Global Journal of Foreign Language Teaching
Volume 16, Issue 1, (2026) 59-74
ISSN 2301-2595

https://un-pub.eu/ojs/index.php/gjflt/index

Evaluating the translation quality and post-editing efficiency of targoman translator

Hamidreza Abdi*, Translation and Language Sciences, Universitat Pompeu Fabra, Spain, hamidreza.abdi@upf.edu , https://orcid.org/0000-0002-0164-2887

Suggested Citation:
Abdi, H. (2026). Evaluating the translation quality and post-editing efficiency of targoman translator. *Global Journal of Foreign Language Teaching, 16*(1), 59-74. https://doi.org/10.18844/gjflt.v16i1.9754

Received from August 12, 2025; revised from October 18, 2025; accepted from February 18, 2026.
Selection and peer review under responsibility of Prof. Dr. Jesus Garcia Laborda, Alcala University, Spain.

Abstract

With the increasing reliance on machine translation (MT) systems, assessing their linguistic accuracy and practical utility has become critical. In Iran, Targoman Translator is a significant local MT project that requires a critical evaluation of its effectiveness in comparison to emerging AI tools such as ChatGPT. The purpose of this study was to critically evaluate the output quality of Targoman Translator, a domestic MT tool, using Wilss' (1982) matrix, with a focus on syntax, semantics and pragmatics. ChatGPT was also used to establish whether Targoman Translator's translations needed post-editing and whether it might serve as a dependable AI tool for end-users, perhaps replacing human editors. The study included a translation test of 80 statements from Mistrík's (1997) taxonomy of text types, narrative, descriptive, and argumentative, along with idiomatic expressions added by the researcher. Targoman Translator initially translated the statements from English to Persian. Evaluators were provided both the original English and Persian translations, and they assessed the translation quality using a Likert-scale questionnaire. The findings revealed that Targoman performed well grammatically on most statements and was moderately successful semantically. However, it regularly mistranslated idiomatic statements, demonstrating poor performance because it was unable to identify functional dependencies, Additionally, the study found that using ChatGPT for post-editing greatly increased translation accuracy. ChatGPT produced translations that were more in line with the intended meaning by fixing grammatical mistakes and clarifying meaning. Because of this, ChatGPT can be regarded as a strong and trustworthy AI editing tool that is advantageous to end users.

Keywords: Translation Quality, Post-Editing, Targoman Translator, Wilss' matrix of transformation/evaluation

* ADDRESS FOR CORRESPONDENCE: **Hamidreza Abdi ,** Translation and Language Sciences, Universitat Pompeu Fabra, Spain
E-mail address: hamidreza.abdi@upf.edu

1. INTRODUCTION

Translation plays a crucial role in our world; without it, many activities would be impossible. Translation has a big influence on daily life, both in written and spoken forms. Because of the controversy surrounding the evaluation of Translation Quality Assessment (TQA) and its emphasis on the relationship between the source text (ST) and the target text (TT), assessing translation quality has always been a crucial topic in Translation Studies (TS) (Bell, 1993).

TQA is an important subject in TS. It helps translators improve skills and understand their work through evaluators' feedback (Abdi, 2021a). As technology advances, TQA is also used to assess technological tools such as MTs. These are becoming increasingly important. Abdi (2019a) states that technology has recently had a major influence on human life. It enables advancements in science, industry, medicine, and communications.

MTs are among the earliest computer programs developed for language study. They are now widely used for translating texts. In the past, translators used traditional tools like paper dictionaries and typewriters. Today, technological advancements allow them to spend less time and energy and produce more cost-effective translations (Nirenburg, 2003).

The rapid advances in MT technologies have driven dramatic improvements in the global translation industry in recent years. Major international MT tools have received significant attention, but little is known about how these advances have influenced countries like Iran. Given Iran's growing interest in developing domestic MT systems to meet the distinctive linguistic and cultural needs of Persian speakers, a focus on the country is especially pertinent.

This study investigates the Iranian MT landscape by analyzing popular domestic tools: Targoman Translator, Faraazin Translator, and Abadis Translator. Progress has been modest, but these tools have begun to affect user attitudes. Their focus is on translating between Persian and widely used languages such as English. Assessing the quality of these domestic tools is critical to judging whether they are as dependable as their international counterparts.

To evaluate the quality of MT outputs, two primary methods are used: automated and human evaluations. Automated evaluations use metrics such as BLEU, NIST, and METEOR. Human evaluations require direct assessment by individuals (Maučec & Donaj, 2019). Automated evaluations analyze how well MT outputs resemble human translations. They offer a quick but less detailed assessment that is useful for research and development, according to Doherty (2016).

In contrast, many researchers, like Callison-Burch et al. (2007), Freitag et al. (2021), and Liu et al. (2024), consider human evaluation more reliable. They use it to measure automatic metrics. Despite the high cost and time required, human evaluation is the gold standard. It provides a comprehensive analysis of translation elements, including sufficiency, acceptability, and fluency (Han, 2018).

Wilss (1982) suggested a matrix for criticizing translations based on three main paradigms: syntax, semantics, and pragmatics. He admits that his matrix is preliminary but sees it as helpful for creating methodologies for trancism. Abdi (2021b) coined the term trancism as an alternative to Translation Criticism. This matrix lets critics judge translations from different perspectives. Syntactic analysis looks at the grammatical arrangement of words. Semantic analysis examines word meanings, and pragmatic analysis explores suggested meanings within contexts.

Wilss' (1982) matrix can thus help develop practical approaches to empirical translation studies that combine descriptive and evaluative components. He thinks this approach is valuable for evaluating MTs because it includes a comparable translation process from SL to TL.

In this study, Wilss's (1982) matrix is used to analyze Targoman Translator, evaluating its syntactic, semantic, and pragmatic aspects to identify its strengths and weaknesses. The purpose is to assist the developer in improving the product by leveraging its strengths and addressing weaknesses. Furthermore, ChatGPT was used to investigate whether Targoman Translator's outputs require post-editing and whether ChatGPT can be a reliable alternative to human editors.

This study is part of a larger project that aims to evaluate Iranian software developers' performance in creating domestic MT tools. Examples include Targoman Translator, Abadis Translator, and Fastdic Translator. The focus is on quality and the need for post-editing. To meet the objectives of this study, the following questions were raised.

1. What was the output quality of Targoman Translator in terms of syntax, semantics, and pragmatics according to Wilss' (1982) matrix?

2. Was ChatGPT deemed a reliable and valid online AI editor for end-users and translators, potentially replacing human editors?

This study is significant for several reasons. It advances theoretical discussions in TS, evaluates a local MT tool, and examines ChatGPT's influence on translation practices. The findings could affect both academic research on TS and translation work, especially in Iran. The study also considers AI tools like ChatGPT for improving translation quality as an AI editor of post-edited MT results. If ChatGPT is accurate, it could make translation faster and cheaper for clients, increasing access to quality translations.

The findings may lead to a preference for domestic MT tools over international ones in Iran. This could affect the local translation industry's economy and user preferences. The study's input and recommendations should help developers improve their MT systems. This can result in better tools that compete globally and meet the needs of Persian-speaking users. The research may also help translators and educational organizations.

The study helps educational institutions update translation programs by offering training on MT technologies and AI-assisted workflows. It adds to scholarly discourse in TS with new research. Most importantly, it prepares students for changing job markets by teaching them to work with, not against, AI.

2. REVIEW OF THE RELATED LITERATURE

2.1. A Brief History of MT

TS often discusses the growing use of technology (as cited in Melby, 1982) states that workstation translation tools must be upgraded to meet modern needs. meet modern needs. Historically, translation was seen as a mechanical task for anyone with basic language skills (Bassnett, 2002). According to Quah (2006), Holmes highlighted the role of technology in translation during two key stages: the development and deployment of translation systems.

The first MT system was developed fifty years ago. This was a significant technological milestone. Nirenburg (2003) describes MT as the oldest type of computer-assisted translation (CAT) tool. It also plays an important role in language research. He considers MT a key topic in computational linguistics and stresses its significance in current translation practices.

Hutchins (2014) dates the origins of translation mechanization back to the seventeenth century, when it was initially advocated as a tool to help translators. However, it wasn't until the twentieth century that mechanized translation became possible. Over the last five decades, the expanding need for diverse translation services has driven the development of a variety of tools, including MT systems, CAT tools, and translation resources (Quah, 2006). Hutchins and Somers (1992) show that early MT systems were inaccurate and required extensive human intervention to obtain useful translations. Despite ongoing research in the United States, these early systems produced low-quality translations that required extensive human post-editing.

The growth of translation technology, from early MT systems to modern CAT tools, has substantially influenced the translation process. Scholars such as Kay, Quah, and Hutchins have noted that technical improvements have consistently altered translation methods, even as integrating these tools has frequently required human expertise. Early MT systems, while pioneering, were flawed and required ongoing refining by translators. These systems have evolved over time, but their reliance on human input remains a key feature of the industry.

2.2. Targoman Translator

Targoman Translator was developed in 2013 by researchers at Amirkabir University of Technology's Natural Language Processing and Machine Learning Laboratory, led by Dr. Shahram Khadevi, with support from the Communications and Information Technology Research Center (Iran Telecommunication Research Center). Initially, the system was built using open-source statistical MT technologies and was intended to process Persian as part of a research project.

The system was modified and customized from January 2013 to January 2014, including redesigns and upgrades of key engine and service components. According to the developers, the new native translation engine has made substantial advances, delivering 270% faster performance and reducing resource consumption by up to 16 times, while maintaining high accuracy.

Targoman Translator comprises a translation engine and language processing features, including pre-processing, post-processing, word root identification, named entity recognition, transliteration, and language modeling. These tools are mostly designed to be language-independent. Furthermore, Targoman Translator's development architecture is designed to be an adaptable platform that can be utilized for a variety of projects, whether language processing-related or not, according to the developers.

From March 2015 to April 2017, the Targoman system was developed, maintained, and enhanced as part of a local welder project, funded by the Research Institute of Communication and Information Technology (Iran Telecommunication Research Center). Both Targoman Intelligence Processing and Vira Afzar Adan were responsible for the technical and executive tasks. Until July 2022, the Targoman system provided English-Persian translation with a single neural translation engine developed by Targoman and Vira Afzar Adan. It also helped a European corporation translate into nine major languages, including English and French (https://targoman.ir).

Using Wilss' (1982) matrix to generate local MT may improve the accuracy and context of Persian translations. On a worldwide scale, strengthening Targoman Translator through this review could help develop MT for Persian and other regional languages, establishing a standard for dealing with linguistic problems in MT and improving the whole MT field.

2.3. Wills' Matrix for MT Evaluation

Trancism is critical in the evaluation, interpretation, and criticism of translated works. According to Abdi (2024), any criticism of the translation must be based on a thorough examination of both the ST and the TT. This can be accomplished through a systematic model of transcendentalism that gives the critic detailed directions for reaching an objective judgment.

Wilss (1982) developed a paradigm for trancism that contains three major components: syntax, semantics, and pragmatics. This paradigm enables critics to examine translated texts systematically. He thinks this approach is valuable for analyzing MTs because it includes a comparable translation process from SL to TL. This approach entails decoding the SL and encoding the TL utilizing technological and algorithmic methods. However, Wilss notes that this strategy frequently falls short because the algorithmic simulation is often imperfect and remains more of a promise than a reality.

According to Wilss (1982), the early MT assessment methods relied on a lexical approach that paid insufficient attention to the TL's syntactic rules. This approach was mostly unproductive because it failed to address issues that MT technology of the 1950s and early 1960s could not handle. As a result, the emphasis shifted to the second generation of MT systems, which focused on examining text-surface features. Despite efforts to address syntactic difficulties and apply grammatical models, this generation remained insufficient, focusing solely on morphosyntactic rule programming and yielding less satisfying results.

Wilss (1982) contends that understanding a text necessitates three categories of knowledge: real-world knowledge, situational knowledge, and text-internal knowledge. Essentially, it is not only about what is stated, but also how it is interpreted. This distinction illustrates the distance between human translators and MT, which lacks an integrated model of reality. To bridge this gap, MT research must transition from an emphasis on surface structures to deep structures. This shift aims to narrow the qualitative gap between human comprehension and MT's ability to interpret text formally, addressing the mental components of translation that MT cannot mechanically reproduce.

Wilss (1982) also explores the fourth generation of MT, which aims to imitate human-like understanding using AI and conceptual dependence theory. This method employs a semantico-conceptual interlingua and focuses on the sentence's function and communicative purpose, which relate to the pragmatic aspect of translation. This requires a cognitive process that recognizes functional dependencies rather than merely formal structures, resulting in more accurate translations. The author observes that this cognitive process is far more difficult for computers than for human translators, who can use their mental powers to make logically consistent translations.

Wilss (1982) argues that human translators can improve their understanding of a text by using extralinguistic or conceptual information as well as macrocontextual analysis. This capacity supports a consistent link between identifying information and its application in translation. He believes that computers lack this talent, limiting their ability to build cognitive methods required for indirect transfer in translation. To overcome this, he believes that MT systems require more than merely syntactic analysis and synthesis routines. They also require a supplementary semantic program. With these modifications, MT systems may be able to produce adequate translations or perhaps fully automatic high-quality translations (FAHQT).

Wilss' (1982) matrix is useful for evaluating local MT systems, such as Targoman Translator, as it covers three crucial levels: lexical, syntactic, and pragmatic. At the lexical level, it enables precise word choice, which is critical when dealing with Persian region-specific terminology and idioms. The syntactic level addresses grammatical accuracy, allowing the system to produce genuine, fluent translations that adhere to the target language's structure. Finally, the pragmatic level emphasizes contextual and cultural appropriateness, ensuring that translations are not only correct but also operationally adequate for Persian-speaking consumers. Wilss' matrix enables refining local MT systems by ensuring that translations are linguistically sound, culturally relevant, and contextually appropriate.

2.4. Recent Studies in the field

Many researchers have recently focused on MT evaluation, making it a hot topic in academia, particularly among international scholars. Numerous recent studies have examined MT evaluation (e.g., Stahlberg, 2020; Tan et al., 2020; Wang et al., 2022; Rivera-Trigueros, 2022; Zappatore & Ruggieri, 2024). Although there isn't much empirical research in this field yet, some have been undertaken to assess MT quality.

Wang et al. (2024) investigated the quality of MTs for figurative language and presented human evaluation criteria specifically tailored to such translations. Their findings revealed that translations of figurative language varied from literal ones. Liu and Zhu (2023) investigated the MT of political materials, using six top NMT systems to evaluate translation quality. They used both the BLEU and NIST algorithms, as well as a manual evaluation approach known as the Score Ranking System, to compare how well these systems translated Chinese-English political documents. Their findings demonstrated a significant improvement in overall MT quality, with IFLYTEK outperforming the other NMT systems in realistic communication settings.

Ulitkin et al. (2021) used recent automatic evaluation metrics such as BLEU, F1, and TER to compare translations produced by the Google and PROMT neural MT systems. Their findings revealed that these evaluation tools help identify and categorize MT errors, potentially leading to improvements in future translations. Munkova et al. (2020) investigated whether all automated metrics of error rate and accuracy are required for evaluating MTs from synthetic Slovak to analytical English. Their findings revealed that, with the exception of the f-measure, all evaluated textual similarity metrics are critical for assessing the quality of MT output on a sentence-by-sentence basis.

In Iran, there has been little research on local MTs and their usefulness. The majority of studies have focused on Google Translate and its performance, with a few recent studies concentrating on CAT tools and information and communication technology (ICT) tools (Abdi, 2019b, 2022; Taghizadeh & Azizi, 2017). For example, Abdi (2021b) evaluated Google Translate to see if it benefits or threatens human translators. His research revealed that Google Translate generally performed well in translating English into Persian in terms of semantic adequacy and understandability, but lacked fluency. As a result, there is little immediate fear that MT will displace human translators.

Similarly, Bonyadi (2020) researched linguistic alterations in texts translated from Persian to English using Google Translate. The study identified problems with tense, literal translation, repetition, collocations, the deletion of major verbs, word choice, and the use of proper nouns, thereby highlighting Google Translate's shortcomings.

Taleghani et al. (2019) investigated the validity of MT evaluation metrics (MTEMs) for Persian texts and their relationship to human judgments. The study discovered a high correlation between automatic and human evaluations of English-to-Persian translations. The GTM metric was established as the most effective predictor of human judgment.

Vaezian and Pakdaman (2018) compared the quality of two MT systems—Google Translate and SDL Free Translation—based on clarity, accuracy, style, and informativeness. Their findings revealed that both MT systems provided low-quality translations in these areas; Google Translate outperformed SDL Free Translation.

Rahimi et al. (2017) compared Google Neural Machine Translation (GNMT) and Google Machine Translation (GMT) in order to examine the suitability of neural machine translation (NMT) for translating between English and Persian. They concluded that Google Translate's GNMT is a helpful tool for translating Persian, noting that it performed better in terms of semantic correctness, especially for translations from Persian to English.

What distinguishes this study from earlier research is its emphasis on the local MT system, Targoman Translator. It assesses Targoman Translator's output quality in three major areas: syntax, semantics, and pragmatics, which had not been investigated before. Furthermore, the current study examined another crucial element of this local MT: its post-editing strength, which poses a significant obstacle for ChatGPT in determining whether it can be considered a trustworthy online AI editor for end-users.

3. METHOD AND MATERIALS

3.1. Participants

Participants in this study were evaluators tasked with assessing the output quality of Targoman Translator, a domestic MT system. They assessed this system using Wilss' (1982) trancism matrix, which encompasses syntax, semantics, and pragmatics. To choose the assessors, the researcher contacted 20 expert translators at random who had translated between English and Persian. These translations were obtained from the websites www.iacti.ir/members.html and www.proz.com. They were informed about the study's goals, the research topic, and their critical role in achieving the intended results and improving the functionality of the domestic MT.

3.2. Instrumentation

A translation test containing 80 statements was developed for data collection. Mistrík's (1997) definition of text kinds, such as narrative, descriptive, and argumentative, guided the selection of 60 statements. These categories have varying linguistic and stylistic characteristics, making them valuable for assessing the quality of any MT system. Each category contained 20 statements extracted from various sources: narrative statements from J.K. Rowling's (1997) Harry Potter and the Sorcerer's Stone, descriptive statements from Burke and Maxwell's (2012) Lonely Planet: Iran, and argumentative statements from scholarly works by Abdi (2021a), Abdi (2021b), and Dalaslan (2015). Additionally, to evaluate Targoman Translator's pragmatic accuracy, 20 idiomatic phrases were selected from the Book of Idioms (Anonymous, 2011), a collection of common American English expressions.

A panel of experts assessed the test to ensure the statements were relevant and effective in assessing Targoman Translator's quality. These five experts, each with over 5 years of experience in translating teaching, provided feedback on the test's domain coverage, relevance, and clarity. Some statements were rearranged or replaced in response to their suggestions.

For the pilot phase, 15 translators with experience and specialization similar to those of the evaluators were chosen. This quantity was deemed sufficient to provide relevant insights within the study's time constraints. These 15 translators were given the test, and the results were compared to the total scores. The average correlation coefficient was 0.76, with a range of 0.54 to 0.83. They all fell within the allowed range of ±0.50 to ±1, demonstrating the test's reliability. Furthermore, the validity of the test was confirmed by the fact that all of the items had p-values below 0.05 ($p < 0.05$).

3.3. Data Collection and Analysis

For data collection, the translation test was first translated from English to Persian using Targoman Translator. The Persian translations and the original 80 English statements were then provided to evaluators, who evaluated Targoman Translator's output quality using three criteria: syntax, semantics, and pragmatics, as included in Wilss's (1982) matrix.

A Likert-scale questionnaire with five response options—wrong, inappropriate, undecidable, correct, and appropriate—was given to the evaluators. These answers were weighted on a scale from 1 to 5 to represent the assessors' assessments of the translation's quality. A score of 1 (Wrong) indicated the lowest quality, a score of 2 (Inappropriate) suggested the worst but marginally better translation, a score of 3 (Undecidable) indicated the evaluator's inability to decide on the quality, a score of 4 (Correct) indicated a good translation, and a score of 5 (Appropriate) indicated the best quality. This metric assessed Targoman Translator's ability to provide accurate translations across syntax, semantics, and pragmatics. The agreement between the participants was measured using Fleiss' kappa (k).

Furthermore, ChatGPT was used to post-edit the statements that the evaluators deemed wrong or inappropriate to determine whether the quality of these translations could be improved. These remarks, along with their Persian translations, were sent to ChatGPT for post-editing, and the resulting versions were assessed for quality improvements. The original, problematic Persian translations were compared with their post-edited counterparts to identify changes in syntax, semantics, and pragmatics.

For data analysis, the percentage and mean scores for each category were computed and illustrated in tables. Furthermore, the Wilcoxon signed-rank test was employed to evaluate ratings between two related situations, as it is effective for ordinal data and when parametric tests, such as the paired t-test, are inapplicable. This test helped ensure that any observed improvements were not due to chance, but rather meaningful. The test was run individually for the syntactic, semantic, and pragmatic paradigms to see if evaluators' ratings of Targoman Translator's Persian translations were consistent within each paradigm.

4. RESULTS AND DISCUSSION

4.1. Inter-rater reliability test

Fleiss' kappa (k) is designed for multiple raters and accounts for chance agreement. Because 20 raters conducted the test, consensus was reached. Following the reliability analysis using SPSS Statistics, Table 1 displays the k value and the underlying data. The test findings revealed that the k value was.735, indicating a strong level of agreement. Additionally, the p-value is less than 0.05 ($p < 0.05$). This indicates that the k coefficient is significantly different from zero.

Table 1

Summary of interrater reliability tests among the 20 raters

Overall Kappa						
k	Asymptotic Standard Error	*z*	*p*	Lower 95% Asymptotic CI Bound	Upper 95% Asymptotic CI Bound	
Overall	.735	.72	6.428	.000	.219	.704

4.2. Syntactic Evaluation.

As shown in Table 1, a significant majority (97%) of the descriptive Persian translations were rated as adequate, with 49% being classified as correct and 47% as appropriate; in contrast, the evaluators agreed that approximately two-thirds (69%) of the argumentative Persian translations were correct, while less than two-thirds (26%) were deemed appropriate; and overall, the evaluators agreed that a large proportion of the narrative statements were syntactically correct (80%), with 54% being deemed correct and 26% being considered appropriate, while a small percentage (20%) were deemed inadequate.

Table 2

Percentages of the Evaluators to Syntactic Evaluation of the 60 Statements

Type of Statements	Wrong	Inappropriate	Undecidable	Correct	Appropriate	*N*	
	%	%	%	%	%	%	*M*
Narrative	4.0	11.0	5.0	54.0	26.0	100.0	3.84
Descriptive	-	-	4.0	49.0	47.0	100.0	4.4
Argumentative	-	-	5.0	69.0	26.0	100.0	4.21
Total	1.0	4.0	5.0	57.0	33.0	100.0	4.15

A one-sample Wilcoxon signed-rank test was used to see if there was a significant relationship between the evaluators' ratings and the syntactic soundness of the Persian translations provided by Targoman Translator for each type of statement. Table 2 displays the test findings, indicating that all statement categories had p-values below.05 ($p < .05$). This, together with mean scores above the midpoint (2.5) for each type of sentence, suggests that evaluators agreed with the grammatical accuracy of the Persian translations. The mean scores were as follows: narrative statements (MN = 3.84), descriptive statements (MD = 4.4), and argumentation statements (MA = 4.21), all of which demonstrated above-average syntactic correctness.

Furthermore, Table 2 shows that the p-value for the overall test of the sixty Persian translations was 0.0, which is less than .05 ($p < .05$). The mean score for all statements combined was 4.15 out of 5, which is above the mid-point of the answers to the sentences (4.15 > 2.5). This indicates that, overall, the evaluators significantly agreed with the syntactic correctness of the Persian translations provided by Targoman Translator.

Table 3

One Sample Wilcoxon Signed Ranks Test for Syntactic Evaluation of the 60 Statements

Statement Types	*N*	*MDN*	*p*
Narrative	20	4	.000
Descriptive	20	4	.002
Argumentative	20	4	.001
Total	60	4	.000

4.3. Semantic Evaluation

The evaluators determined that 60% of the narrative statements were semantically adequate, 49% correct, and 11% appropriate. Most Persian translations were classified as inappropriate (12%), undecidable (5%), correct (53%), or appropriate (31%). The evaluators assessed the bulk of the argumentative translations (77%) as appropriate. Overall, the evaluators found that more than half (59%) of the 60 translations were semantically adequate.

Table 4

Percentages of the Evaluators for the Semantic Evaluation of the 60 Statements

Type of Statements	Wrong	Inappropriate	Undecidable	Correct	Appropriate	*N*	
	%	*%*	*%*	*%*	*%*	*%*	*M*
Narrative	5.0	28.0	8.0	49.0	11.0	100.0	3.33
Descriptive	-	12.0	5.0	53.0	31.0	100.0	4.01
Argumentative	-	12.0	11.0	65.0	12.0	100.0	3.75
Total	2.0	17.0	8.0	56.0	18.0	100.0	3.69

A one-sample Wilcoxon test was used to examine the extent to which evaluators agreed with the semantic adequacy of each type of statement, as well as whether there was a significant relationship between evaluators' judgments and the semantic adequacy of Persian translations. The narrative, descriptive, and argumentative Persian translations had p-values of.000, .000, and .003, respectively, all less than.05 ($p < .05$) (see Table 4). This suggests a strong relationship between the evaluators' ratings and the semantic soundness of the translations. The mean scores for these translations were likewise higher than the theoretical mean/median of 2.5, indicating above-average semantic adequacy: narrative translations (MN = 3.33), descriptive translations (MD = 4.01), and argumentative translations (MA = 3.75).

Table 4 indicates that all 60 Persian translations had a p-value <.05. The aggregate mean score of 3.69 exceeded the predicted mean/median of 2.5. This means that, overall, evaluators agreed with the semantic adequacy of Targoman Translator's Persian translations, indicating greater-than-average semantic adequacy.

Table 5

One Sample Wilcoxon Signed Ranks Test for Semantic evaluation of the 60 Statements

Statement Types	*N*	*MDN*	*p*
Narrative	20	4	.000
Descriptive	20	4	.000
Argumentative	20	4	.003
Total	60	4	.000

4.4. Pragmatic Evaluation

According to Table 5, the evaluators disagreed with the pragmatic correctness of all 20 idiomatic statements (100%) produced by the Targoman Translator.

Table 6

Percentages of the Evaluators for Pragmatic Evaluation of the 20 Statements

Type of Statements	Wrong	Inappropriate	Undecidable	Correct	Appropriate	*N*	
	%	*%*	*%*	*%*	*%*	*%*	*M*
Idiomatic	80.0	20.0	-	-	-	100.0	1.2

A one-sample Wilcoxon signed-rank test was used to determine whether there was a significant relationship between the evaluators' ratings and the pragmatic adequacy of the Persian translations provided by Targoman Translator. Table 6 indicates that the p-values for idiomatic statements were less than.05 ($p < .05$). This, together with the fact that the mean score for this statement type was lower than the theoretical mean/median, indicates that evaluators did not find the idiomatic translations pragmatically appropriate.

Table 7

One Sample Wilcoxon Signed Ranks Test for Pragmatic Evaluation of the 20 Statement

Statement Types	*N*	*MDN*	*p*
Idiomatic	20	2	.002

4.5. Discussion

The findings indicate that Targoman Translator performed well in terms of grammar when translating the majority of narrative, descriptive, and argumentative sentences into Persian. In other words, the translations with the highest quality ratings were accurate and appropriate. This means that the statements were translated with correct grammar in general, though a few translations were awkward due to Targoman Translator's grammatical order (SOV) and subject-verb agreement rule deficiencies. For example, the English sentence "Harry woke up at five o'clock the next morning" was rendered as "ساعت پنج صبح روز بعد هری از خواب بیدار شد," which is poor grammar. This is because the Persian grammar order (SOV) was not followed. The exact translation is هری ساعت پنج صبح روز بعد از خواب بیدار شد, where هری (Harry) is the subject and comes first, rather than صبح روز بعد ساعت پنج (at five o'clock the next morning).

The results show that Targoman Translator was quite successful in conveying the meaning of the 60 statements. That is to say, the semantic quality of the Persian translations generated by Targoman Translator was a bit above average. For instance, the translation of Mr. and Mrs. Dursley, of number four, Privet Drive, was not intelligible because it was translated literally as خانم و آقای دورسلی، مرد شماره چهار پرایوت درایو. This word-for-word translation didn't quite clearly state where Mr. and Mrs. Dursley live. After ChatGPT post-edited the sentence, the correct translation was آقای و خانم دورسلی، ساکن شماره چهار پرایوت درایو, which exactly shows their address.

For another example, descriptive sentence glassware was usually green, lapis lazuli, light blue or clear with a tinge of yellow, and decorationswere cut into the glass was translated into ظروف شیشه ای معمولا سبز، لاجوردی، آبی روشن یا شفاف با ته رنگ زرد بودند و تزیینات آن ها به شیشه می رسید. This translation did not convey the meaning correctly because it failed to transfer the meaning of the coordinate clause, and decorations were cut into the glass. When ChatGPT post-edited the sentence, the correct translation that correctly conveyed the meaning was:و تزئینات به صورت حکاکی روی شیشه اجرا می‌شد

In the argumentative sentence, the translations appear questionable as there are no English onomatopoeic words translated in the target text, and any English onomatopoeic expressions have been translated as هیچ عبارت آنومالی انگلیسی, which was not quite understandable to target readers. This occurred as the word onomatopoeic remained in English, which may be unfamiliar to the majority of the readers. When ChatGPT retranslated the translation, the phrase "any English onomatopoeic expressions" was rendered well as "اصطلاحات صوتی انگلیسی," which greatly improved readers' clarity.

The results of the pragmatic evaluation indicated that Targoman Translator was poor, as all such statements were translated incorrectly with respect to pragmatic adequacy. For example, the idiom the graveyard shift in the sentence Benjamin isn't happy now that he's working the graveyard shift was translated literally as شیفت قبرستان, which confused target readers. This is because graveyard shift refers to the night shift as used in an employment setting, which Targoman Translator failed to recognize. Once ChatGPT analyzed the phrase, it accurately recognized the implied meaning, and the phrase was post-edited to شیفت شب (the night shift). The sentence's new translation became بنجامین اکنون که در شیفت شب کار می‌کند، خوشحال نیست, which maintained the correct meaning.

In another instance, the idiomatic expression "a knockout" in the sentence "Jan's daughter is a knockout" was literally translated as "ناک اوت" by Targoman Translator, altering the sentence's meaning. That is because a knockout is someone who is very beautiful. A more appropriate translation of the word would be فوق‌العاده جذاب (extremely attractive), which ChatGPT correctly chose and used in the post-edited text.

Overall, the results indicate that Targoman Translator performed well with grammatical constructions across the vast majority of the 60 sentences. Its performance was also comparatively inconsistent in semantics and completely lacking in pragmatics, especially with idiomatic expressions. That is, more than half of the 60 sentences were translated with an acceptable degree of quality, but the translation of idiomatic items and participle constructions was problematic. This is because these items are inherently implicit and require greater linguistic complexity, according to Wilss (1982).

These weaknesses point to Targoman Translator's reliance on literal translation, which often produces awkward or incorrect translations in the TL. Furthermore, the MT struggled to preserve the original meaning of complex phrases with cultural or idiomatic significance. This weakness was evident in examples such as translating across the board or the graveyard shift, where literal translation confuses the target readership.

The limitations of this study are that it is based on a single MT tool, Targoman Translator, which limits the generalizability of the findings to other MT systems. A second limitation is that post-editing relies on ChatGPT, which may not reflect the subtle judgment of human editors, particularly when dealing with cultural and contextual subtleties. In addition, the limited number of text types (narrative, descriptive, argumentative, and idiomatic expressions) may not encompass the wide variety of texts that one is confronted with in real translation practice. Apart from this, the study is primarily focused on syntax, semantics, and pragmatics, perhaps overlooking other important considerations, such as fluency and domain-specific translation accuracy. Future research can address these constraints by comparing Targoman Translator with other MT tools across different languages and text types to yield more generalizable results. A closer analysis of human versus AI post-editing could compare the effectiveness of AI tools like ChatGPT with that of human translators, considering quality, efficiency, and user satisfaction. Extending the range of text types and incorporating diverse evaluation frameworks would provide a more comprehensive analysis of MT quality.

5. CONCLUSION

The aim of the present study was to critically examine the syntax, semantics, and pragmatics of the output of Targoman Translator, a domestic MT system, according to Wilss's (1982) matrix of trancism, in order to determine its strengths and weaknesses. The present study is carried out as part of a research project examining the success of Iranian software developers in developing three domestic MT systems—Targoman Translator, Abadis Translator, and Fastdic Translator—based on translation quality and post-editing.

The results showed that Targoman Translator performed well grammatically in translating a significant percentage of the 60 sentences into Persian. It also demonstrated a fair ability for semantically accurate translations. The output quality was typically good for descriptive and argumentative sentences and quite good for narrative. The results indicated that the Targoman Translator performed grammatically well, translating most of the 60 statements into Persian. Its performance was also quite good at generating semantically correct translations. Its quality was overall acceptable for descriptive and argumentative statements and quite acceptable for narrative statements.

As thAs the results of the pragmatic evaluation revealed, Targoman Translator, in general, failed to translate idiomatic expressions and participle structures well, as all such sentences were translated inappropriately. The main weakness of Targoman Translator is its incapacity to process implied meaning, which requires higher-level cognitive processes. Wilss (1982) states that in order to reproduce these kinds of items efficiently in TL, one needs to "recognize functional dependencies rather than formal structures" (p. 241). This lack means that post-editing is necessary for the translations produced by Targoman Translator, despite the developer's claim that it has post-editing capabilities. Furthermore, the results demonstrated that post-editing with tools like ChatGPT significantly improved translations by correcting errors, ensuring grammatical consistency, and better communicating the intended message. This indicates that ChatGPT can serve as a capable and reliable

AI editor for translators and end-users. It also highlights the crucial roles of contextual knowledge and cognitive processing in translation, especially in participle constructions, where literal translation cannot fully convey the meaning of the SL statement. Based on the study findings, the following could be recommended to domestic MT developers, professional translators, and institutions:

To domestic MT developers, although the linguistic ability of Targoman Translator was remarkable, the area for improvement is pragmatics, specifically the translation of idiomatic statements. Developers need to enhance the software's capacity to interpret context and semantics to improve the quality of advanced-level translations. One such solution is for local MT developers to integrate with or collaborate with advanced AI models like ChatGPT to create hybrid translation products or enhance post-ediFurther, developers must focus on building a variety of text types the software is fine-tuned to process so that it can be more generally applicable across different domains and genres.domains and genres.

For educational institutions, post-editing courses must be part of their curriculum so they can assist future translators in hybrid processes that combine human and MT to convey implied meanings, such as proverbs and idioms. This will benefit students in both manual and machine-assisted translation. Furthermore, institutions should encourage student research to improve local MT systems, particularly in areas such as pragmatics and cognitive processing, which are imperative for developing more sophisticated translation systems.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

Abdi, H. (2019a). *Translation and technology: A study of Iranian freelance translators*. LAMBERT Academic Publishing.

Abdi, H. (2019b). The status of ICT employment among Iranian freelance translators. *International Journal of Innovation and Research in Educational Sciences, 6*(3), 339–349.

Abdi, H. (2021a). Examining the appropriateness of Reiss's functionalist-oriented approach to translation criticism. *Theory and Practice in Language Studies, 11*(5), 561–567. https://doi.org/10.17507/tpls.1105.155

Abdi, H. (2021b). Considering machine translation (MT) an aid or a threat to the human translator: The case of Google Translate. *Journal of Translation and Language Studies, 2*(1), 19–32. https://doi.org/10.48185/jtls.v2i1.1222

Abdi, H. (2022). Inquiry into students' familiarity with computer-assisted translation tools. *International Journal of New Trends in Social Sciences, 6*(2), 53–63. https://doi.org/10.18844/ijss.v6i2.6537

Abdi, H. (2024). Negative analytic of the Persian translation of Atwood's *The Blind Assassin* using Berman's model of criticism. *International Journal of Philology and Translation Studies, 6*(1), 27–40. https://doi.org/10.55036/ufced.1434250

Anonymous. (2011). *Book of idioms*. Defense Language Institute English Language Center.

Bassnett, S. (2002). *Translation studies* (3rd ed.). Routledge.

Bell, R. T. (1993). *Translation and translating: Theory and practice* (2nd ed.). Longman.

Bonyadi, A. (2020). Exploring linguistic modifications of machine-translated literary articles: The case of Google Translate. *Journal of Foreign Language Teaching and Translation Studies, 5*(3), 89–104. https://doi.org/10.22034/efl.2020.250576.1057

Burke, A., & Maxwell, V. (2012). *Lonely Planet: Iran*. Lonely Planet Publications.

Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., & Schroeder, J. (2007). (Meta-)evaluation of machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation* (pp. 136–158).

Dalaslan, D. (2015). *An analysis of the English translation of Erzurum folk riddles in the light of Raymond Van den Broeck's translation criticism model* (Master's thesis, Hacettepe University).

Doherty, S. (2016). The impact of translation technologies on the process and product of translation. *International Journal of Communication, 10*, 947–969.

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics, 9*, 1460–1474. https://doi.org/10.1162/tacl_a_00437

Han, L. (2018, September 19). Machine translation evaluation resources and methods: A survey [Conference presentation]. Ireland Postgraduate Research Conference, Dublin, Ireland.

Hutchins, J. (2014). *The history of machine translation in a nutshell*.

Hutchins, J., & Somers, H. (1992). *An introduction to machine translation*. Academic Press.

Liu, S., & Zhu, W. (2023). An analysis of the evaluation of translation quality of neural machine translation application systems. *Applied Artificial Intelligence, 37*(1), 1–28. https://doi.org/10.1080/08839514.2023.2214460

Liu, T., Lo, C., Marshman, E., & Knowles, R. (2024, September 30). Evaluation briefs: Drawing on translation studies for human evaluation of MT [Conference paper]. *Association for Machine Translation in the Americas Conference*, Chicago, USA.

Maučec, M. S., & Donaj, G. (2019). *Machine translation and the evaluation of its quality*. IntechOpen.

Melby, A. (1982). The translator workstation. In J. Newton (Ed.), *Computers in translation: A practical appraisal* (pp. 147–165). Routledge.

Mistrík, J. (1997). *Štylistika*. SPN.

Munková, S., Hájek, P., Munk, M., & Skalka, J. (2020). Evaluation of machine translation quality through error rates and accuracy metrics. *Procedia Computer Science, 171*, 1327–1336. https://doi.org/10.1016/j.procs.2020.04.142

Nirenburg, S. (2003). Introduction. In S. Nirenburg, H. Somers, & Y. Wilks (Eds.), *Readings in machine translation* (pp. 3–11). MIT Press.

Quah, C. K. (2006). *Translation and technology*. Palgrave Macmillan.

Rahimi, M. Z., Madayenzadeh, M., & Alizadeh, M. (2017). A comparative study of English-Persian neural machine translation. *Iranian Journal of Applied Language Studies, 9*, 279–286.

Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: A systematic review. *Language Resources and Evaluation, 56*, 593–619. https://doi.org/10.1007/s10579-021-09537-5

Rowling, J. K. (1997). *Harry Potter and the sorcerer's stone*. Scholastic Press.

Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research, 69*, 343–418. https://doi.org/10.48550/arXiv.1912.02047

Taghizadeh, M., & Azizi, M. (2017). Exploring computer-aided translation competences of Iranian translators. *International Journal of English Language & Translation Studies, 5*(1), 78–87.

Taleghani, M., Pazouki, E., & Ghahraman, V. (2019). Correlation of machine translation evaluation metrics with human judgment in Persian. *Journal of Language and Translation, 9*(3), 43–55.

Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, S., & Liu, Y. (2020). Neural machine translation: Methods, resources, and tools. *AI Open, 1*, 5–21. https://doi.org/10.1016/j.aiopen.2020.11.001

Ulitkin, I., Filippova, I., Ivanova, N., & Poroykov, A. (2021). Automatic evaluation of machine translation quality of scientific texts. *E3S Web of Conferences, 284*, 1–12. https://doi.org/10.1051/e3sconf/202128408001

Vaezian, H., & Pakdaman, A. (2018). Comparative study of Persian translations by Google Translate and SDL Free Translation. *Translation Studies Quarterly, 15*(60), 61–78.

Abdi, H. (2026). Evaluating the translation quality and post-editing efficiency of targoman translator. *Global Journal of Foreign Language Teaching, 16*(1), 59-74. https://doi.org/10.18844/gjflt.v16i1.9754

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in machine translation. *Engineering, 18*, 143–153. https://doi.org/10.1016/j.eng.2021.03.023

Wang, S., Zhang, G., Wu, H., Loakman, T., Huang, W., & Lin, C. (2024). MMTE: Corpus and metrics for evaluating machine translation quality of metaphorical language. *arXiv*. https://doi.org/10.48550/arXiv.2406.13698

Wilss, W. (1982). *The science of translation*. Shanghai Foreign Language Press.

Zappatore, M., & Ruggieri, G. (2024). Adopting machine translation in healthcare: A multi-criteria review. *Computer Speech & Language, 84*, 1–46. https://doi.org/10.1016/j.csl.2023.101582