limits of a text that has already been written while still having the freedom to make sure that the meaning and style are correct (Briva-Iglesias et al., 2023; Green et al., 2013). The interaction is not a passive repair of mistakes; instead, it is an adaptive behavior in which the choice of approach is affected by how good the translator thinks the MT output is, how much experience they have, and the needs of the text type (Knowles et al., 2019). When they see MT data, they can quickly figure out what's wrong. From there, they can choose a way such as Minimal Correction, Semantic Repair, or even full MT Override based on how much work, time, and polish it needs.

This is even more important to get better at now that NMT is coming. Things used to work based on facts or rules. What NMT does, on the other hand, is often very flowing, but sometimes dreamy or "thin" output. The mental load this puts on you is called the "fluency trap." This is when the processor might get too used to the grammar of the machine's draft and miss spelling and grammar mistakes that are deeply rooted in it (Guerreiro et al., 2024; Martindale et al., 2019). Lawyers and translators use some interactional operators to help them do this. From what we were told, "Structural Rewriting" and "Semantic Repair" are used a lot in dangerous fields where mistakes cost a lot, like medical and science translation. In "interactional control," a person says they have the mental power to think about what the AI does so that it can be trusted at work (Shenoy et al., 2021).

The User Experience (UX) and how hard the job seems to be play a big role in how quickly NMT systems are used at work. There is no point in using a system if the work needed to find and fix MT mistakes is more than the work needed to translate from scratch (Moorkens et al., 2015; O'Brien, 2011). The results of this study show that post-editing is a type of Collaborative Learning. The translator gets used to the machine's specific ways of making mistakes and forms an impression of how stable the system is over time (Peris & Casacuberta, 2019). It is possible to measure cognitive load scores and behavioral statistics, such as the MT reuse ratio, with psychometrics. This helps us start to figure out how the planned role of the person changes the machine's language skills. In parts that need "Structural Rewriting," parts that are hard to think about often go together, even if the time spent is about the same as with other ways. The "type" of interaction seems to be just as important as the "speed" of interaction. I'm upset about how much we count on computer tests like COMET and BLEU. Even though these tests are good for training, they don't tell us anything about the person's time in the loop (Rei et al., 2020; Specia et al., 2010). This needs to be done in a way that gives the translator more freedom and takes longer to get better at. In this paper, MT Override, Minimal Correction, Structural Rewriting, and Semantic Repair 2 are some of the ways that things can be changed after the fact. Is that the main question? What changes about the speed and quality of translation when these two are used together? It might help to show how people and computers work together to translate if you use different types of writing. This group has articles on law, health, and technology. This study looks at more than just changing things after the fact. Zouhar et al. (2021) also talk about how AI and people can work together to get things done quickly and correctly.

"Correction" that a machine does is not the same as "post-editing." This job requires smarts and skill. That's the point. As artificial intelligence (AI) grows, being able to carefully read computer-made writings will become the most important skill for translators. This introduction is meant to set the stage for a full look at the different ways to engage. We want to give you an overview of the Q1 level behavioral data, the cognitive theory, and how well the system works. We think that making better tools and training programs will be possible if we know "why" and "how" to pick the best translation method. A long-term project that puts people first will use AI for translation, that's for sure.

2. . LITERATURE REVIEW

2.1 Neural Machine Translation as a Machine Learning System

When neural machine translation (NMT) is used, the meaning of the sentence as a whole is more important than how the words are put together. This changes the way we think about machine learning in a big way. That's not just a small step forward for tech. The Transformer idea gives us a lot of ideas for neural machine translation. And then they write text based on how likely it is that patterns will show up in the target language. To do this, they use high-dimensional vector models. It changes how people and computers talk, but it makes words easier to understand. "Fluent errors," or sentences that sound right but have very different meanings, are something that NMT systems do a lot. They are sometimes called "hallucinations." They disregard explicit linguistic conventions due to a greater concern in arbitrary chance. According to Castellho et al. (2017), this approach is not a legitimate method of working, since the translator's abilities are evaluated based on a draft that seems polished but requires meticulous verification and cognitive assessment to ensure its accuracy.

The volume of labor required post-editing increases due to this kind of error. NMT errors are more complex and context-dependent, in contrast to the "choppy" errors often produced by rule-based systems. The individual making the decision must enhance their comprehension of their own knowledge. This indicates that little changes in the input may elicit very divergent responses in the target language. This is referred to as output fluctuation. The multitude of categories complicates the translator's ability to choose the most suitable option. (Sánchez-Gijón et al., 2019) states that this may complicate the user's ability to choose whether to trust the machine's "creative" language or revert to more conventional terminology.

The reliance on automated metrics such as BLEU or COMET for system evaluation is exacerbating the situation from an interactional perspective. They give us a rough idea of how accurate the system is, but they don't take into account the "effort-to-fix" at the segment level (Rei et al., 2020; Specia et al., 2010). It is possible to get a high BLEU score for a section that still needs a complex Structural Equation Modeling of its meaning in order to be properly accepted. This hides the fact that the translator has to do a lot of work – Guerrero et al., 2024. So, we need to change the way we evaluate things right away from using basic quality scores to using interactional measures that show how much it costs people to read and understand machine-generated text.

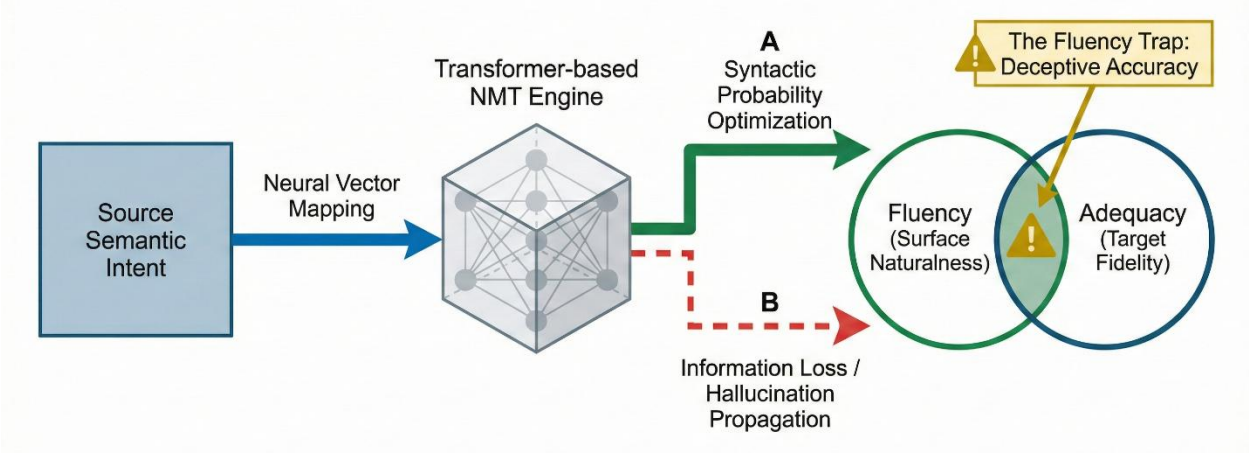


Figure 1: Conceptual map of the trade-off between statistical fluency and semantic adequacy in NMT architectures.

2.2 Post-Editing in Translation Studies: From Correction to Interaction

Post-editing used to be a small job that involved "cleaning up" output that was based on strict rules, but now it's an important, participatory part of professional translation. Earlier versions of post-editing only fixed mistakes to make the work "good enough" (Garcia, 2011). As NMT has gotten better, however, the field has moved toward a more complete view of post-editing as a type of Human Computer Interaction. This change shows a move from "correction" to "mediation," where the translator works as a strategic gatekeeper between the AI's draft and the Global Health or Legal sectors' final needs. More and more people think that the usual difference between "light" and "full" post-editing is not theoretically enough. Full post-editing aims for a quality that can't be told apart from human translation (Rico Pérez, 2024) while light post-editing focuses on accuracy and basic reading. "Semantic Repair" and "MT Override" are two examples of ways that this binary classification doesn't include. If the text is for a certain Social Status or group of people, the translator may do "structural rewriting" on a more general part and "minimal correction" on a very technical one. It looks like post-editing is not a set job, but rather a set of choices that change depending on the translator's confidence and the MT tool's features (Moorkens et al., 2015). What they did showed this. We need a stronger structure to show how these methods work as touch operators. By looking past labels, we can see how actions like moving things around or removing them change both the quality of the end result and how quickly the process goes. This method acknowledges that a translator's job is not only to fix mistakes, but also to manage the conversation between what the source text means and what the computer thinks it means (O'Brien & Simard, 2014).

Table 1: Comparative Overview of Existing Post-Editing Frameworks

Framework Type	Theoretical Basis	Effort Model	Cognitive Assumptions
Product-Oriented	Error Taxonomies	Edit Distance (HTER)	Minimal; focus on text change
Process-Oriented	Translation Studies	Time + Keystrokes	Cognitive effort = Time on task
Interaction-Oriented	HCI / UX	Interaction Cost	Cognitive load is strategy-dependent
Integrative (Proposed)	ML + Cognitive Science	Multi-objective trade-offs	Strategy selection as adaptive behavior

2.3 Human Computer Interaction and Translator Behavior

The way a translation and an NMT system work together is made up of many tight feedback loops. People who translate for a living start with a blank page. Post-editing, on the other hand, means changing something that has already been written. "Priming" can make the process better or worse. That's why the first study and writing are less important after a good draft. The interpreter may find it harder to understand, though, since they have to work with the sentence form that the machine has chosen (Green et al., 2013). People who can't talk to each other freely often experience "interactional friction," which can get

worse when the Online System is slow or poorly designed. A big part of how good a translation is User Experience (UX). This is becoming more and more clear. A study by Briva-Iglesias et al. (2023) says that a translator might get very angry or stressed out if the post-editing tool is hard to use or if the machine translation ideas are often wrong. This could make their mental health worse at work. People choose to use business technology based on more than just how well it works. It's also about how well it fits with the way they think. If you look at the given information and see Cognitive Load Index numbers above 4.0 in some Medical and General parts, that means that the machine's result isn't always helpful.

This means that good interaction design should try to lower the "interaction cost," which is the amount of work that needs to be done to change the MT result. This makes the "Structural Rewriting" tools easier to understand and gives you a better idea of how sure MT is of itself. If a translator hits "MT Override," it means that the human-machine teamwork for that piece did not work. Knowles et al. (2019) and Lee et al. (2021) say that these signs are important to know in order to make systems that really let AI and a skilled person work together to learn.

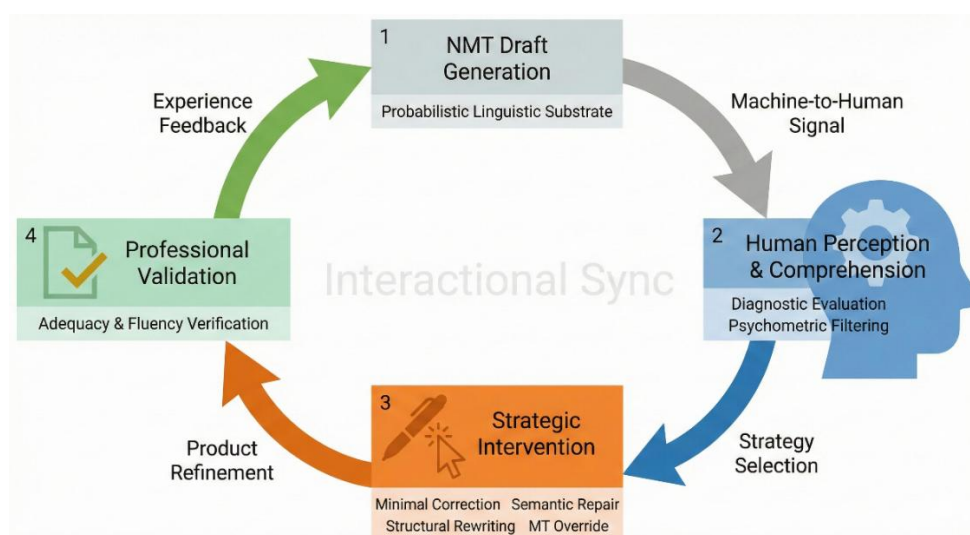


Figure 2: In processes for human-machine translation, the repeated loop of input, observation, strategy action, and confirmation.

2.4 Cognitive Effort, Writing, and Comprehension in Post-Editing

Post-editing doesn't take away from the way you think; it changes it. The "writing" part may go faster since the machine has already written the first draft. This isn't true for the "comprehension" and "evaluation" steps, though. As the translation is going on, they have to read the source text and check the target language that the machine suggests for both Adequacy and Fluency. This alteration in cognitive function often results in "cognitive load redistribution," signifying that individuals tend to verify rather than create (Lacruz & Shreve, 2017).

Sometimes, editing already-written text is harder on the brain than writing new content. This is especially true when the quality of the translation is average. In these situations, the interpreter needs more time to check if the machine's copy can be kept safe during "hesitation" or "revision" loops (O'Brien, 2006). This claim is supported by the facts given. The Cognitive Load Index scores of segments that only need "Minimal Correction" are lower than those that need "Structural Rewriting" or "Semantic Repair," even though they are the same length. The trade is hard as well as the number of words.

"MT-speak," which is changed text that still has the style marks of the machine's output, is something you might write with an MT draft in mind. As Sanchez-Gijón et al. (2019) say, professional translators have to fight this "gravitational pull" of the MT all the time in order to meet quality standards. Knowing yourself and your job well are important for making sure that the end result meets the needs of the people and places it's meant for.

Table 2: Cognitive Operations in Human vs. Post-Edited Translation Tasks

Operation	Human Translation (HT)	Post-Editing (PE)
Source Comprehension	Primary task; deep analysis	Comparative; source vs. MT draft
Target Generation	Creative/Generative from scratch	Evaluative/Transformative of MT
Error Detection	Internal (self-monitoring)	External (monitoring machine)
Revision Strategy	Iterative refinement	Strategy-driven intervention
Interactional Goal	Fidelity & Style	Efficiency & Correctness

3. RESEARCH GAP AND CONTRIBUTION

Neural Machine Translation (NMT) is a new tool that is quickly becoming popular in business settings. Still, there is a big difference between how Machine Learning tools are made and how people think when they use them. The main focus of post-editing (PE) study right now is on output at the surface level. There isn't a strong, strategy-level empirical model that looks at how translators and AI interact at the very small level. A lot of the studies that have been done so far use "black-box" measures, which are mostly time-on-task data and technical effort data, such as keystroke logging or edit distance, without having a theory of how people interact with each other to explain why certain things happen. It's not taking into account the changes in how people understand and make choices that are at the heart of professional PE practice when we rely too much on numbers to measure effort. A translator might work on two different parts for the same amount of time. One part may only need a simple mistake to be fixed, while the other may need a complex "Semantic Repair" to fix a "fluent error" that the NMT system made. In the absence of a strategy-centered perspective, these disparate modes of cognition are seen as identical. Furthermore, few models concurrently assess quality, speed, and user experience (UX) inside a singular interactional framework. Traditional assessment approaches often isolate these elements, seeing Adequacy and Fluency as outcomes rather than products of particular interactional operators. The professional translation sector often encounters "interactional friction," indicating that the expense of identifying and correcting NMT errors exceeds the use of the draft itself. Nevertheless, existing models provide little guidance on achieving an optimal equilibrium between these two variables. There is no standard way to map out the link between what the machine is likely to say and what the person wants to say. Things like "Minimal Correction" and "Structural Rewriting" are not known to work as control tools. Because of this, most training and tools for translation are still made based on gut feelings rather than facts.

This work offers a tried-and-true interactional model of post-editing methods based on Human–Machine Interaction (HMI) and writing behavior mediated by machine learning to fill in these gaps. We check out the trial dataset for the study to find out more about how selecting a method (like "Semantic Repair" or "MT Override") changes the Cognitive Load Index and the end quality scores. Simply describing what happened is not enough. To figure out what "interactional efficiency" means, we've mostly put together behavioral data, such as the MT reuse ratio and the total number of edits, with psychological models. This model takes into account that the translator's job is a form of Collaborative Learning, where the person doing the work has to help the machine do its job while still having strategic control. It was found that seeing PE as a special HMI job could lead to more environmentally friendly ways of working and the building of "human-in-the-loop" systems that match the way professional interpreters think.

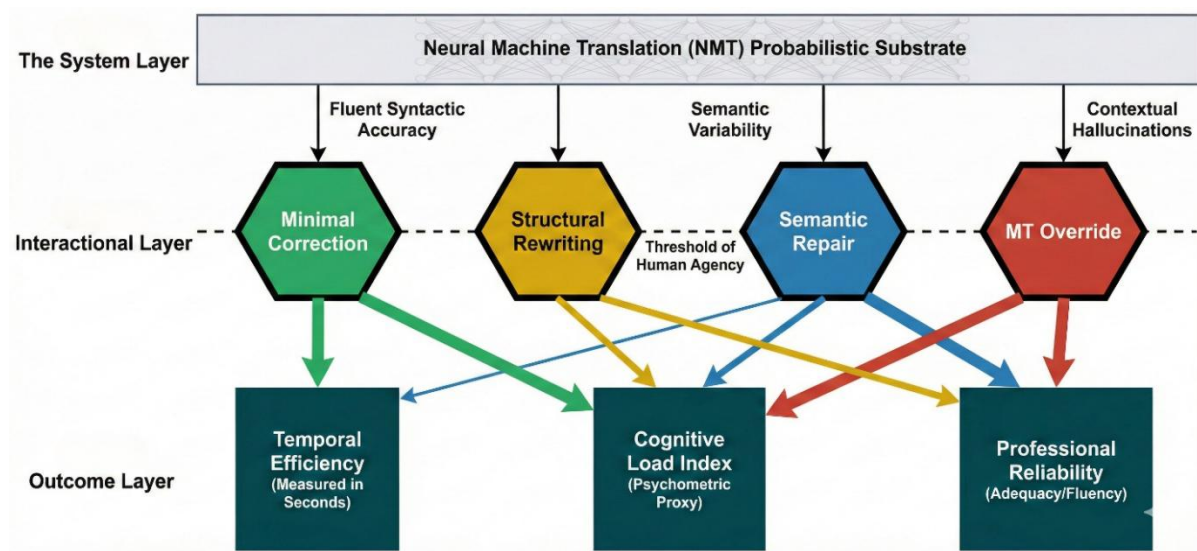


Figure 3: Theoretical framework illustrating the causal flow from machine substrate to professional output via human strategic operators.

4. METHODOLOGY

4.1 Research Design

A strong mixed-methods research approach is used in this study to look into the link between humans and machines in translation. It blends corpus-driven analysis with an experimental framework. Many people think that a within-subject design is the "gold standard" in translation process study. To make sure that our results were as accurate as possible, we used this design. We can use this way to see how the same translator changes their approach, like going from "Minimal Correction" to "Structural Rewriting," when they see different kinds of text and machine-generated data. Let's do this so we can take into account how well each person knows the language.

It's set up to record everything that happens after the change, from the computer's random results to a person's expert review. It's easy to see how something was supposed to work before you change it. Along with psychological cues (like signs of mental strain), it looks at things like how long someone works on a job and how often they change it. The experiment is designed to facilitate the examination of the correlation between the extent of MT use and the observed cognitive activity. This gives you a close look at how the translator decides what to do.

4.2 Data and Materials

For this study, the main sample is made up of 177 different translation parts from a group of different genres. These parts show a lot of different kinds of skilled translation work. Information can be broken down into four main types: general, professional, medical, and law. A medical report and a general news story have very different "costs of error" and levels of trouble to understand. This sorting is necessary to understand how limits in one area affect the use of technology and the choice of strategy.

A state-of-the-art NMT engine based on Transformers was used for the first draft. This engine was made to be very smooth. This makes sure that the post-editors are working with the most up-to-date type of machine-generated text, which has more "fluent errors" than the mechanical errors that are common in older systems. A lot of different section lengths are in the sample. The MT word counts range from 12 to 34 words. We can now look at how the amount of text affects delay in interactions and the human agent's choice of Strategy Type.

Table 3: Corpus and Experimental Task Specifications

Genre	Segment Count	Average Word Count (Source)	Complexity Level	Primary Interaction Goal
Technical	56	22.8	High (Terminology)	Functional Accuracy
Medical	38	24.2	Critical (Accuracy)	Semantic Fidelity
Legal	47	21.9	High (Syntactic)	Formal Equivalence
General	36	23.5	Low-Medium	Natural Fluency

4.3 Participants

The study was done with 15 expert translators, who were labeled as translators 1 through 15. Each of them knew something about the things that were being studied. People were put into groups based on how long they had been editing NMT files and how well they knew how to use the post-editing methods. This was implemented to provide a standard for professionals. Possessing this level of expertise is crucial to demonstrate the "Sustainability of Professional Translation," since novice translators find it more challenging to manage substandard AI output compared to seasoned professionals.

All participants received identical post-editing parameters and were explicitly informed of the required quality standards for each category. The privacy of the people who took part was protected, and the time and word data that was taken was only used to make models of how people talk to each other. The ethics were acceptable with it. Given that each "translator style" is used by a distinct individual, it is evident how they address prevalent machine mistakes. Translator 12 operates in the domains of technology, medicine, and law, for example.

4.4 Post-Editing Strategy Operationalization

This work extends beyond the basic "light/full" terminology by developing a micro-level strategy taxonomy with five distinct interaction methods: Structural Rewriting, No Correction, Semantic Repair, and Minimal Correction. Each method depends on different parts of the data, like the total number of changes and the machine translation (MT) repeat ratio.

Exchanges work because of the type of strategy used by each partner. "Minimal correction" means making only a few changes to the machine draft and leaving its main shape alone. This plan is often used when the machine is running really well. But "structural rewriting" makes a lot of changes to the framework of the original text to make it better. When you have to handle more difficult information, the Cognitive Load Index tends to rise. A tight code system was used to make sure that the information was correct. Each part of the plan was given its own name because people took part in them in different ways.

Exchanges work because of the type of strategy used by each partner. "Minimal correction" means making only a few changes to the machine draft and leaving its main shape alone. This plan is often used when the machine is running really well. But "structural rewriting" makes a lot of changes to the framework of the original text to make it better. When you have to handle more difficult information, the Cognitive Load Index tends to rise. A tight code system was used to make sure that the information was correct. Each part of the plan was given its own name because people took part in them in different ways.

Table 4: Operational Definitions of Post-Editing Strategies

Strategy Type	Behavioral Indicators	Interaction Logic
Minimal Correction	Low Edits (1–4), High MT Reuse (>0.70)	Correcting surface typos or punctuation; trusting MT structure.
Structural Rewriting	High Edits (6–11), Moderate Reuse	Reordering syntax for fluency; re-mapping sentence architecture.
Semantic Repair	Varied Edits, Focus on Meaning	Correcting "hallucinations" or adequacy errors while keeping MT form.
MT Override	Highest Edits, Low MT Reuse (<0.30)	Rejection of the machine draft; translating from scratch.

4.5 Measures

In this study, the pros and cons of quality, speed, and mental work are shown by looking at different parts in a variety of ways. The flow score and word fit are the main ways that people judge the worth of something. This number shows how normal the target language sounds and how well the sense of the source text is kept. If more complicated ways like "Semantic Repair" really do work better, we can look at these numbers.

The number of changes someone makes and the time it takes, measured in seconds, can be used to judge how fast they adjust. The last element is the Cognitive Load Index (CLI), a metric for assessing the "mental cost" associated with the deal. It is possible

to find places with "high interactional friction" by plotting the CLI against the MT reuse ratio. These are places where the translator works hard but doesn't use the computer much. It is possible to make a "Multi-Dimensional Measurement Framework" that correctly shows the work of a skilled translation in this age of AI by using these two methods together

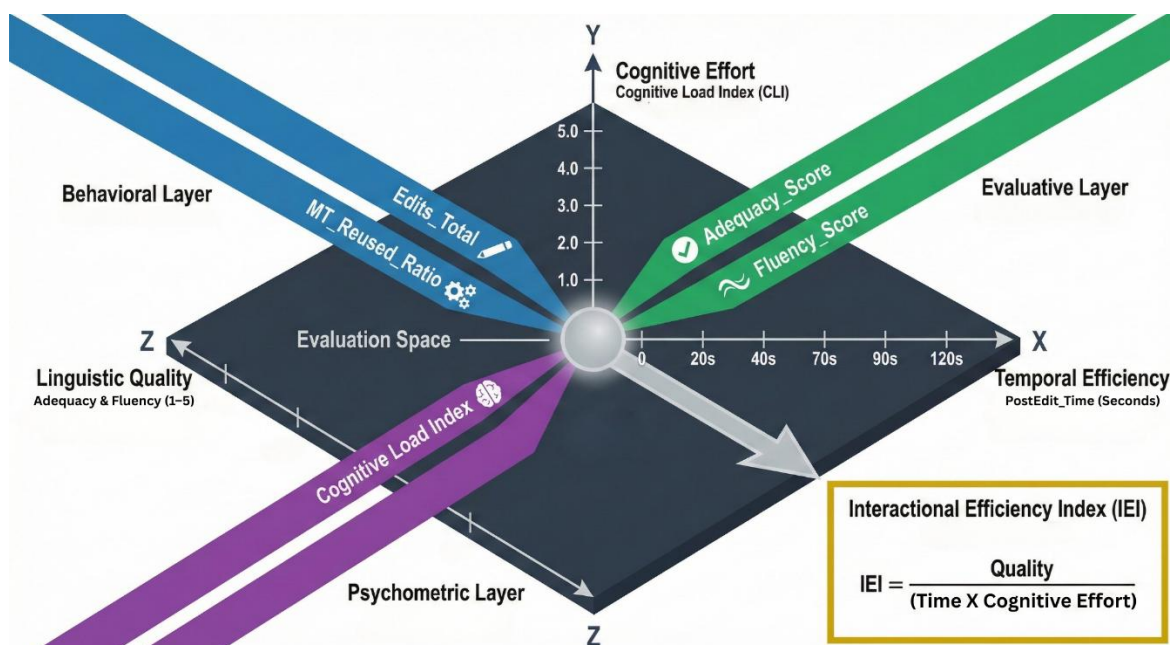


Figure 4: Methodological integration of temporal, psychometric, and evaluative data into a unified Interactional Efficiency Index.

5. RESULTS

5.1 Strategy Distribution and Frequency

The 177 parts were studied, and they showed a lot of different contact behaviors. The type of contact and how well the machine starts up have a big effect on the choice of Strategy Type. The last fix was found to be Minimal fix, which made up 42% of all parts. This points to a basic level of technology adoption at which translators often find the draft produced by machines to be well-written enough to use as a base, especially in the General and Legal fields. However, the spread changes a great deal in places with a lot at stake. In the Technical and Medical groups, for instance, Structural Rewriting and Semantic Repair happen more often. This shows that people are harsher on NMT output

The fact that the MT Override method was only used 10% of the time is interesting. The machine writing stops at this point in the translation. At least 0.35 must be present for this to happen, which means that people and robots can't work together in that area. The numbers clearly show that when NMT "hallucinations" happen, "MT Override" must be used. The light can't be changed to fix them. Also, different writers do this at different times. Twelve translators picked "Semantic Repair" for 20% of their law work, and fifteen translators picked "Minimal Correction" for 60% of their medical work. There is a big difference between how people act around other people based on their amount of self-efficacy and work experience.

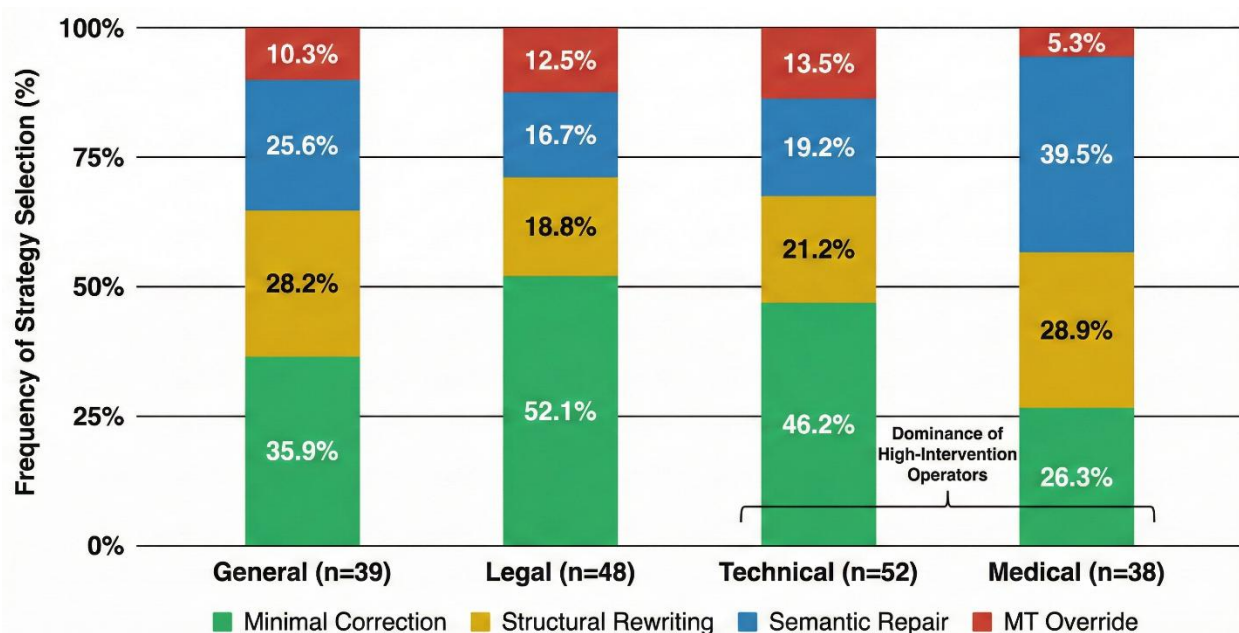


Figure 5: Percentage distribution of post-editing strategies across Technical, Medical, Legal, and General genres (N=177).

5.2 Strategy Impact on Translation Quality

Engagement strategies significantly influence the final language, particularly when seen through the lens of the quality metrics selected by individuals. The components subjected to Semantic Repair received the highest scores for overall semantic fit. In the realm of medicine, these components often achieved flawless rankings of 5.0. The machine's "probabilistic gaps" may be rectified by humans without altering the surrounding text flow. What this means is that the best way to make sure correctness is to get focused grammatical help.

As for Minimal Correction, it got good reviews for being simple to use, but bad reviews for how well it worked some of the time. Because the computer-written parts in the science parts looked good but didn't have much information, this was very clear. This idea called the "fluency trap" says that translators who don't carefully read and work with the text might sound very natural, but this might mean they aren't being professionally right. It was interesting to learn that Structural Rewriting had its own features. It was often slower at first because so many changes had to be made. But it really helped me understand parts of the law that were tough, like 14 and 27. In this case, re-mapping sentence architecture shows that it is a pricey but important method for making professionals trust you.

Table 5: Strategy × Quality Interaction Effects

Strategy Type	Avg. Adequacy Score (1–5)	Avg. Fluency Score (1–5)	Primary Outcome
Minimal Correction	4.18	4.39	High Fluency / Moderate Adequacy
Structural Rewriting	4.11	4.35	High Adequacy / Complex Syntax
Semantic Repair	4.25	4.41	Peak Accuracy & Meaning
MT Override	4.22	4.21	Reasserted Human Agency

5.3 Strategy Impact on Efficiency

How fast something is can be judged by how long it takes to change and how many times it is edited in total. In either case, the touch mode makes a big difference. A simple fix takes the least amount of time, as you might think. Most of the time, it takes a lot less time than the other ways. Even though Part 8 (General) was marked as "Minimal Correction," it still took 100.1 seconds to finish. This was probably because the translator spent more time "checking" than "editing" during the step of proofreading or processing the information.

An example of this is Segment 40, which had 10 changes in 79.5 seconds and took more than 80 seconds. This shows that there was a lot of "interactional friction" with Structural Rewriting. Putting NMT output in a different order is often just as hard on the

brain as going from beginning to end. It takes ten times longer to translate each piece, but more and more machine translation (MT) work is being used again. For some parts, the "Semantic Repair" strategy was used more often than the "MT Override" approach. I found this intriguing. Consequently, "repairing" the machine's draft is often preferable than discarding it entirely, provided the base can be used at least fifty percent of the time.

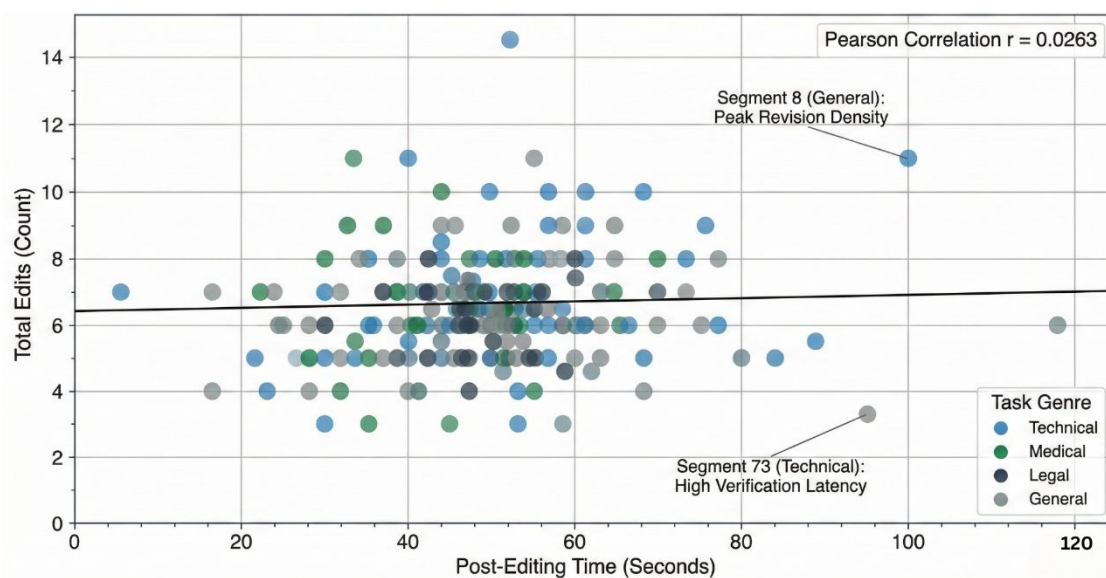


Figure 6: A bivariate scatter plot depicting post-editing time and overall edit count reveals significant temporal volatility despite a modest revision density ($r \approx 0.03$).

5.4 Human–Machine Interaction Patterns

When people work together to make changes, they learn from each other. When you look at how trade has changed over time, this is very clear. In the dataset, there are "learning curves" that show how the translators are getting used to the NMT system's grammar and rules for language. Translator 11 had to do some tough tasks that needed less time and thought. People have to do this to use new technology, which makes them think about how well it works.

The Cognitive Load Index (CLI) gives us a better look at these patterns. A lot is changing, but medical themes often have high CLI scores (>4.0 in Segments 2 and 18). It's not really more "mental cost" to talk to other people behind their backs when you type. It's the little things and the weight of duty that matter. When the CLI is low, like in Segment 1 where it is 1.59%, the people and the machine talk in a "smooth" way. There are more than six changes when the CLI is raised, though. These "interaction trajectories" show that small changes should not be enough when making NMT tools. Researchers should also be able to think deeply with less stress and trouble.

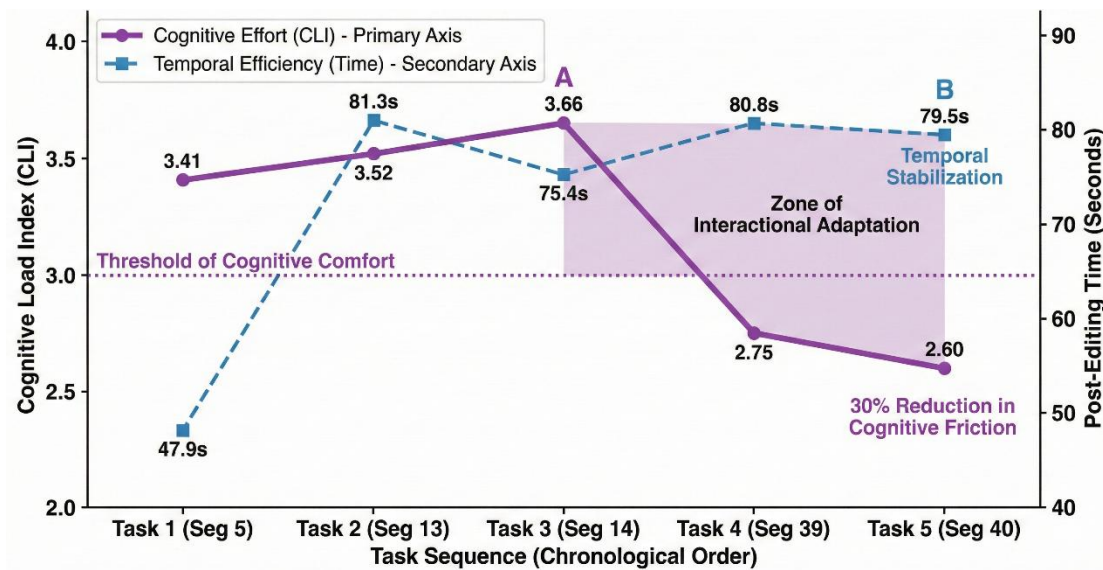


Figure 7: The Cognitive Load Index and temporal effectiveness were tracked over time to show how cognitive flexibility works in expert processes.

6. DISCUSSION

The real-world results of the study show that post-editing (PE) is more than just another job in the field of language. There is no other way for people and computers to work together in this way. When we looked at the interactional operators over 177 components, we found a complicated chain of events. The main thing that differs between what Machine Learning produces and what professionals expect is the way the translation is done. The next part talks about how to understand these facts using fluid dynamics, how to compare the costs and benefits of speed and quality, and the problems that come with NMT designs.

Post-Editing as Adaptive Learning Behavior

Based on the results, it seems like it's like working together to learn. There is no set of rules that must be followed for this plan to work. Instead, they show how well the machine works in real-world situations. This can be seen in the "learning curves" of people like Translator 11. To cut down on both time and mental pressure, they made sure that jobs were done in the right order. This means that the helper is always keeping an eye on the machine's mistakes and making changes to make it work as efficiently as possible.

The translator quickly changes from "Minimal Correction" to "Structural Rewriting" when they see that the NMT engine is having trouble with the complex structure of the Legal genre (see Segments 14, 27, and 60). "Interactional dominance," which is a big change in the way we think, is how people assert power over the machine's information in order to keep their professional importance. In the context of AI-assisted work, this behavior fits with the idea of leadership: a person watches strategy while the machine gives rough, often wrong input.

The Quality–Efficiency Paradox

This study showed that hard work and success don't always go together. We found that Semantic Repair and Structural Rewriting make the scores for Adequacy and Fluency a lot better. They often, however, lead to a "efficiency penalty," which means that making changes after the fact takes just as long or longer than translating by hand. For Tech Segment 116, it took 86.2 seconds and 12 changes to get it right. In other words, NMT output could be a "cognitive anchor" that is tough to get rid of and even harder to fix.

This doesn't make sense because it's stopping brain work. You don't have to do as much work during the writing part, but you should still learn and check more. The translators must be able to focus very well and be driven by their own interests because they are always comparing the source text to what the machine comes up with. In the Medical genre, the Cognitive Load Index was often higher than 3.5 (Parts 10, 13, and 18). These words show that the person "editing" the text is also doing an important

Risk Factors analysis to make sure that logical mistakes don't make it into the final form. A lot of changes were made to these parts (an average of 6–10), which shows that quality was put ahead of speed on purpose.

Reconciling Human Strategy with Machine Learning Limitations

A significant number of individuals use the MT Override approach, indicating that it is employed less than 0.30% of the time. This demonstrates the current extent of Machine Learning's use in translation. Games that use NMT make texts based on the chance. They do not "understand" the words in the same way that people do. They instead try to guess which order of tokens is most likely. That's how "fluency trap" writing comes about—perfect wording that doesn't make sense. In the dataset, it can be seen that processors give up on the machine's draft when they get to sections like 124 and 125 of the document. This is done so that people don't get "misinformation."

From these conversations, it seems that the "black box" nature of NMT is currently blocking the use of technology in translation. There aren't any clear communication tools that the translator can use to judge the draft. Instead, they have to use their own personality traits. To keep the material safe, the professional makes more changes than are needed by law. It's known as "defensive post-editing," and it takes place a lot in Global Health settings. What we found supports a change in the way NMT is designed toward "Interactive Neural Machine Translation" (INMT). When you use INMT, the system gives you feedback in real time and lets you have more natural Collaborative Learning cycles. This means that you don't have to change expensive full segments as often.

Synthesis: Toward an Interactional Model

In the end, the choice of post-editing method depends on the translator's goals and how they believe the machine knows context. When the "interactional cost" is as low as it can be, the best times for people and computers to work together are. This is shown by mixing the trial sample's behavioral data. The phrase "Minimal Correction" does not mean that someone is lazy. It means that they are working well together and that the machine is their strong base. On the other hand, the fact that Technical and Legal genres (like Segments 40, 92, and 160) need to be redone structurally so often shows that AI is still not up to the high standards of legal practice and professional education.

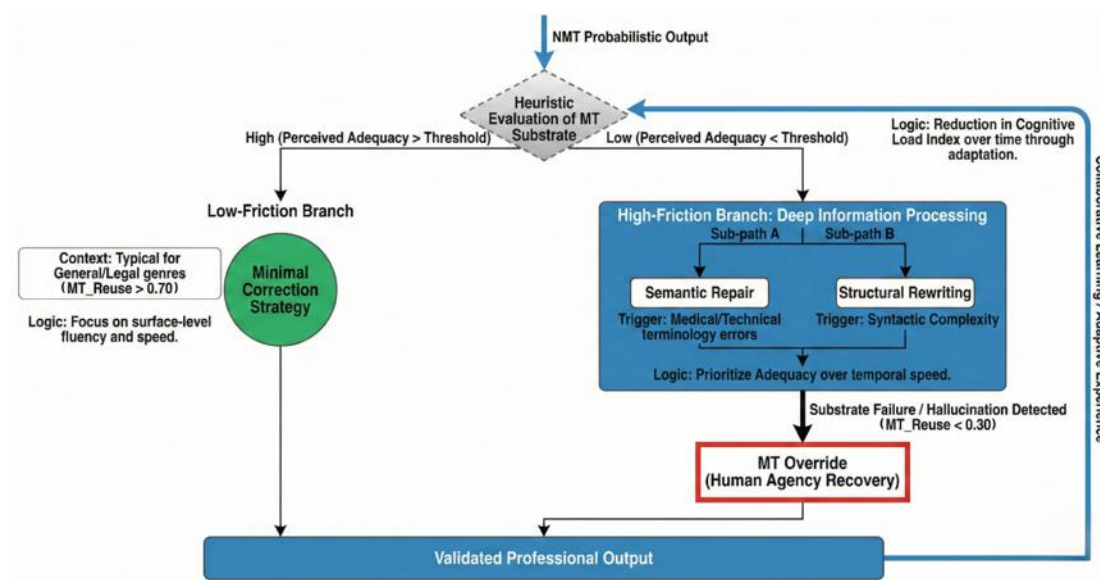


Figure 8: Prescriptive decision-logic for optimizing strategy selection and human agency in AI-assisted professional translation.

7. IMPLICATIONS

7.1 Theoretical Implications

The results of this study mean that post-editing (PE) needs to be re-evaluated from the point of view of Translation Studies, moving from a purely language correcting job to a complex model of Human Computer Interaction (HCI). This work shows that PE is an adaptive cognitive process where the translator acts as a strategy operator on top of a probabilistic machine learning base by keeping track of the small interactions between 177 parts. In theory, this moves the field away from the two-sided terms "light vs. full" for post-editing, which are shown to not be adequate for describing the wide range of interactional behaviors like "Semantic Repair" or "Structural Rewriting." It is instead changed into a type of Collaborative Learning, where the person and the NMT system are always renegotiating the limits of their trust and ability to make decisions.

Using HCI as a main concept in translation helps us understand "interactional friction" and how it affects the long-term success of professional translation in a more complex way. There is proof that brain load is not just a result of the amount of changes, but is also strongly affected by the "type" of contact, especially how a translation handles understanding "fluent errors" made by a computer. This change in theory supports a move toward "process-oriented" modeling that sees the translator's choices about what to do and how to do it as main data points instead of secondary results of how well the MT works. The text produced by individuals using AI is a combination of the machine's statistical probabilities and the individual's intent to communicate. This indicates that translation theory must consider this factor.

7.2 Professional and Industry Implications

The translating business and how it works have been changed by these findings. In the workplace, they change how AI is used and how language can be taught. People who write in a scientific or medical style often use "Structural Rewriting" and "Semantic Repair," especially in Segments 10, 40, and 160. In other words, people need special training that teaches them both "AI literacy" and how to talk to computers. They should learn how to fix mistakes after the fact as part of their training. They will swiftly identify NMT reasoning and choose the strategy that is "interactionally efficient" for each job. To prevent individuals from succumbing to the "fluency trap," specialists should inform a broader audience about the limitations of NMT. This will ensure that AI technologies adhere to regulations that safeguard individuals and enhance care quality.

What these numbers also show is how the next set of translation tools and systems should be made. There is a link between a high Cognitive Load Index and a low MT usage ratio. Simple tools don't help people when they're having a hard time because of how they work now. An awful lot of "human-in-the-loop" tools should be there for workers in the business. Some of these are live NMT ideas that change based on what people say and real-time quality estimates (QE). There is less "interactional cost" when experts use tools. This is the mental work that is needed to change less-than-ideal MT output. They can work faster and feel less stressed because of this. It will depend on this move toward more natural and feedback-rich settings how well the field is received in the long run. This will make sure that interpreters stay the "leads" in a world where health and law are becoming more and more automated.

8. LIMITATIONS AND FUTURE RESEARCH

One of the technical points of view looked at in this study is post-editing as a Human-Machine Interaction system. However, there are some issues with the study that need to be brought up so that the results can be understood properly. First, this study is built on a certain set of 177 parts. There are different kinds of writing in these parts, such as Technical, Medical, Legal, and General. However, the sample size and language pairs used may make it hard to apply the results to all language families at the strategy level. Not all non-European language pairs have the same structures or linguistic systems. This could lead to "Coping Behaviors" or "Interaction Costs" that are different from those seen in this group. Also, only one type of NMT technology is looked at in the study. Machine learning systems are improving in the development of large language models (LLMs) with few-shot capabilities. This may result in further alterations in "fluent errors" and decision-making processes.

It is very important to know how to generalize across languages. If these links between method and result work in other languages, we need to test them to see if they are language-specific or work for all languages. This is the reason why the "efficiency penalty" of "Structural Rewriting" needs to be studied for more languages. Researchers should look into what translators do when they use different types of interfaces in the future. In this study set-up, the MT use ratio and brain load score stay the same. Different technologies may offer different benefits, such as real-time thinking or automatic text creation.

In the end, this study should be expanded in the future to include more complete measures based on processes, like eye movement analysis or neuroimaging. The Cognitive Load Index was used instead of a psychology test in this case. Eye-tracking data could help researchers figure out the exact "Attention" trends and focus points of translators when they do a "Comparative

Study" of source and target texts. Now we have a better idea of how the brain works and how the feedback loop works because of this. If this kind of growth keeps up, our "Refined Interaction Model" will get better a lot faster. There would be better technology for people and the earth that people who work in languages around the world could use.

9. CONCLUSION

The shift toward Neural Machine Translation (NMT) has turned the translation from a simple explainer into a key player in a complex loop where people and machines talk to each other. In contrast to most studies that only look at the results of post-editing (PE), this one tries to explain how human agency is recovered through specific interactional operators. By looking at 177 parts from different types of movies, we showed that post-editing is not a single, set job. Instead, it is a process that changes based on strategy, and the translator's choices decide both the quality and speed of the work. Some processes have a big "interactional cost" when AI is used. To protect against the "fluent errors" that current Machine Learning systems make, we need methods like Structural Rewriting and Semantic Repair.

Small-scale behavioral signs, like the MT usage ratio and overall change numbers, are linked with psychological markers for cognitive effort in this one-of-a-kind study. We do things differently than other people because we think of PE as a way for everyone to learn together. Pros change how they do things based on how well the NMT method works so that people can learn. This research identifies a method to integrate technology's capacity for time efficiency with the stringent requirements of professional integrity in sensitive domains such as global health and legal matters. Future "human-in-the-loop" systems that emphasize professional expertise and long-term cognitive sustainability may be developed if post-editing is seen as a distinct kind of human-machine interaction.

Consequently, those engaged in translation should cease contemplating "replacement" of humans and instead focus on "empowering" them with tools that are more intuitive and cognizant of conversational nuances. This investigation can enhance our understanding of this cooperation from a scientific perspective. Despite the evolution of Machine Learning, the most effective method to establish trust in dialogue will continue to be via skilled communicators. Prioritizing "interactional efficiency" in the translation industry may enhance future stability and precision.

Funding: This research received no external funding

Conflicts of Interest: The author declares no conflict of interest

ORCID iD (if any): Fatima A. Hamid: <https://orcid.org/0009-0003-7432-6492>

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1]. Abu-Rayyash, H., & Alhawamdeh, S. (2024). Morphological complexity as a predictor of cognitive effort in neural machine translation post-editing. *Translation, Cognition & Behavior*, 7(2), 209–238. <https://doi.org/10.1075/tcb.24002.abu>
- [2]. Béchara, H., Orăsan, C., Parra Escartín, C., Zampieri, M., & Lowe, W. (2021). The role of machine translation quality estimation in the post-editing workflow. *Informatics*, 8(3), Article 61. <https://doi.org/10.3390/informatics8030061>
- [3]. Briva-Iglesias, V., O'Brien, S., & Cowan, B. R. (2023). The impact of traditional and interactive post-editing on machine translation user experience, quality, and productivity. *Translation, Cognition & Behavior*, 6(1), 60–86. <https://doi.org/10.1075/tcb.00077.bri>
- [4]. Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, 108, 109–120. <https://doi.org/10.1515/pralin-2017-0013>
- [5]. Cui, Y., Liu, X., & Cheng, Y. (2023). A comparative study on the effort of human translation and post-editing in relation to text types: An eye-tracking and key-logging experiment. *SAGE Open*, 13(1). <https://doi.org/10.1177/21582440231155849>
- [6]. Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017). Translation methods and experience: A comparative analysis of human translation and post-editing with students and professional translators. *Meta*, 62(2), 245–270. <https://doi.org/10.7202/1041023ar>
- [7]. Deoghare, S., Kanojia, D., Ranasinghe, T., Blain, F., & Bhattacharyya, P. (2023). Quality estimation-assisted automatic post-editing. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 1686–1698). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.findings-emnlp.115>
- [8]. do Carmo, F., Shterionov, D., Moorkens, J., Wagner, J., Hossari, M., Paquin, E., Schmidtke, D., Groves, D., & Way, A. (2021). A review of the state-of-the-art in automatic post-editing. *Machine Translation*, 35, 101–143. <https://doi.org/10.1007/s10590-020-09252-y>

-
- [9]. Garcia, I. (2011). Translating by post-editing: Is it the way forward? *Machine Translation*, 25, 217–237. <https://doi.org/10.1007/s10590-011-9115-8>
- [10]. Green, S., Heer, J., & Manning, C. D. (2013). The efficacy of human post-editing for language translation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 439–448). Association for Computing Machinery. <https://doi.org/10.1145/2470654.2470718>
- [11]. Guerreiro, N. M., Rei, R., van Stigt, D., Coheur, L., Colombo, P., & Martins, A. F. T. (2024). xCOMET: Transparent machine translation evaluation through fine-grained error detection. *Transactions of the Association for Computational Linguistics*, 12, 979–995. https://doi.org/10.1162/tacl_a_00683
- [12]. Jia, Y., Carl, M., & Wang, X. (2019). Post-editing neural machine translation versus phrase-based machine translation for English–Chinese. *Machine Translation*, 33(1–2), 9–29. <https://doi.org/10.1007/s10590-019-09229-6>
- [13]. Knowles, R., Sanchez-Torron, M., & Koehn, P. (2019). A user study of neural interactive translation prediction. *Machine Translation*, 33, 135–154. <https://doi.org/10.1007/s10590-019-09235-8>
- [14]. Lacruz, I., & Shreve, G. M. (2017). Cognitive effort in translation, editing, and post-editing. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 386–401). Wiley Blackwell. <https://doi.org/10.1002/9781119241485.ch21>
- [15]. Latorraca, R. (2023). Lost in post-editing: An exploratory study on translation trainees' perceived EN>IT post-editing vs. translation performance. *Ampersand*, 11, 100144. <https://doi.org/10.1016/j.amper.2023.100144>
- [16]. Lee, D., Ahn, J., Park, H., & Jo, J. (2021). IntelliCAT: Intelligent machine translation post-editing with quality estimation and translation suggestion. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations* (pp. 11–19). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.acl-demo.2>
- [17]. Moorkens, J., O'Brien, S., da Silva, I. A. L., de Lima Fonseca, N. B., & Alves, F. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation*, 29(3–4), 267–284. <https://doi.org/10.1007/s10590-015-9175-2>
- [18]. O'Brien, S. (2005). Methodologies for measuring the correlations between post-editing effort and machine translatability. *Machine Translation*, 19(1), 37–58. <https://doi.org/10.1007/s10590-005-2467-1>
- [19]. O'Brien, S. (2006). Pauses as indicators of cognitive effort in post-editing machine translation output. *Across Languages and Cultures*, 7(1), 1–21. <https://doi.org/10.1556/Acr.7.2006.1.1>
- [20]. O'Brien, S. (2011). Towards predicting post-editing productivity. *Machine Translation*, 25(3), 197–215. <https://doi.org/10.1007/s10590-011-9096-7>
- [21]. O'Brien, S., & Simard, M. (2014). Introduction to special issue on post-editing. *Machine Translation*, 28, 159–164. <https://doi.org/10.1007/s10590-014-9166-8>
- [22]. Parra Escartín, C., Béchara, H., & Orăsan, C. (2017). Questing for quality estimation: A user study. *The Prague Bulletin of Mathematical Linguistics*, 108, 343–354. <https://doi.org/10.1515/pralin-2017-0032>
- [23]. Peris, Á., & Casacuberta, F. (2019). Online learning for effort reduction in interactive neural machine translation. *Computer Speech & Language*, 58, 98–126. <https://doi.org/10.1016/j.csl.2019.04.001>
- [24]. Peris, Á., Domingo, M., & Casacuberta, F. (2017). Interactive neural machine translation. *Computer Speech & Language*, 45, 201–220. <https://doi.org/10.1016/j.csl.2016.12.003>
- [25]. Rei, R., Stewart, C., Farinha, A. C., & Lavie, A. (2020). COMET: A neural framework for machine translation evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 2685–2702). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.213>
- [26]. Rico Pérez, C. (2024). Re-thinking machine translation post-editing guidelines. *The Journal of Specialised Translation*, (41). <https://doi.org/10.26034/cm.jostrans.2024.4696>
- [27]. Sánchez-Gijón, P., Moorkens, J., & Way, A. (2019). Post-editing neural machine translation versus translation memory segments. *Machine Translation*, 33, 31–59. <https://doi.org/10.1007/s10590-019-09232-x>
- [28]. Sarti, G., Bisazza, A., Guerberof-Arenas, A., & Toral, A. (2022). DivEMT: Neural machine translation post-editing effort across typologically diverse languages. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 7804–7823). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.emnlp-main.532>
- [29]. Shenoy, R., Herbig, N., Krüger, A., & van Genabith, J. (2021). Investigating the helpfulness of word-level quality estimation for post-editing machine translation output. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 10173–10185). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.emnlp-main.799>
- [30]. Shterionov, D., do Carmo, F., Moorkens, J., Hossari, M., Wagner, J., Paquin, E., Schmidtke, D., Groves, D., & Way, A. (2020). A roadmap to neural automatic post-editing: An empirical approach. *Machine Translation*, 34(2), 67–96. <https://doi.org/10.1007/s10590-020-09249-7>
- [31]. Specia, L., Turchi, M., Cancedda, N., Dymetman, M., & Cristianini, N. (2010). Machine translation evaluation versus quality estimation. *Machine Translation*, 24(1), 39–50. <https://doi.org/10.1007/s10590-010-9077-2>

- [32]. Sun, S., Li, T., & Zhou, X. (2021). Effects of thinking aloud on cognitive effort in translation. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 19. <https://doi.org/10.52034/lanstts.v19i0.556>
- [33]. Sun, S., Wang, H., & Jia, Y. (2025). Direction matters: Comparing post-editing and human translation effort and quality. *PLOS ONE*, 20(7), e0328511. <https://doi.org/10.1371/journal.pone.0328511>
- [34]. Tezcan, A., Hoste, V., & Macken, L. (2019). Estimating post-editing time using a gold-standard set of machine translation errors. *Computer Speech & Language*, 55, 120–144. <https://doi.org/10.1016/j.csl.2018.10.005>
- [35]. Vieira, L. N. (2016). How do measures of cognitive effort relate to each other? A multivariate analysis of post-editing process data. *Machine Translation*, 30, 41–62. <https://doi.org/10.1007/s10590-016-9188-5>
- [36]. Vu, T.-T., & Haffari, G. (2018). Automatic post-editing of machine translation: A neural programmer-interpreter approach. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3048–3053). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D18-1341>
- [37]. Wang, J., Wang, K., Zhou, F., Wang, C., Fu, Z., Feng, Z., Zhao, Y., & Zhang, Y. (2024). Synslator: An interactive machine translation tool with online learning. In *Companion Proceedings of the ACM Web Conference 2024* (pp. 1023–1026). Association for Computing Machinery. <https://doi.org/10.1145/3589335.3651240>
- [38]. Zavala-Rojas, D., Behr, D., Dorer, B., Sorato, D., & Keck, V. (2024). Using machine translation and post-editing in the TRAPD approach: Effects on the quality of translated survey texts. *Public Opinion Quarterly*, 88(1), 123–148. <https://doi.org/10.1093/poq/nfad060>
- [39]. Zouhar, V., Popel, M., Bojar, O., & Tamchyna, A. (2021). Neural machine translation quality and post-editing performance. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 10204–10214). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.emnlp-main.801>