



The Effect of ChatGPT-Assisted Pre-Editing on Naturalness and Readability of Google Translate's Persian Output

Hamideh Nemati Lafmejani* 

Corresponding Author, Assistant Professor of Translation Studies, Department of English Language, SR.C., Islamic Azad University, Tehran, Iran;
hamideh.nemati@iau.ac.ir

Seyed Mehdi Mousavian 

MA in Translation Studies, Department of English Language, SR.C., Islamic Azad University, Tehran, Iran;
seyedmehdi.mousavian@iau.ir

Abstract

The quality of machine translation depends not only on the capabilities of the translation system but also on the linguistic characteristics and complexity of the source text. Although pre-editing has been widely explored in relation to controlled language and machine translation, the potential of generative artificial intelligence, particularly ChatGPT, as a pre-editing tool remains underexplored. This study investigates the effectiveness of ChatGPT-5 in pre-editing English news texts prior to machine translation into Persian. Thirty-five English news articles were processed using ChatGPT-5 to generate controlled language optimized for machine translation. Both the original and ChatGPT-pre-edited versions were subsequently translated into Persian using Google Translate and evaluated by two professional translators for naturalness and readability on a three-point scale. The results demonstrate notable improvements following ChatGPT-5 pre-editing: mean naturalness increased from 1.45 to 2.34, while mean readability rose from 1.77 to 2.71. These findings suggest that ChatGPT-assisted pre-editing can enhance the naturalness and readability of machine-translated English-to-Persian journalistic texts. The study highlights generative AI as a promising addition to the machine translation workflow, extending the role of pre-editing from a traditionally human-intensive practice toward an AI-assisted process and contributing to the growing body of research on generative AI in translation.

Keywords: AI-assisted translation, machine translation, pre-editing naturalness, readability

Cite this article: Nemati Lafmejani, H., & Mousavian, S. M. (2026). The effect of ChatGPT-assisted pre-editing on naturalness and readability of Google Translate's Persian output. *Translation and Interpreting Research*, 3(12), 41-55. <https://doi.org/10.22054/tir.2026.96338.1100>



© 2026 – Allameh Tabataba'i University Press
Translation and Interpreting Research is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license.

Print ISSN: 3092-6599

Online ISSN: 3060-7019

Publisher: Allameh Tabataba'i University Press

Published: 30.12.2026

Accepted: 29.10.2026

Received: 27.09.2026

1. Introduction

Machine translation (MT) has undergone major technological transformations, moving from early rule-based approaches to statistical and, more recently, neural approaches. While earlier systems relied heavily on manually designed linguistic rules and statistical models, neural machine translation (NMT) uses deep neural networks to model the relationship between source and target languages within an end-to-end framework (Koehn, 2020; Stahlberg, 2020). This development has substantially improved the fluency and overall quality of machine-generated translations, contributing to the widespread adoption of MT in both professional and everyday translation contexts. Nevertheless, the quality of MT output depends not only on the architecture and training of the MT system but also on characteristics of the source text. Recent research has therefore increasingly drawn attention to the preparation and modification of source texts before they are submitted to MT systems (Sánchez-Gijón & Kenny, 2022).

The relationship between source-text characteristics and MT performance is particularly important because MT systems may encounter difficulties when processing ambiguous, syntactically complex, inconsistent, or otherwise difficult-to-process source texts. One approach to addressing such problems is the use of controlled language, which imposes explicit restrictions on aspects of vocabulary, grammar, and style in order to improve comprehensibility and machine translatability (Huijsen, 1998). Research has shown that controlled-language rules can influence MT accuracy and reduce the effort required for subsequent post-editing (Krüger, 2022; O'Brien, 2006). Thus, rather than focusing exclusively on improving the MT system itself, researchers have also investigated whether modifying the source text can facilitate the translation process.

Within this context, pre-editing has emerged as a strategy for improving the suitability of source texts for MT. Sánchez-Gijón and Kenny (2022) define pre-editing as “rewriting parts of source texts in a way that is supposed to ensure better quality outputs when those texts are translated by machine” (p. 81). They further explain that pre-editing may involve applying formal rules, sometimes referred to as controlled-language rules, which specify permitted words or structures and prohibit others. Pre-editing can therefore involve a range of operations, including simplifying syntactic structures, resolving ambiguities, making implicit information more explicit, standardizing terminology, and improving the overall clarity of the source text. Importantly, pre-editing is not necessarily synonymous with merely simplifying a text. Miyata and Fujita (2021), for example, found that making the meaning and syntactic structure of the source text more explicit can be more important for improving NMT output than simply making the source text shorter or simpler.

Pre-editing is traditionally a human-intensive process. It may require substantial linguistic knowledge, familiarity with the requirements of MT systems, and considerable time, particularly when source texts are long or contain complex linguistic structures. Hudáková (2024), for instance, compared source-text pre-editing and MT post-editing and reported that combining pre-editing with post-editing improved translation quality and reduced cognitive load during post-editing, although pre-editing required additional time at the initial stage. Such findings highlight an important practical tension: although pre-editing can improve the conditions under which MT operates, the resources required to perform it may limit its large-scale application.

The emergence of artificial intelligence has opened new possibilities for addressing this limitation. Large language models such as ChatGPT are capable of analyzing and rewriting natural-language texts according to specified linguistic instructions, potentially making them suitable tools for automating or assisting source-text preparation before MT. The use of generative AI for pre-editing raises an empirical question concerning whether AI-generated modifications can actually improve the quality of subsequent MT output.

The study is motivated by the need to explore more efficient approaches to preparing texts for MT. The increasing capabilities of generative AI therefore raise the possibility that some aspects of pre-editing could be automated or supported by AI, potentially reducing the time and effort required to prepare source texts for machine translation. The study seeks to answer the following questions:

1. Does ChatGPT-assisted pre-editing affect the naturalness of Google Translate's English-to-Persian output?
2. Does ChatGPT-assisted pre-editing affect the readability of Google Translate's English-to-Persian output?

2. Literature Review

Controlled languages (CLs) are subsets of natural languages constrained by explicit regulations governing lexicon, syntax, and stylistic conventions. They are fundamentally employed to “mitigate ambiguity and intricacy” in written procedural or descriptive technical documents and can thus be perceived as augmenting commendable authoring practices (Ryan, 2018, p. 289). Their primary objective is to realize cost efficiencies through the simplification and standardization of document production, together with enhanced document usability (Ryan, 2018). As Nyberg et al. (2003) illustrate, controlled language “is an explicitly defined restriction of a natural language that specifies constraints on lexicon, grammar, and style” (p. 245). They (2003) further state that “a common goal of controlled language is a reduction in the number of words that may be used, [...] to reduce or eliminate the use of ambiguous and complex sentence structures” (p. 246).

Controlled language is employed by both humans and computational systems to make source texts more comprehensible and intelligible. These objectives are addressed primarily at the lexical and syntactic levels. As Sánchez-Gijón and Kenny, (2022) illustrate, at the lexical level, controlled language involves the elimination of lexical ambiguity and the reduction of homonymy, synonymy, and lexical complexity. At the syntactic level, it concerns the reduction of syntactic ambiguity, simplification of sentence constructions, constraints on sentence length, and limitations on voice, tense, and other grammatical elements. In this respect, controlled language can be regarded as a form of source-text optimization, as researchers have explored ways of modifying source texts before MT through controlled-language rules and other forms of source-text optimization (Kuhn, 2014; Nyberg et al., 2003; Sánchez-Gijón & Kenny, 2022).

The relationship between controlled language and translation quality has been examined across different generations of MT. Early research primarily investigated controlled-language rules in rule-based and statistical MT environments. For example, O'Brien and Roturier (2007) examined the portability of controlled-language rules across MT systems, while Aikawa et al. (2007) investigated the effect of controlled language on translation quality and post-editing in a statistical MT environment. More recently, Marzouk and Hansen-Schirra (2019) examined the impact of controlled language on neural machine translation and compared its effects across different MT architectures. Taken together, controlled language has been considered not only a means of improving source-text clarity and consistency but also a strategy for facilitating machine translation and potentially improving translation quality and post-editing outcomes.

Pre-editing emerged as a practical extension of controlled-language approaches and refers broadly to modifying the source text before it is submitted to a machine translation system. Pym (1990) is among the early scholars associated with systematic pre-editing for MT, proposing ten key principles for simplifying source texts in order to improve their machine translatability. Thus, the primary goal of pre-editing is to enhance the translatability of the source text and increase the likelihood of producing higher-quality MT output. In more recent work, Sánchez-Gijón and Kenny (2022) explain that pre-

editing may involve either the application of language rules or a relatively short list of simple corrections, such as correcting spelling errors or imposing standard punctuation.

Pre-editing can range from the application of comprehensive controlled-language rules to relatively minor corrections of problematic source-text features. Controlled languages employ explicit guidelines that specify which lexical items, syntactic structures, and stylistic features are permitted or prohibited in a text (O'Brien, 2003). Manual pre-editing has therefore been closely associated with controlled-language approaches for many years (Kuhn, 2014; Nyberg et al., 2003; Pym, 1990), and this approach has been tested with different generations of MT systems, including rule-based and statistical MT (Aikawa et al., 2007; O'Brien & Roturier, 2007). Sánchez-Gijón and Kenny (2022) distinguish several dimensions of pre-editing, including lexical choices, structural and stylistic features, and referential elements. Lexical pre-editing can involve maintaining a consistent register, avoiding uncommon abbreviations, eliminating unnecessary words, and ensuring terminological consistency. Structural and stylistic pre-editing may involve using relatively straightforward sentence structures, complete sentences, parallel constructions, active voice, and less complex verb forms. Referential pre-editing, meanwhile, aims to resolve unclear pronoun references and clarify cultural or contextual references (Sánchez-Gijón & Kenny, 2022, p. 91). These categories provide a useful framework for understanding pre-editing as a systematic process rather than as general proofreading.

Importantly, empirical research has demonstrated that effective pre-editing does not necessarily mean making every sentence shorter or grammatically simpler. Miyata and Fujita (2017), in a large-scale investigation of human pre-editing for black-box MT, collected 6,652 pre-editing instances across three translation directions, two MT systems, and four text domains. Their findings indicated that making the source text shorter and simpler was not always the most important strategy. Instead, "enhancing the explicitness of the meaning of an ST and its syntactic structure" was more important for obtaining improved translations (Miyata & Fujita, 2017, p. 20). This finding is particularly significant because it challenges a simplistic conceptualization of pre-editing as sentence shortening. It suggests that the central concern should be the unambiguous representation of meaning and syntactic relationships. Consequently, pre-editing may involve adding information or restructuring a sentence when such modifications make the intended meaning or syntactic relations more explicit.

Subsequent research has examined pre-editing in neural MT environments and has increasingly explored computational support and automation. Hiraoka and Yamada (2019), for example, investigated pre-editing combined with neural machine translation in the subtitling of TED Talks, demonstrating the relevance of source-text modification to specialized audiovisual translation contexts. These studies demonstrate a gradual movement from manually formulated pre-editing rules toward computationally supported and increasingly automated approaches. Semi-automatic tools, including controlled-language checkers and interactive rewriting assistants, have also been developed to reduce the manual effort involved in pre-editing (Bernth & Gdaniec, 2001; Gulati et al., 2015; Mitamura & Nyberg, 2001). Moreover, fully automatic pre-editing has long been explored, while automatic text simplification techniques have been investigated as a means of enhancing NMT output quality (Mitamura & Nyberg, 2001; Wang et al., 2016).

The efficiency dimension of pre-editing has also received attention. Hudáková (2024) directly compared workflows involving post-editing alone with workflows combining pre-editing and post-editing. In a controlled study involving 20 translation students, the researcher reported that the pre-editing-plus-post-editing group required more time overall but produced fewer translation errors. Hudáková (2024) consequently concluded that "pre-editing enhances translation quality by reducing errors and cognitive load during post-editing, though it requires more time upfront" (p. 42). This finding is important because it reveals a trade-off between the additional time required before MT and the potential reduction in effort required after MT. It also indicates that the effectiveness of a translation

workflow cannot be evaluated solely in terms of the time required to produce the initial machine translation; the overall quality and post-editing effort must also be considered.

Miyata and Fujita (2017) found that the effects of source-text modifications on MT outputs could be difficult to predict, while Hudáková (2024) acknowledged that participant experience and text type could affect the observed relationship between pre-editing, quality, and efficiency. Thus, pre-editing should be understood as a context-dependent intervention whose effectiveness may vary according to language pair, MT architecture, genre, linguistic characteristics of the source text, and the nature of the editing rules employed. The development of semi-automatic and fully automatic approaches further suggests that pre-editing is evolving from a predominantly manual practice toward an increasingly technology-mediated stage of the MT workflow. Taken together, the literature indicates that the fundamental purpose of pre-editing is not simply to simplify source texts, but to make their meaning, structure, terminology, and references sufficiently explicit and consistent for the MT system to process them more effectively.

In the past, pre-editing was mainly performed by human translators or more advanced specialized controlled-language tools, but now generative AI systems can be applied for many more functions in modifying source texts like lexical replacement, syntax modification, clarification of references, or reformulation of ambiguous phrases. The significance of this achievement is that it links the actual research in pre-editing with the current trend of human-machine cooperation in translation.

According to Fan and Wang (2023), AI-based translation research includes translation products, translation processes, and translation teaching with particular attention to the significance of human-computer interaction in modern translation processes. Fan and Wang (2023) note that research increasingly considers how translators interact with MT systems and how technological environments affect translation processes. Within this perspective, AI-assisted pre-editing can be conceptualized not simply as another form of automatic text simplification but as an interactive stage in which human-defined linguistic requirements are operationalized through an AI system.

The emergence of large language models is particularly relevant to pre-editing because such systems can operate at several linguistic levels simultaneously. Unlike rule-based pre-editing tools that typically identify predetermined lexical or syntactic patterns, generative AI can be instructed to modify a text according to a set of linguistic objectives while preserving its propositional content. In this respect, the framework proposed by Sánchez-Gijón and Kenny (2022) provides a useful basis for AI-assisted pre-editing because it organizes source-text modifications into lexical, structural/stylistic, and referential dimensions. The framework can therefore be transformed into explicit prompts that guide an AI system in preparing source texts for MT.

3. Methodology

The study's corpus consisted of 35 English news articles from the *Daily Mail*, *CNN*, and *BBC*, chosen through random sampling. These articles were selected to cover a wide range of journalistic content, featuring complex syntax, idiomatic expressions, and culturally specific references.

The data collection process included three key stages: in the first stage, thirty-five news articles were selected. A small chunk from each article was chosen, typically up to 50 words, with care taken to ensure the chunk is meaningful even out of context. The second stage involved pre-editing with ChatGPT-5 (September 2025). Each article was input into ChatGPT-5 using prompts based on Sánchez-Gijón and Kenny's (2022) pre-editing guidelines. These prompts guided ChatGPT-5 to simplify word choices, refine sentence structures, and clear up referential ambiguities, creating texts that are more machine-translation-friendly. In the final stage, both the original and pre-edited English texts were

translated into Persian using Google Translate. This step allowed for a direct comparison of translation quality between the raw and pre-edited texts.

This study employed the pre-editing framework proposed by Sánchez-Gijón and Kenny (2022), which focuses on three key areas: lexical choices, structure and style, and referential elements.

Lexical guidelines address word choice and vocabulary consistency to optimize MT performance and maintain clarity, e.g. “avoiding lexical shifts in register”, “avoiding uncommon abbreviations”, “avoiding unnecessary words”, “maintaining consistency” (Sánchez & Kenny, 2022, p. 91).

Structural and stylistic guidelines address structural and stylistic features to optimize MT performance and maintain clarity, e.g., “using short and simple sentences”, “ensuring complete sentences”, “using parallel structures in related sentences”, “preferring active voice”, “maintaining a homogeneous style”, “using simple verb tense forms”, “avoiding concatenated verbs” (Sánchez & Kenny, 2022, p. 93 – 94).

Referential elements include: “pronouns”, “possessive determiners”, and “cultural references”, e.g., “clarifying pronoun references”, “addressing the reader consistently”, “clarifying cultural references” (Sánchez & Kenny, 2022, p. 94 – 95).

Based on the aforementioned pre-editing model, a prompt was developed to be used in ChatGPT-5. Different modifications and revisions have led to the following prompt:

You are a professional editor tasked with pre-editing English news articles to optimize their suitability for machine translation into Persian using Google Translate. Your objective is to modify the text according to the following guidelines, ensuring clarity, accuracy, and readability while addressing challenges such as complex syntax, idiomatic expressions, and culturally specific references. Apply the changes systematically and preserve the original meaning of the text.

A) Lexical Guidelines

- Maintain a consistent register: Ensure a uniform style and tone throughout the text. Replace any sudden shifts in formality or complexity with appropriate alternatives that align with the overall register.
- Replace uncommon abbreviations: Substitute uncommon or ambiguous abbreviations with their full forms or widely recognized abbreviations to prevent mistranslation.
- Remove unnecessary words: Eliminate extra words that do not contribute to the meaning, making the text concise and reducing processing complexity for machine translation.
- Ensure consistent terminology: Use the same terms consistently throughout the text. Replace synonyms or varying terminology with a single, uniform term where appropriate to maintain coherence.

B) Structural and Stylistic Guidelines

- Simplify sentence structure: Break down long, complex sentences into shorter, simpler ones to minimize ambiguity and enhance translation accuracy.
- Ensure complete sentences: Verify that all sentences contain all necessary elements (e.g., subject, verb, object). Add missing components if needed to avoid incomplete constructions.
- Use parallel structures: Apply consistent sentence structures in lists, headings, or related sentences to improve readability and translation consistency.
- Prefer active voice: Convert passive voice constructions to active voice wherever possible to provide clarity and facilitate accurate translation.

- Maintain a homogeneous style: Ensure a consistent tone and structure throughout the text, avoiding stylistic shifts that could disrupt translation reliability.
- Use simple verb tenses: Replace complex verb tense forms with present or past simple tenses to improve translation accuracy across languages.
- Avoid concatenated verbs: Simplify clauses containing multiple verbs by breaking them into separate sentences or using simpler verb constructions to reduce translation complexity.

C) Referential Elements

- Clarify pronoun references: Resolve ambiguous pronouns (e.g., "he", "she", "it") by explicitly restating the subject where necessary to eliminate potential misinterpretation.
- Address the reader consistently: If the text directly addresses the reader, maintain a uniform form of address (e.g., "you" or "one") throughout to ensure coherence.
- Clarify cultural references: For culturally specific references or idiomatic expressions, rephrase them into literal, universally understandable language or provide brief explanatory context to aid accurate translation.

Apply these pre-editing guidelines to the provided English text. Submit your output as a revised version of the text, optimized for machine translation into Persian.

The evaluation process utilized a modified version of Taufik (2020, p. 66) three-grade scale, focusing on two primary criteria: naturalness and readability (Table 1).

Table 1. Taufik's (2020) Rating Scale for Assessing the Naturalness and Readability of Translations

Score	Description of Naturalness	Description of Readability
3	It is very natural. Structural naturalness is achieved.	It is very easy to read.
2	It is acceptable. Source language structure is still visible.	The text is difficult to read and requires considerable effort to understand.
1	It is not a good translation.	It is not readable.

Naturalness assesses the extent to which a translation conforms to the conventions of natural language use. A score of 3 indicates fluent and natural language, while a score of 1 reflects strange and unnatural phrasing. Readability measures the ease with which the translated text can be read and understood. The score of 3 reflects clear and straightforward text, while a score of 1 reflects significant difficulty in comprehension.

The results were evaluated by two professional translators using the three-grade translation evaluation scale proposed by Taufik (2020). The unit of analysis was the article, with each article receiving one naturalness score and one readability score under each experimental condition. To enhance the reliability of the assessment, both evaluators had more than five years of professional experience specializing in English-to-Persian translation. Both raters provided their opinions on all translated texts according to the existing standard scale with three points of assessment regarding their naturalness and readability. In general, the raters demonstrated a high level of coherence in their estimations with only rare exceptions. To measure the level of consensus among the two professional translators, weighted Cohen's kappa was applied since the evaluation had an ordinal nature and the level of discrepancies could be evaluated. The resulting value of weighted Cohen's Kappa is $k_w = 0.82$, meaning excellent. If raters had different assessment scores, the relevant translation segments were re-examined with reference to the operational definitions and scoring criteria, and differences in opinions were solved. The value of Kappa was obtained from the independent assessments done by the raters.

The analysis process involved four stages: (1) translating the original texts into Persian, (2) pre-editing the English texts using ChatGPT-5 guided by Sánchez-Gijón and Kenny's (2022) framework (3) translating the pre-edited versions into Persian, and (4) scoring both sets of translations. The comparison of scores between the original and pre-edited translations formed the basis for determining ChatGPT's effectiveness as a pre-editing tool. The results were classified based on the quality of their translations.

4. Results and Discussion

After the Google Translate outputs, both before and after pre-editing, were rated, the results were classified into five categories: 1) improved naturalness and readability; 2) improved readability, same naturalness; 3) improved naturalness, same readability; 4) same readability, same naturalness; and 5) decreased naturalness, same readability.

4.1. Improved Naturalness and Readability

In this category, the output of the Google Translate after being pre-edited by ChatGPT-5 was assessed to be more natural and readable in comparison to the source text without being pre-edited.

Example 1:

Source text (without pre-editing):

Southport killer Axel Rudakubana's unexpected guilty pleas have today caused 'surprise and shock' on Merseyside after the teenage child-killer admitted murdering three young girls at a Taylor Swift-themed dance class.

Machine translation output (without pre-editing):

اعترافات غیرمنتظره اکسل روداکوبانا، قاتل ساوثپورت، امروز پس از اعتراف قاتل نوجوان کودک به قتل سه دختر جوان در یک کلاس رقص با مضمون تیلور سویفت، باعث «تعجب و شوک» در مرسی ساید شد.

(Back Translation: The unexpected admissions of Axel Rudakubana, the Southport killer, today, after the teenage child killer confessed to the murder of three young girls in a Taylor Swift-themed dance class, caused "surprise and shock" in Merseyside.)

Pre-edited text using ChatGPT-5:

Axel Rudakubana, a Southport resident, unexpectedly pleaded guilty today to the murder of three young girls. The incident, which occurred during a Taylor Swift-themed dance class, has caused widespread surprise and shock across Merseyside.

Machine translation output (with pre-editing):

اکسل روداکوبانا، ساکن ساوثپورت، امروز به طور غیرمنتظره‌ای به قتل سه دختر جوان اعتراف کرد. این حادثه که در طول یک کلاس رقص با موضوع تیلور سویفت رخ داد، باعث تعجب و شوک گسترده در سراسر مرسی ساید شده است.

(Back Translation: Axel Rudakubana, a resident of Southport, unexpectedly confessed today to the murder of three young girls. This incident, which occurred during a dance class with the theme of Taylor Swift, has caused widespread surprise and shock throughout Merseyside.)

For the non-pre-edited texts, the Google Translate output received a score of 2 for both naturalness and readability, indicating a moderate level of performance on both criteria. Following pre-editing, the Google Translate outputs received the highest score of 3 for both naturalness and readability. Thus,

the ratings indicate that pre-editing was associated with an improvement in the perceived naturalness and readability of the Google Translate outputs, with both criteria increasing from 2 to 3.

This improvement was observed in 5 instances of the news articles, in which pre-editing enhanced both naturalness and readability. In these cases, the translated texts became more fluent and idiomatic while also becoming easier to read and comprehend.

4.2. Improved Readability, Same Naturalness

In this category, the output of the Google Translate after being pre-edited by ChatGPT-5 was assessed to be more readable and have the same naturalness in comparison to the source text without being pre-edited.

Example 2:

Source text (without pre-editing):

Tesla is recalling nearly all Cybertrucks in the United States to fix an exterior panel that could detach during driving, the latest in a series of call-backs for the pickup truck.

Machine translation output (without pre-editing):

تسلا تقریباً تمام سایبرتراک‌های ایالات متحده را فرا می‌خواند تا یک پنل بیرونی را که می‌تواند در حین رانندگی جدا شود، تعمیر کند، آخرین مورد از یک سری تماس‌های برگشتی برای این وانت.

(Back Translation: Tesla is calling back almost all Cybertrucks in the United States to repair an exterior panel that could become detached while driving, the latest in a series of recalls for this pickup.)

Pre-edited text:

Tesla is recalling almost all Cybertrucks in the United States to fix an exterior panel that could detach while driving. This is the latest recall for the pickup truck.

Machine translation output (with pre-editing):

تسلا تقریباً تمام سایبرتراک‌های ایالات متحده را برای تعمیر یک پانل بیرونی که می‌تواند در حین رانندگی جدا شود، فرا می‌خواند. این آخرین فراخوان برای پیکاپ تراک است.

(Back Translation: Tesla is calling back almost all Cybertrucks in the United States to repair an exterior panel that could become detached while driving. This is the latest recall for the pickup truck.)

The output of the non-pre-edited source text received a score of 2 for naturalness and 1 for readability, indicating moderate naturalness but low readability. In contrast, the output of the pre-edited text received a score of 2 for both naturalness and readability, indicating moderate performance on both criteria. Thus, pre-editing was associated with an improvement in readability, with the score increasing from 1 to 2, while the naturalness score remained unchanged at 2.

This pattern was observed in 9 instances in the news articles. In these cases, pre-editing improved readability only: the translated text became easier to read and comprehend, while its level of naturalness remained unchanged.

4.3. Improved Naturalness, Same Readability

In this category, the output of the Google Translate after being pre-edited by ChatGPT-5 was assessed to be more natural and have the same readability in comparison to the source text without being pre-edited.

Example 3:**Source text (without pre-editing):**

The Department of Defense is investigating “recent unauthorized disclosures of national security information” and plans to use polygraphs as part of the probe, according to a memo published to the department’s website Friday.

Machine translation output (without pre-editing):

بر اساس یادداشتی که روز جمعه در وبسایت این وزارتخانه منتشر شد، وزارت دفاع در حال بررسی «افشای غیرمجاز اخیر اطلاعات امنیت ملی» است و قصد دارد از پلی‌گراف به عنوان بخشی از تحقیقات استفاده کند.

(Back Translation: The Department of Defense is investigating “the recent unauthorized disclosure of national security information” and intends to use a polygraph as part of the investigation.)

Pre-edited text:

The Department of Defense is investigating recent leaks of national security information. It plans to use polygraphs as part of the investigation, according to a memo published on the department’s website on Friday.

Machine translation output (with pre-editing):

وزارت دفاع در حال بررسی افشاگری‌های اخیر اطلاعات امنیت ملی است. بر اساس یادداشتی که روز جمعه در وبسایت این وزارتخانه منتشر شد، این سازمان قصد دارد از پلی‌گراف به عنوان بخشی از تحقیقات استفاده کند.

(Back Translation: The Department of Defense is investigating recent disclosures of national security information. According to a memo published on the department’s website on Friday, the organization intends to use a polygraph as part of the investigation.)

The Google Translate output of the non-pre-edited source text received a score of 2 for naturalness and 3 for readability, indicating moderate naturalness and high readability. In contrast, the output of the pre-edited text received the highest score of 3 for both naturalness and readability. Thus, pre-editing was associated with an improvement in naturalness, which increased from 2 to 3, whereas the readability score remained unchanged at 3.

This pattern was observed in 9 instances in the news articles. In these cases, pre-editing improved naturalness only: the translated texts became more fluent and idiomatic, while their readability remained unchanged.

4.4. Same Readability, Same Naturalness

In this category, the output of Google Translate after being pre-edited by ChatGPT-5 was assessed as equally natural and readable in comparison to the source text without being pre-edited.

Example 4:**Source text (without pre-editing):**

When violence broke out in Syria's coastal area recently, including mass killings of civilians allegedly carried out as revenge for attacks on Syrian security forces, volunteer rescuers quickly came to help.

Machine translation output (without pre-editing):

هنگامی که اخیراً خشونت در منطقه ساحلی سوریه رخ داد، از جمله کشتار دسته جمعی غیرنظامیان که گفته می‌شود به عنوان انتقام حملات به نیروهای امنیتی سوریه انجام شده است، امدادگران داوطلب به سرعت به کمک آمدند.

(Back Translation: When violence recently occurred in Syria's coastal area, including the mass killing of civilians that is said to have been carried out as revenge for attacks on Syrian security forces, volunteer rescuers quickly came to help.)

Pre-edited text:

When violence erupted recently in Syria's coastal area, including alleged mass killings of civilians in retaliation for attacks on Syrian security forces, volunteer rescuers responded quickly.

Machine translation output (with pre-editing):

هنگامی که اخیراً خشونت در منطقه ساحلی سوریه از جمله کشتار دسته جمعی غیرنظامیان در تلافی حملات به نیروهای امنیتی سوریه رخ داد، امدادگران داوطلب به سرعت واکنش نشان دادند.

(Back Translation: When violence recently occurred in Syria's coastal area, including the mass killing of civilians in retaliation for attacks on Syrian security forces, volunteer rescuers quickly responded.)

The ratings indicate that pre-editing did not produce an observable change in either naturalness or readability. The Google Translate output of the non-pre-edited source text received a score of 1 for naturalness and 2 for readability, indicating low naturalness and moderate readability. The output of the pre-edited text received the same scores: 1 for naturalness and 2 for readability. Thus, the ratings remained unchanged for both criteria, suggesting that pre-editing did not improve the naturalness or readability of this particular Google Translate output.

Although the pre-edited version is structurally shorter and uses a more concise sentence, these changes did not result in a measurable improvement according to the rating criteria. The translation remained relatively difficult to read and did not achieve a higher level of naturalness after pre-editing.

This pattern was observed in 10 instances in the news articles. In these cases, pre-editing produced no change in either naturalness or readability, meaning that the translated texts received the same ratings before and after pre-editing.

4.5. Decreased Naturalness, Same Readability

In this category, the output of the Google Translate after being pre-edited by ChatGPT-5 was assessed to be less natural and equally readable in comparison to the source text without being pre-edited.

Example 5:

Source text (without pre-editing):

Warships from Iran, China and Russia kicked off their annual joint exercises in the Gulf of Oman on Monday, showing off their military ties as US President Donald Trump upends longstanding Western alliances.

Machine translation output (without pre-editing):

کشتی‌های جنگی ایران، چین و روسیه رزمایش‌های مشترک سالانه خود را روز دوشنبه در خلیج عمان آغاز کردند تا روابط نظامی خود را در حالی که دونالد ترامپ، رئیس‌جمهور آمریکا، اتحادهای دیرینه غربی را برهم می‌زند، به نمایش بگذارند.

(Back Translation: Warships from Iran, China, and Russia began their annual joint exercises in the Gulf of Oman on Monday to demonstrate their military relations while Donald Trump, the President of the United States, is disrupting longstanding Western alliances.)

Pre-edited text:

Warships from Iran, China, and Russia began their annual joint exercises in the Gulf of Oman on Monday. The drills highlighted their military cooperation as U.S. President Donald Trump disrupts long-standing Western alliances.

Machine translation output (with pre-editing):

ناوهای جنگی ایران، چین و روسیه رزمایش مشترک سالانه خود را از روز دوشنبه در خلیج عمان آغاز کردند. این مانورها بر همکاری نظامی آنها تاکید کردند، زیرا دونالد ترامپ، رئیس جمهور ایالات متحده، اتحادهای طولانی مدت غربی را مختل می کند.

(Back Translation: Warships from Iran, China, and Russia began their annual joint exercises from Monday in the Gulf of Oman. These maneuvers emphasized their military cooperation because Donald Trump, the President of the United States, is disrupting long-term Western alliances.)

The Google Translate output of the non-pre-edited source text received a score of 2 for both naturalness and readability, indicating a moderate level of performance on both criteria. In contrast, the output of the pre-edited text received a score of 1 for naturalness and 2 for readability. Thus, pre-editing was associated with a decrease in naturalness from 2 to 1, whereas the readability score remained unchanged at 2.

This pattern was observed in 2 instances in the news articles. In these cases, pre-editing decreased naturalness while readability remained unchanged, meaning that the translated texts were assessed as less fluent or idiomatic after pre-editing, despite showing no change in their readability.

Table 2 illustrates the average naturalness and readability scores for the original and pre-edited texts.

Table 2. Average Translations Scores

Category	Original Texts	Pre-Edited Texts
Naturalness	1.45	2.34
Readability	1.77	2.71

The data reveal a consistent pattern, i.e., pre-edited texts outperform their original counterparts across all examples. Original texts received scores of 1 or 2, with few instances reaching the highest score of 3. In contrast, pre-edited texts achieved scores of 2 or 3, with a large majority (25 out of 35) earning the maximum score of 3 for both naturalness and readability. The average scores increased from 1.45 to 2.34 for naturalness and 1.77 to 2.71 for readability, indicating a marked improvement in translation quality following ChatGPT-5 pre-editing.

A quantitative summary of the improvements in naturalness and readability resulting from the pre-editing process is presented in Table 3. The original naturalness score was 1.45 out of 3, corresponding to 48.33% of the maximum score, whereas the pre-edited texts achieved a higher score of 2.34 out of 3, corresponding to 78% of the maximum possible score. The mean naturalness score increased by 61.38% relative to the original mean demonstrating the effectiveness of pre-editing in producing more natural and fluent translations. Similarly, readability followed the same trend, with the original texts scoring 1.77 out of 3 (59%), which improved to 2.71 out of 3 (90.33%) after pre-editing; this means naturalness score increased by 53.11 % relative to the original mean. The increases in both metrics emphasize the importance of pre-editing as a useful method for improving translation quality.

Table 3. Summary of Naturalness and Readability Improvements

Measure	Original Score	Pre-edited Score	Percentage (original)	Percentage (pre-edited)	Improvement (%)
Naturalness	1.45	2.34	48.33%	78%	61.38%
Readability	1.77	2.71	59%	90.33%	53.11%

A Wilcoxon signed-rank test showed that pre-editing significantly improved naturalness ($Z = -2.32$, $p = .020$, $r = .39$) and readability ($Z = -3.38$, $p < .001$, $r = .57$). Both effects are medium to large by Cohen's benchmarks, indicating that the gains are not only statistically significant but also practically meaningful.

5. Conclusion

The present study investigated the potential of ChatGPT-5 as an AI-assisted pre-editing tool for English-to-Persian translation of journalistic texts using Google Translate. Guided by the pre-editing principles proposed by Sánchez-Gijón and Kenny (2022), the study focused on three areas of source-text modification: lexical choices, structural and stylistic features, and referential elements. The comparison of the Google Translate outputs produced from the original and ChatGPT-pre-edited source texts showed higher mean scores for both naturalness and readability in the pre-edited condition. These findings suggest that modifying the source text before machine translation may contribute to improvements in the perceived fluency and accessibility of the resulting Persian translations.

The findings demonstrate that pre-editing with ChatGPT-5, guided by Sánchez-Gijón and Kenny's (2022) framework, enhances the quality of machine translations of English journalistic texts into Persian. The improvements in naturalness and readability highlight the potential of AI-assisted pre-editing as a potentially useful tool in translation practice. These findings are consistent with previous studies that have shown the application of controlled languages and pre-editing has a positive impact in terms of improved readability, translatability, accuracy and quality (Hiraoka & Yamada, 2019; Marzouk & Hansen-Schirra, 2019; Miyata & Fujita, 2017; Reuther, 2003).

The findings are also in line with research examining controlled-language effects across different MT architectures. Marzouk and Hansen-Schirra (2019) investigated the effects of controlled-language rules on neural, statistical, rule-based, and hybrid MT systems and found that controlled-language interventions can affect MT quality. Hiraoka and Yamada (2019) likewise examined the combination of pre-editing and neural machine translation in subtitling and demonstrated the relevance of appropriate pre-editing rules to MT output. These studies, together with the present findings, indicate that the relationship between source-text preparation and MT quality deserves attention alongside developments in the MT systems themselves.

Future studies could also extend the investigation to larger and more diverse journalistic corpora, other language pairs, and different MT and generative-AI systems. A particularly productive direction would be to examine which individual pre-editing operations produce the greatest improvements in MT output. Such research could build on the typology of pre-editing operations developed by Miyata and Fujita (2017) and investigate whether specific lexical, syntactic, or referential interventions have different effects on different types of journalistic texts.

This research contributes to the evolving fields of Translation Studies and AI applications in language processing by evaluating ChatGPT's role as a pre-editing tool. Nonetheless, the limitations underscore the need for further exploration to fully realize its scope and impact.

References

- Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M., & Lozano, C. (2007). Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. In B. Maegaard (Ed.), *Proceedings of Machine Translation Summit XI: Papers* (pp. 1–7). <https://aclanthology.org/2007.mtsummit-papers.1>

- Bernth, A., & Gdaniec, C. (2001). MTranslatability. *Machine Translation*, 16(3), 175–218. <https://doi.org/10.1023/A:1019867030786>
- Fan, K., & Wang, C. (2023). Translation studies in the era of AI: Characteristics, fields and significance. *International Journal of Translation and Interpretation Studies*, 3(4), 58–67. <https://doi.org/10.32996/ijtis.2023.3.4.7>
- Gulati, A., Bouillon, P., Gerlach, J., Porro, V., & Seretan, V. (2015). The ACCEPT academic portal: A user-centred online platform for pre-editing and post-editing. In *Proceedings of the 7th International Conference of the Iberian Association of Translation and Interpreting Studies (AIETI)*. Málaga, Spain. <https://www.accept.unige.ch/Products/AIETI-Gulati-al-2015.pdf>
- Hiraoka, Y., & Yamada, M. (2019). Pre-editing plus neural machine translation for subtitling: Effective pre-editing rules for subtitling of TED Talks. In M. L. Forcada, A. Way, J. Tinsley, D. Shterionov, C. Rico, & F. Gaspari (Eds.), *Proceedings of Machine Translation Summit XVII: Translator, Project and User Tracks* (pp. 64–72). European Association for Machine Translation. <https://aclanthology.org/W19-6710/>
- Hudáková, Z. (2024). Comparing the efficiency of source text pre-editing vs. machine translation post-editing. *L10N Journal*, 3(2), 42–59. <https://l10njournal.net/index.php/home/article/view/44>
- Huijsen, W.-O. (1998). Controlled language—An introduction. In *Proceedings of the Second International Workshop on Controlled Language Applications (CLAW 98)* (pp. 1–15). Language Technologies Institute, Carnegie Mellon University.
- Koehn, P. (2020). *Neural machine translation*. Cambridge University Press.
- Kuhn, T. (2014). A survey and classification of controlled natural languages. *Computational Linguistics*, 40(1), 121–170. https://doi.org/10.1162/COLI_a_00168
- Krüger, R. (2022). Integrating professional machine translation literacy and data literacy. *Lebende Sprachen*, 67(2), 247–282. <https://doi.org/10.1515/les-2022-1022>
- Marzouk, S., & Hansen-Schirra, S. (2019). Evaluation of the impact of controlled language on neural machine translation compared to other MT architectures. *Machine Translation*, 33, 179–203. <https://doi.org/10.1007/s10590-019-09233-w>
- Mitamura, T., & Nyberg, E. (2001). Automatic rewriting for controlled language translation. In *Proceedings of the NLPRS2001 Workshop on Automatic Paraphrasing: Theories and Applications* (pp. 1–12).
- Miyata, R., & Fujita, A. (2017). Dissecting human pre-editing toward better use of off-the-shelf machine translation systems. In *Proceedings of the 20th Annual Conference of the European Association for Machine Translation (EAMT)* (pp. 54–59).
- Miyata, R., & Fujita, A. (2021). Understanding pre-editing for black-box neural machine translation. In P. Merlo, J. Tiedemann, & R. Tsarfaty (Eds.), *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 1539–1550). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.eacl-main.132>
- Nyberg, E., Mitamura, T., & Huijsen, W.-O. (2003). Controlled language for authoring and translation. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 245–281). John Benjamins. <https://doi.org/10.1075/btl.35.17nyb>

- O'Brien, S. (2003). Controlling controlled English. In *EAMT Workshop: Improving MT through other language technology tools: Resources and tools for building MT* (pp. 105–114). European Association for Machine Translation. <https://aclanthology.org/2003.eamt-1.12/>
- O'Brien, S. (2006). Controlled language and post-editing. *MultiLingual*, 17(7), 17–19.
- O'Brien, S., & Roturier, J. (2007). How portable are controlled language rules? A comparison of two empirical MT studies. In B. Maegaard (Ed.), *Proceedings of Machine Translation Summit XI: Papers* (pp. 345–352). <https://aclanthology.org/2007.mtsummit-papers.46/>
- Pym, P. (1990). Pre-editing and the use of simplified writing for MT. In P. Mayorcás (Ed.), *Translating and the computer 10: The translation environment 10 years on* (pp. 80–96). Aslib.
- Reuther, U. (2003). Two in one—can it work? Readability and translatability by means of controlled language. In *EAMT Workshop: Improving MT through other language technology tools: Resources and tools for building MT*. European Association for Machine Translation. <https://aclanthology.org/2003.eamt-1.14/>
- Ryan, R. (2018). Controlled Language. In G. Budin, J. Humbley, & C. Laurén (Eds.), *Languages for special purposes: An international handbook* (pp. 289–306). De Gruyter.
- Sánchez-Gijón, P., & Kenny, D. (2022). Selecting and preparing texts for machine translation: Pre-editing and writing for a global audience. In D. Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 81–103). Language Science Press. <https://doi.org/10.5281/zenodo.6759980>
- Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69, 343–418. <https://doi.org/10.1613/jair.1.12007>
- Taufik, A. (2020). Pre-editing of Google neural machine translation. *Journal of English Language and Culture*, 10(2), 64–74. <https://doi.org/10.30813/jelc.v10i2.2137>
- Wang, T., Chen, P., Rochford, J., & Qiang, J. (2016). Text simplification using neural machine translation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), 4270–4271. <https://doi.org/10.1609/aaai.v30i1.9933>