BD CENTER 2013

Global Journal of Computer Sciences: Theory and Research
ISSN 2301-2587

Global Journal of Computer Sciences Theory and Research

Volume 15, Issue 1, (2025) 1-14

www.gjcs.eu

The level of achievement of Iranian developers in local machine translation systems: Focusing on quality and post-editing

Hamidreza Abdi[a1], Universitat Pompeu Fabra, Carrer de la Mercè, 12, Ciutat Vella, 08002 Barcelona, Spain, hamidreza.abdi@upf.edu

Suggested Citation:

Abdi, H. (2025). The level of achievement of Iranian developers in local machine translation systems: Focusing on quality and post-editing. *Global Journal of Computer Sciences: Theory and Research*, *15*(1), 1-14. https://doi.org/10.18844/gjcs.v15i1.9874

Received from; October 30, 2024, revised from; February 12, 2025 and accepted from March 15.

Selection and peer review under the responsibility of Assist. Prof. Dr. Ezgi Pelin YILDIZ, Kars Kafkas University, Department of Computer Technology, Turkey.

Abstract

Localized machine translation technologies have been developed to address region specific linguistic needs, yet they remain less widely adopted than dominant global systems. Despite their potential, limited evidence exists regarding their translation quality, particularly in complex communicative contexts. This study addresses this gap by evaluating the performance of three domestic machine translation systems and examining their strengths, limitations, and overall reliability in English to Persian translation. A set of eighty test statements representing narrative, descriptive, argumentative, and idiomatic forms was used to assess linguistic accuracy across grammatical, semantic, and pragmatic dimensions. The findings indicate that although the systems generated broadly intelligible translations, they showed recurrent weaknesses in subject verb agreement, tense consistency, and especially pragmatic rendering, where idiomatic expressions posed substantial challenges. One system demonstrated comparatively stronger syntactic and semantic performance, yet none provided consistently adequate output without the need for substantial post editing. For comparative purposes, translations edited through a large-scale language model displayed improved grammatical correctness and reduced error rates, although occasional literal renderings required minor adjustments. The overall analysis suggests that current local technologies do not yet deliver dependable high quality translations independently and that significant technological refinement is required to enhance their usability for professional or high stakes applications.

Keywords: Artificial intelligence; language technology; machine translation; translation evaluation; translation quality.

* ADDRESS FOR CORRESPONDENCE: Hamidreza Abdi, Universitat Pompeu Fabra, Carrer de la Mercè, 12, Ciutat Vella, 08002 Barcelona, Spain. *E-mail address*: hamidreza.abdi@upf.edu

1. INTRODUCTION

Machine translation (MT) is a modern computer-based tool that, as Lin and Chien (2009) explain, makes simple substitutions using important translated words. This assists non-native speakers in understanding the main material of a foreign language that needs to be evaluated. Furthermore, MT enables translators to do translations more quickly and at a cheaper cost. According to Granell (2014), clients are increasingly demanding higher-quality translations in shorter time frames. According to Macías et al., (2020), MT is now considered a mature technology in the translation industry. However, as Abdi (2021b) points out, the main problem concerning MT is its ability to provide high-quality results. In other words, MT must create translations that are *competent*, *fluent*, and *comprehensible* without the need for further editing, which is a challenge that recent MT quality-assessment studies continue to highlight (Rivera-Trigueros, 2022; Araghi & Palangkaraya, 2024)

Traditionally, translation was done using manual dictionaries and dictation, which required a significant amount of time and energy. Without technology, translators had to use typewriters or record their voices for later transcription by typists (Melby, 2002). According to Granell (2014), the invention of MT has made translation faster and more economical. MT mechanically translates texts between languages. This has led to the creation of many MT tools, such as Google Translate and Microsoft Bing Translator, which are free to use and assist translators in producing accurate translations more quickly and affordably, with neural MT becoming increasingly dominant in recent years (Lee et al., 2023)

Along with global MT systems, local MT tools such as Targoman Translator and Abadis Translator have emerged in Iran, indicating a considerable advance in translation technology. These domestic tools were designed for local users as an alternative to global options. Because these local MT systems are relatively new, their quality needs to be rigorously evaluated to improve their performance and increase their use by both local and foreign users; a concern similarly raised in contemporary MT evaluation research (Wang, 2023)

There are two main methods for evaluating MT systems: *automatic* and *human* evaluations. Automated evaluation uses metrics such as BLUE, NIST, and METEOR, whereas human evaluation involves individuals assessing translations. There have been numerous discussions concerning which method is more reliable. Despite the speed, convenience, and cost-effectiveness of automated measures, some researchers feel that human evaluation can be replaced by automated evaluation since they are "very reliable even when just one reference translation is available" (Coughlin, 2003). However, recent work demonstrates that traditional automatic metrics often underperform when compared with human judgments, especially on complex or idiomatic texts, suggesting that automated metrics cannot fully replace human evaluation in many contexts (He et al., 2025).

However, many researchers place more reliance on human evaluation. Han (2016) portrays human evaluation as being widely accepted as the *gold standard*, despite being more time-consuming and costly. This is because it allows for a more precise assessment of factors such as *sufficiency*, *fluency*, and *acceptability*. According to Callison-Burch et al. (2007), human evaluation is a superior and more authoritative alternative to automated evaluation due to "substantial variations in the results of human and automatic evaluations and much higher inter-annotator agreement".

Wilss (1982) suggested a matrix for evaluating translations using three essential paradigms: *syntax*, *semantics*, and *pragmatics*. Although Wilss (1982) acknowledged that his matrix was still in its early phases, he believes it has the potential to help create trancism, a term coined by Abdi (2021a) and used as an alternative to Translation Criticism. The matrix enables a multidimensional examination of translations. It allows evaluators to examine phrase structure (syntax), word meanings (semantics), and inferred meanings in certain settings (pragmatics). As Wilss (1982) explains, his matrix provides a realistic approach to conducting empirical studies in the field, which naturally includes both descriptive and evaluative parts.

Wilss (1982) included a chapter on MT in his book *The Science of Translation* and used the same matrix to evaluate it. Using his trancism/evaluation framework, a study effort was conducted to evaluate the quality of three MT systems created by Iranian software developers: Targoman Translator, Abadis Translator, and

Fastdic Translator. The aim was to discover each system's strengths and shortcomings so that shortcomings may be reduced and quality improved. The project also examined the performance of different MTs in translating sentences from English into Persian to discover which one performed best and whether they could be trusted by Iranian end-users. Furthermore, ChatGPT was employed to determine whether these MT outputs required post-editing and whether ChatGPT may serve as a viable replacement for human editors for both end-users. To meet the objectives of the current study, the following questions are raised:

1. How did Targoman, Abadis, and Fastdic Translators perform in terms of syntax, semantics, and pragmatics using Wilss' (1982) matrix?
2. Which MT (Targoman Translator, Abadis Translator, or Fastdic Translator) produces superior quality translations?
3. Are Targoman Translator, Abadis Translator, and Fastdic Translator trustworthy tools for Iranian translators and users?
4. Is ChatGPT a dependable and legitimate online AI editor for end-users, capable of replacing human editors?
5. Did Iranian software developers meet acceptable standards for quality and post-editing?

The study provides an objective assessment of Iranian MT systems such as Targoman, Abadis, and Fastdic. Local MT systems in countries like Iran have not received as much attention as global ones, such as Google Translate and Microsoft Bing. This study advances the area of Translation Studies by analyzing translations at the syntax, semantics, and pragmatic levels. It also offers helpful suggestions for improving technology, cutting down on mistakes, and boosting user confidence.

The current study also examines ChatGPT as a potential replacement for human editors that could have a considerable impact on translation and editing in the future. If effective, ChatGPT could reduce the cost and improve accessibility of translation by doing away with the need for expensive human post-editing.

The study's findings should be useful to Iranian MT developers, particularly those behind Targoman, Abadis, and Fastdic, as the feedback identifies areas of strength and potential growth. These observations can help guide future updates and localization efforts. The results may also be helpful to Iranian educational institutions, allowing universities and research centers to incorporate them into their translation studies programs. This increases their comprehension of computational linguistics and AI development.

1.1. Review of the related literature

1.1.1. Mechanism of global MTs

According to Kituku et al. (2016), MT is based on three major approaches. These include the knowledge-driven approach used in rule-based MT (RBMT), the data-driven approach used in statistical-based MT (SBMT), and the hybrid approach, which blends the two and is used in hybrid MT. RBMT was the first mechanism for MT systems, and as Chan (2004) describes, it was based on linguistic knowledge of both the source language (SL) and target language (TL), including semantic and contextual comprehension.

According to Okpor (2014), RBMT translates text word-for-word with basic grammatical changes, getting equivalents from multilingual dictionaries. Gough (2005) clarifies that SBMT was a more sophisticated version of RBMT that addressed several of the latter's linguistic constraints. As Menezes (2002) points out, SBMT concentrated on a huge corpus of bilingual texts rather than linguistic rules. Brown (1999) describes SBMT as statistical since it uses a statistical model to increase translation accuracy.

HMT combines the strengths of both knowledge-driven and data-driven techniques to solve each method's deficiencies, as Almahasees (2020) argues. Okpor (2014) emphasizes that the MT research community employs HMT to produce higher-quality translations. Somers (2014) claims that HMT was created because prior MT versions, despite their strengths, were not always useful in all situations.

While MT has evolved, each approach presents obstacles. RBMT requires a considerable amount of linguistic data, whereas SBMT relies on big bilingual corpora and may be inaccurate. HMT enhances translation quality but has difficulties with smooth integration. With improvements in AI, Neural Machine Translation (NMT) has arisen, improving fluency, context awareness, and accuracy in modern MT systems.

1.1.2. Mechanism of domestic MTs

1.1.2.1. Targoman translator

Researchers at Amirkabir University of Technology's Natural Language Processing and Machine Learning Laboratory, directed by Dr. Shahram Khadevi, created Targoman Translator in 2013. The Communications and Information Technology Research Center (Iran Telecommunication Research Center) provided support for the project. Initially, it used open-source statistical MT technologies and concentrated on Persian as part of a research effort.

Between January 2013 and January 2014, the system underwent improvements and reconfiguration. The developers said that the new translation engine was 270% faster and consumed up to 16 times fewer resources while maintaining good accuracy. Targoman includes a variety of language processing techniques, including pre-processing, word root searching, named entity recognition, transliteration, and language modeling, all of which are language-independent.

Targoman was further developed from March 2015 to April 2017 as part of a local initiative funded by the Research Institute of Communication and Information Technology. Targoman Intelligence Processing and Vira Afzar Adan handled the technological parts. Until August 2022, Targoman delivered English-Persian translations via a neural translation engine. It also provided translation services in nine global languages, including English and French, with the assistance of a European partner.

1.1.2.2. Abadis translator

Abadis initiated its activities in information technology in 2005 with the launch of its first project, the Abadis Dictionary website. Since that time, the online dictionary has undergone extensive revisions and has incorporated a wide range of additional functionalities. The organization has also produced a browser add-on for Firefox and Chrome, a mobile application, and a dictionary and translation bot designed for Telegram. The Abadis dictionary is organized into several components, including multilingual dictionaries in English, Farsi, and Arabic, along with twenty domain-specific dictionaries. It provides both English and American pronunciation, translations of abbreviations and acronyms, and an online text translation service. Users are able to access authoritative sources such as the Dehkhoda and Moin dictionaries.

The platform further offers a dictionary of synonyms and antonyms, proposed Persian equivalents for foreign lexical items, expressions endorsed by the Persian Language and Literature Academy, a general encyclopedia, an Islamic encyclopedia, a dictionary of personal names, a tool for identifying related family names, and visual resources associated with lexical entries.

1.1.2.3. Fastdic translator

Fastdic's operations began in 2006, with the establishment of its website. It was later extended by providing apps for iOS (iPhone and iPad) and Android to enhance the user experience. The dictionary has over 250,000 terms and thousands of phrase examples, and Fastdic's expert editorial team adds new words daily. With over two million active users on its website and applications, Fastdic is dedicated to improving the quality of its English and Persian content. The platform also includes useful features and tools to enhance the user experience. Fastdic has received numerous awards over the previous decade, including Best Website at the 12th, 13th, and 14th Iran Web and Mobile Festivals. It was also awarded as the Best Software in the Books and References category based on public votes, and earned the Best iOS Software award at the Iran Web Festival (iOS category).

Applying the methodological framework proposed by Wilss (1982) to the development of local machine translation systems has the potential to enhance the quality of Persian translations by producing outputs that are more precise and contextually appropriate. At a broader international level, systematically improving platforms such as Targoman, Abadis, and Fastdic through this approach could contribute to the advancement of MT technologies for Persian as well as other languages in the region. Such work would support the creation of a consistent standard for addressing linguistic complexity in MT and would strengthen overall progress in the global MT environment.

1.1.3. Wilss' matrix for MT evaluation

Wilss (1982) criticizes early MT approaches for depending too much on a lexical approach that ignored the target language's syntactic principles. This strategy was mainly ineffective, especially considering the limitations of MT technology in the 1950s and 1960s. As a result, MT research evolved into a second generation that focused on assessing surface structures and applying grammatical models. However, these attempts were insufficient, focusing solely on morphosyntactic rule programming, resulting in unsatisfactory translation outputs.

Wilss (1982) contends that true text comprehension requires three forms of information: *real-world knowledge*, *situational knowledge*, and *text-internal knowledge*. This complexity highlights a key gap between human translators and MT systems: machines lack a profound, integrated understanding of reality. He proposes that MT studies expand beyond surface structures to investigate deeper, more abstract structures. This change could assist MT systems in addressing the mental and cognitive parts of translation that human translators naturally use but that robots cannot mimic.

Wilss (1982) discusses fourth-generation MT systems and the possibility of artificial intelligence and conceptual dependence theory to produce more human-like translation procedures. This generation intends to apply a semantico-conceptual interlingua and emphasize the communication function of phrases, in line with the pragmatic features of translation. The author highlights that human translators may use extralinguistic information and macro contextual analysis to make correct translations, which computers cannot match. To improve this, he proposes that MT systems include both syntactic analysis and semantic programming, which could eventually lead to fully automatic high-quality translations (FAHQT).

Utilizing the methodological approach outlined by Wilss (1982) in the development of local machine translation systems may significantly improve the accuracy and contextual appropriateness of Persian translations. On a wider international scale, the systematic enhancement of MT platforms such as Targoman, Abadis, and Fastdic through this framework could advance machine translation capabilities for Persian and other regional languages. This approach would facilitate the establishment of standardized procedures for managing linguistic complexity in MT and promote broader advancements within the global machine translation landscape.

1.1.4. Recent studies in the field

Compared to local academics, foreign researchers have concentrated more on evaluating the quality of MTs. Stahlberg (2020), Tan et al. (2020), Deng and Yu (2022), Rivera-Trigueros (2022), Zappatore and Ruggieri (2024), and Mercan et al. (2024) are only a few of the recent review papers on MT that emphasize this area of research. The majority of empirical research has been on Google Translate and its effectiveness as a widely used, free MT service. For instance, Wang et al. (2024) developed human evaluation measures for figurative language translations and investigated how well MTs translate them. According to their findings, literal and figurative translations are very different.

Hendy et al. (2023) examined Generative Pre-Trained Transformer (GPT) models for translation. They found that these models work well for languages with a lot of resources but struggle with lesser-known languages. They also mentioned that combining GPT models with other systems can boost translation quality. Yan et al. (2023) investigated various automatic metrics used in training machine learning systems, revealing issues with robustness and biases in training datasets.

Zerva and Martins (2024) explored biases in translation quality evaluation and discovered that current techniques frequently underestimate uncertainty. Their findings suggested that conformal prediction could increase accuracy and fairness when judging translation quality across language pairs. Munkova et al. (2020)

investigated whether all automated error rate and accuracy measurements are required when assessing MTs, specifically those from synthetic Slovak to analytical English. Their findings revealed that all analyzed measures, with the exception of the *f*-measure, are required for sentence-by-sentence evaluation of MT quality.

In Iran, research on native MTs is limited. The majority of research has concentrated on Google Translate, with just a few recent studies on computer-assisted translation (CAT) and information and communication technology (ICT) tools (Abdi, 2019, 2022; Taghizadeh & Azizi, 2017). With an emphasis on accuracy and cultural understanding, Aghai's (2024) study examined how well ChatGPT and Google Translate handle literary translations from Persian to English. In their comparison of Google Translate and Bing Translator, Safari et al. (2023) found that Google's translations were more accurate, especially when it came to medical contexts. These studies emphasize how important human oversight is to maintaining cultural integrity and translation quality.

Mirzaeian and Oskoui (2022) probed how Iranian EFL student teachers see MT's role in academic language acquisition and found that many were comfortable using MT tools. Abdi (2021b) assessed Google Translate to determine whether it helps or threatens human translators. The findings indicated that Google Translate worked well when translating English to Persian, particularly in terms of semantic adequacy and understandability, but suffered with fluency. Thus, there is no immediate prospect of machine translation displacing human translators.

Taleghani et al. (2019) investigated the validity of MT evaluation metrics (MTEMs) for Persian texts and their relationship to human judgments. The findings revealed that a substantial correlation between automatic and human evaluations of English-to-Persian translations, with the GTM measure providing the greatest 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 showed that both systems produced low-quality translations, while Google Translate outperformed SDL Free Translation.

1.2. Purpose of study

This study addresses a gap by evaluating the performance of three domestic machine translation systems and examining their strengths, limitations, and overall reliability in English-to-Persian translation. The current study differs from previous ones in numerous crucial respects. First, whereas much of the existing literature has focused on global MT systems such as Google Translate and GPT models, the current study particularly evaluates the quality of three developed Iranian MT systems: Targoman Translator, Abadis Translator, and Fastdic Translator. By focusing on these domestic systems, the study addresses an existing gap in the literature, which frequently overlooks regionally specialized technologies. Second, the study examines MT outputs in three crucial areas: *syntax*, *semantics*, and *pragmatics*, offering a more thorough evaluation of translation quality that has not been investigated in earlier research. Furthermore, the study takes a unique approach by including ChatGPT as a post-editing tool and examining its potential as a dependable AI editor for refining MT outputs. Beyond the conventional focus on raw translation accuracy, this exploration of ChatGPT's potential as a post-editing tool contributes a substantial dimension to the conversation about human-machine collaboration in MT. To sum up, the current work differs from earlier studies in that it offers a localized viewpoint, a comprehensive assessment of MT quality, and a fresh investigation of AI-assisted post-editing.

2. METHODS AND MATERIALS

2.1. Participants

The participants in this study were evaluators invited to assess the output quality of three domestic MT systems: Targoman Translator, Abadis Translator, and Fastdic Translator. They used Wilss's (1982) evaluation matrix, which includes syntax, semantics, and pragmatics. To select the evaluators, the researcher first contacted 20 professional translators. These translators were chosen randomly from those who work with English and Persian. They were selected from two websites: www.iacti.ir/members.html and www.proz.com.

The translators were informed about the study's objectives, the research topic, and their important role in achieving the desired results and improving the functionality of the domestic MT systems.

3.2. Data collection instrument

A translation assessment comprising 80 statements was designed to collect relevant data. Of these, 60 statements were selected based on Mistrík's (1997) classification of text types, encompassing narrative, descriptive, and argumentative texts, with each category containing 20 statements drawn from diverse sources. Narrative statements were sourced from Rowling's (2003) *Harry Potter and the Sorcerer's Stone*, descriptive statements were extracted from Burke and Maxwell's (2012) *Lonely Planet: Iran*, which provides insights into Iranian culture and tourist attractions, and argumentative statements were obtained from academic articles by Abdi (2021a), Abdi (2021b), and Dalaslan (2015). To evaluate the performance of the three MT systems and examine their handling of figurative language, 20 idiomatic statements with implied meanings were included, derived from a compendium of contemporary American English idioms.

The test underwent validation by a panel of experts who reviewed the statements to ensure their appropriateness, organization, and effectiveness in assessing the output quality of the three domestic MT systems. This panel consisted of specialists in relevant disciplines. Based on their recommendations, several statements were either reordered or replaced. Additionally, the test was administered to 15 translators whose backgrounds were comparable to those of the expert evaluators. Comparison of their responses with the overall scores revealed a significant correlation ranging from 0.54 to 0.83, with a mean of 0.76, supporting the content validity of the test. Furthermore, the overall validity was reinforced by the statistical significance of each statement, with p-values below 0.05 ($p < 0.05$).

2.3. Data collection procedure and analysis

For data collection, the translation test was initially rendered from English into Persian using Targoman Translator, Abadis Translator, and Fastdic Translator. The resulting Persian translations, together with the original 80 English statements, were provided to evaluators who assessed the translation quality across three dimensions: syntax, semantics, and pragmatics, employing Wilss's (1982) evaluation matrix. Evaluators utilized a Likert-scale questionnaire comprising five response options: wrong, inappropriate, undecidable, correct, and appropriate. This scale facilitated a systematic evaluation of the domestic MT systems' ability to generate accurate Persian translations with respect to syntactic, semantic, and pragmatic correctness. In addition, ChatGPT was employed to revise the statements classified as wrong or inappropriate by the evaluators, in order to determine whether the quality of these translations could be enhanced. Data analysis consisted of calculating the percentage and mean score for each category, which were subsequently presented in tabular form. The Wilcoxon signed-rank test was applied to determine whether the evaluators' assessments of the 80 Persian translations produced by Targoman Translator, Abadis Translator, and Fastdic Translator were significantly consistent within each of the three evaluation areas.

3. RESULTS

3.1. Syntactic evaluation

Table 1 shows that the evaluators agreed on the syntactic correctness of a large majority (90%) of the 60 statements produced by Targoman Translator. They also agreed that a significant majority (85%) of the 60 Persian translations generated by Abadis Translator were syntactically correct. Additionally, the evaluators found that 87% of the 60 Persian translations produced by Fastdic Translator were syntactically correct as well.

Table 1

Percentages of the Evaluators to Syntactic Evaluation of the 60 Statements

Type of MTs	Wrong	Inappropriate	Undecidable	Correct	Appropriate	*N*	
	%	*%*	*%*	*%*	*%*	*%*	*M*
Targoman	1.0	4.0	5.0	57.0	33.0	100.0	4.15
Abadis	6.0	9.0	1.0	32.0	53.0	100.0	4.17
Fastdic	2.0	6.0	5.0	59.0	28.0	100.0	4.03

A one-sample Wilcoxon signed-rank test was conducted to examine whether a significant relationship existed between the evaluators' assessments and the syntactic accuracy of the Persian translations produced by Targoman Translator, Abadis Translator, and Fastdic Translator. The p-value for the overall evaluation of the 60 Persian translations for each MT system was 0.0, which is below the 0.05 threshold ($p < 0.05$), as presented in Table 2. The mean scores for all statements were 4.15, 4.17, and 4.03, each exceeding the midpoint of 2.5. These results indicate a general consensus among the evaluators regarding the syntactic correctness of the Persian translations generated by Targoman Translator, Abadis Translator, and Fastdic Translator.

Table 2

One-Sample Wilcoxon *signed-rank test* for syntactic evaluation of the 60 statements

Type of MTs	*N*	*MDN*	*P*
Targoman	20	4	.000
Abadis	20	5	.000
Fastdic	20	4	.002

3.2. Semantic evaluation

Table 3 shows that the evaluators rated over half of the 60 translations (59%) as correct in terms of semantic adequacy. They also agreed that Abadi's Translator produced only half of the 60 statements (50%) correctly. The evaluators noted that the number of semantically correct translations produced by Fastdic Translator (49%) was almost equal to the number of incorrect translations (48%).

Table 3

Percentages of the evaluators for the semantic evaluation of the 60 statements

Type of MTs	Wrong	Inappropriate	Undecidable	Correct	Appropriate	*N*	
	%	*%*	*%*	*%*	*%*	*%*	*M*
Targoman	2.0	17.0	8.0	56.0	18.0	100.0	3.69
Abadis	16.0	30.0	4.0	32.0	18.0	100.0	3.02
Fastdic	18.0	30.0	3.0	35.0	14.0	100.0	2.93

A one-sample Wilcoxon test was conducted to see how much the evaluators agreed on the semantic adequacy of the Persian translations. It also aimed to find out if there was a significant relationship between their opinions and the semantic adequacy of translations from Targoman Translator, Abadis Translator, and Fastdic Translator. The results showed that the *p*-values for each MT type were 0.000, 0.003, and 0.001, which are all less than 0.05 ($p < .05$) (see Table 4). Additionally, the average scores for each translation type were above the theoretical mean/median of 2.5. This means that the evaluators significantly agreed on the semantic adequacy of more than half of the Persian translations produced by the three domestic MTs, indicating a higher-than-average level of semantic adequacy.

Table 4

One-Sample Wilcoxon *signed-rank test for the semantic evaluation of the 60 statements*

Type of MTs	*N*	*MDN*	*P*
Targoman	20	4	.000
Abadis	20	4	.003
Fastdic	20	4	.001

3.3. Pragmatic evaluation

As shown in Table 5, the evaluators found that none of the 20 idiomatic statements (100 percent) generated by Targoman Translator met the criteria for pragmatic correctness. Similarly, all 20 idiomatic statements (100 percent) produced by Abadis Translator were judged as pragmatically incorrect. The same pattern was observed for Fastdic Translator, where the evaluators did not consider any of the 20 idiomatic statements (100 percent) to be pragmatically accurate.

Table 5

Percentages of the evaluators for pragmatic evaluation of the 20 idiomatic statements

Type of MTs	Wrong	Inappropriate	Undecidable	Correct	Appropriate	N	
	%	%	%	%	%	%	M
Targoman	80.0	20.0	-	-	-	100.0	1.2
Abadis	40.0	60.0	-	-	-	100.0	1.59
Fastdic	37.0	63.0	-	-	-	100.0	1.63

A one-sample Wilcoxon signed-rank test was conducted to see if there was a significant relationship between the evaluators' assessments and the pragmatic adequacy of the 20 Persian translations from Targoman Translator, Abadis Translator, and Fastdic Translator. Table 6 shows that the *p*-values for the idiomatic statements from the three machine translators were below 0.05 ($p < .05$). Also, the average scores for these statements were below the theoretical mean/median. This indicates that the evaluators did not consider the idiomatic translations to be pragmatically adequate.

Table 6

One-Sample Wilcoxon *signed-rank test for pragmatic evaluation of the 20 idiomatic statements*

Type of MTs	*N*	*MDN*	*P*
Targoman	20	2	.002
Abadis	20	1	.001
Fastdic	20	2	.01

4. DISCUSSION

The results showed that Targoman Translator, Abadis Translator, and Fastdic Translator generally performed well in translating Narrative, documentaries, and argumentative statements into Persian. They produced grammatically correct translations that mostly followed the SOV structure of Persian. However, there were some awkward translations with incorrect tenses and subject-verb agreements, especially in complex sentences. This issue was more common in Abadis Translator and Fastdic Translator than in Targoman Translator.

For instance, there was improper subject–verb agreement in the Persian translation of the original statement, as these items are related to Erzurum culture; their renderings are challenging because they may not be present in the target culture. Since such notions may not exist in the target culture, translating this sentence, which refers specifically to Erzurum culture, is difficult. Abadis Translator and Fastdic Translator rendered the source-language (SL) item as *az ānjāyi ke in mavāred marbūt be farhang-e Erzurum ast, renderhā-ye ānhā chālesh-barangiz ast zīrā momken ast dar farhang-e hadaf vojud nadāshte bāshad* (since these items are related to Erzurum culture, their renderings are challenging because they may not be present in the target culture). This example uses the singular verb *ast* ("is") with the plural subject in *mavāred* ("these items"). To ensure correct subject–verb agreement, the verb should be plural, *hastand* ("are"), because *mavāred* ("items") is plural. The translations of Abadis Translator and Fastdic Translator also used the singular verb ast ("is") with the plural subject *renderhā-ye ānhā* ("their renderings"), which is incorrect. Although Targoman Translator made the same mistake in the second position, it maintained correct agreement in the first.

According to the SL statement, Bazar-e Tabriz is the world's largest covered bazaar and was once one of the most important trading centers on the Silk Road. The SL word *zamānī* ("once") indicates a past event and therefore requires a past verb. To convey this meaning appropriately, the correct tense is *bude ast* ("has been"), which denotes a past state (its importance on the Silk Road) while linking it to the present.

The results also showed that the three machine translators were moderately successful in producing semantically correct Persian translations. Their semantic quality was rated slightly above average. For example, the SL phrase keep your noses out if you know what's good was translated incorrectly by all three systems as *agar mīdānid che chizī khob ast, bīnītān rā birun negah* dārid ("if you know what is good, keep your nose outside"). The problem is that the systems translate every word literally, ignoring the idiomatic meaning in English. In Persian, the literal expression keep your nose outside sounds odd, fails to convey the intended meaning, and does not make sense to the audience. Similarly, *che chizī khob* ast ("what is good") is overly literal and does not communicate the implied warning ("you should know what's good for you"). To convey the correct sense, an idiomatic equivalent is necessary, one that expresses both staying out of others' business and the cautionary tone. An appropriate rendering is *agar mīdānid che be salāhatetān ast, dekhālat nakonid* ("if you know what is in your best interest, do not interfere"), which is the version provided through ChatGPT's post-editing.

Another example concerns the sentence Behesht-e Zahra is Tehran's biggest cemetery and is interesting primarily because …, which was also translated inaccurately. Abadis Translator and Fastdic Translator rendered it as *Behesht-e Zahra bozorgtarīn qabrastān-e Tehrān ast va dar daraje-ye avval jāleb ast* ("Behesht-e Zahra is Tehran's largest cemetery and is primarily interesting"), which is inappropriate. The phrase jāleb ast ("is interesting") often refers to something entertaining or fascinating, making it unsuitable, possibly even offensive, when describing a cemetery. Furthermore, the phrase *dar daraje-ye avval* ("primarily") does not semantically capture the nuance of the SL expression, which implies that the cemetery is notable for a particular reason. ChatGPT's post-edited version provides a more culturally sensitive and semantically accurate translation: *Behesht-e Zahra bozorgtarīn qabrastān-e Tehrān ast va bishtar be dalil-e ahamīyat-e tārīkhī-ash mored-e tavajjoh ast* ("Behesht-e Zahra is Tehran's largest cemetery and is mainly notable for its historical significance"). This translation replaces interesting primarily with a more appropriate, culturally sensitive structure that conveys both meaning and tone.

From a pragmatic perspective, the three MT systems performed poorly. All idiomatic expressions were translated inaccurately and with low quality. For example, in the SL sentence, Dan became a basket case after the horrible disaster, the idiom basket case was mistranslated as *zanbīl/jabʻe shodan* ("to turn into a basket/box"): *Dan baʻd az in hādese-ye gham'angiz tabdil be yek zanbīl/jabʻe shod*. The English idiom basket case refers to someone who becomes emotionally or psychologically unstable due to trauma or stress; it does not literally mean "basket" or "box." Thus, the translations failed to convey the intended meaning. A correct rendering would be *Dan baʻd az in hādese-ye gham'angiz az nazar-e rūhī be ham rīkht* ("Dan experienced an emotional breakdown after this tragic incident"). This post-edited translation correctly renders basket case as az nazar-e rūhī be ham rīkht ("broke down emotionally").

Overall, the three MTs seem to perform like a basic MT, similar to the second generation of MTs described by Wilss (1982), which focused on surface structure analysis of texts. They rely on predefined grammatical rules or statistical probabilities based on word co-occurrence, even though Targoman Translator's developers claim their system now uses a common neural translation engine.

This rule-based system leads to grammatical issues, like problems with subject-verb agreement and tense, especially in complex sentences. Targoman Translator performed better in handling these grammatical issues. However, the weaknesses in grammar mean these domestic MTs struggle to break down and analyze sentence structures, including the relationships between subjects, verbs, and objects. This requires strong morphological analysis to ensure that a single subject matches with a single verb, and vice versa.

The reliance on literal translation techniques causes semantic deficiencies in Targoman Translator, Abadis Translator, and Fastdic Translator because they use inappropriate equivalents without considering the proper context. Wilss (1982) considers this a major shortcoming of such MTs. In other words, these three MTs produced translations that didn't fully convey the exact meanings of the SL items. They translated words based on their literal meanings rather than their connotative or implied meanings, leading to awkward translations, although Targoman Translator performed somewhat better in this area.

In pragmatic evaluation, the shortcomings of the three MTs come from misunderstanding idioms, which again relates to their inability to recognize context. This suggests that the three MTs may not have advanced enough algorithms to detect contextual cues and manage pragmatic nuances, such as understanding when a phrase is used metaphorically or colloquially. As a result, they tend to translate idiomatic expressions literally.

5. CONCLUSION

The present research aimed to critically evaluate the output quality of three MTs - Targoman Translator, Abadis Translator, and Fastdic Translator - developed by Iranian software developers. Three paradigms are included in this matrix: *pragmatics*, *semantics*, and *syntax*. The objective was to determine each MT type's advantages and disadvantages to lower errors and improve their quality. In order to ascertain which of the three MT types was more effective at translating statements from English into Persian and whether end-users could rely on them, a comparison was also made between them. ChatGPT was also used to assess whether the output of these three domestic MTs required post-editing and to explore if ChatGPT could serve as a reliable substitute for human editors for end-users.

The results showed that Targoman Translator, Abadis Translator, and Fastdic Translator generally performed acceptably in producing grammatically correct Persian translations, although they did show deficiencies in subject-verb agreement and tense in some cases. They were able to produce semantically accurate Persian translations, with Targoman Translator showing better performance in terms of syntax and semantics. However, none of the domestic MTs demonstrated acceptable performance in pragmatics, producing translations of poor quality.

These limitations underscore the need for post-editing of the output from Targoman Translator, Abadis Translator, and Fastdic Translator in translation tasks, particularly for complex sentences or idiomatic expressions. Therefore, these MTs cannot be considered reliable for end users and translators. Furthermore, the results suggested that ChatGPT greatly enhanced translations by eliminating errors and assuring adherence to grammatical standards while effectively delivering the intended message. Although there were instances of literal translation and inappropriate translations, which were rectified upon retranslation. Nonetheless, ChatGPT remains a powerful and reliable AI editor for end-users.

In conclusion, the MTs developed by Iranian MT developers have not yet achieved the level of producing high-quality translations that do not need post-editing. The findings of the study provide valuable recommendations for Iranian developers to enhance the output quality of the three MTs and make them competitive with global MTs such as Google Translate and Microsoft Bing Translator. Iranian developers should focus on improving syntactic parsing and morphological analysis to recognize correct subject-verb relationships.

Moreover, they should enhance contextual analysis to avoid semantic and pragmatic errors, leading to the development of better mechanisms for understanding sentence context. Implementing NMT techniques that consider entire sentence structures rather than translating phrase by phrase would significantly reduce errors. Recognizing functional dependencies and deep structures rather than merely formal or surface structures is crucial for effectively reproducing complex statements and idiomatic expressions in the TL.

The study only assessed three specific MT systems developed by Iranian software developers (Targoman Translator, Abadis Translator, and Fastdic Translator). This limited focus may not give a complete picture of how well MT systems perform overall, especially when developed outside of Iran or by international developers. Additionally, the study only evaluated translation from English to Persian, which limits the findings' generalizability to other language pairs. Furthermore, the study only examined MT systems that were in use at the time. However, because MT technology is evolving so quickly, the systems evaluated might not be relevant for future applications by the time the study's conclusions are released.

Future studies should evaluate a greater range of MT systems, including popular global systems, namely Google Translate, DeepL, etc., and contrast them with locally developed systems. This would give MT a more global perspective. Expanding the study to include more language pairs (Persian to English, English to other languages, etc.) would provide for a more comprehensive and diversified knowledge of the benefits and drawbacks of MT systems across different languages.

A comparison of AI-assisted post-editing (like ChatGPT) and human post-editing might be done to assess the speed, cost-effectiveness, and quality of edits. Examining the operation of AI tools such as ChatGPT with different levels of post-editing complexity (from simple changes to more extensive adjustments) may be helpful.

Conflict of Interest: The authors have no competing interests to declare that are relevant to the content of this article.

Ethical Approval: The study adheres to the ethical guidelines for conducting research.

Funding: This research received no external funding.

REFERENCES

Abdi, H. (2021a). Examining the Appropriateness of Reiss's Functionalist-oriented Approach to Trancism. *Theory and Practice in Language Studies*, *11*(5), 561-567. https://www.researchgate.net/profile/Hamidreza-Abdi-3/publication/351329734_Examining_the_Appropriateness_of_Reiss's_Functionalist-oriented_Approach_to_Trancism/links/6240de205e2f8c7a0344187f/Examining-the-Appropriateness-of-Reisss-Functionalist-oriented-Approach-to-Trancism.pdf

Abdi, H. (2021b). Considering machine translation (MT) as 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.v2i1.122

Abdi, H. (2022). Inquiry into students' familiarity with computer-assisted translation tools. *International Journal of New Trends in Social Sciences*, *6*(2), 53-63.

Aghai, M. (2024). ChatGPT vs. Google Translate: Comparative analysis of translation quality. *Iranian Journal of Translation Studies*, *22*(85). https://www.journal.translationstudies.ir/ts/article/view/1156

Almahasees, Z. (2020). Diachronic evaluation of Google Translate, Microsoft Translator, and Sakhr in English-Arabic translation. https://research-repository.uwa.edu.au/en/publications/diachronic-evaluation-of-google-translate-microsoft-translator-an

Araghi, S., & Palangkaraya, A. (2024). The link between translation difficulty and the quality of machine translation: a literature review and empirical investigation. *Language Resources and Evaluation*, *58*(4), 1093-1114. https://link.springer.com/article/10.1007/s10579-024-09735-x

Brown, R. D. (1999). Adding linguistic knowledge to a lexical example-based translation system. In *Proceedings of the 8th Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*. https://aclanthology.org/1999.tmi-1.3.pdf

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

Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C, & Schroeder, J. (2007). *(Meta-) Evaluation of machine translation* [Conference session]. The Second Workshop on Statistical Machine Translation, Prague, Czech Republic. https://dl.acm.org/doi/10.5555/1626355.1626373

Chan, S. W. (2004). *A dictionary of translation technology*. Chinese University Press. https://books.google.com/books?hl=en&lr=&id=3gwOFvbxMGcC&oi=fnd&pg=PR7&dq=Chan,+S.+W.+(2004).+A+dictionary+of+translation+technology.+Chinese+University+Press.&ots=s79xTrrEKX&sig=QWyaVsGD74361XCGe2BoNM7Kk24

Coughlin, D. (2003). Correlating automated and human assessments of machine translation quality. *ACM Transactions on Software Engineering and Methodology, 31*(4), 1-28. https://doi.org/10.1145/3502853

Dalaslan, D. (2015). An Analysis of the English translations of Erzurum folk riddles in the light of Raymond van den Broeck’s translation criticism model.

Deng, X., & Yu, Z. (2022). A systematic review of machine-translation-assisted language learning for sustainable education. *Sustainability*, *14*(13), 7598. https://www.mdpi.com/2071-1050/14/13/7598

Gough, N. (2005). *Example-based machine translation using the marker hypothesis* (Doctoral dissertation, Dublin City University). https://doras.dcu.ie/17366/

Granell, X. (2014). *Multilingual information management: Information, technology, and translators*. Chandos Publishing.

Han, L. (2016). Machine translation evaluation resources and methods: A survey. *arXiv preprint arXiv:1605.04515*. https://arxiv.org/abs/1605.04515

He, B., Wang, J., & Wang, Y. (2025). Research on quantitative assessment of translation quality from the perspective of phraseology. *PloS one*, *20*(2), e0318804. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0318804

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., ... & Awadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation. *arXiv preprint arXiv:2302.09210*. https://arxiv.org/abs/2302.09210

Kituku, B., Muchemi, L., & Nganga, W. (2016). A review of machine translation approaches. *Indonesian Journal of Electrical Engineering and Computer Science*, *1*(1), 182-190. https://www.academia.edu/download/94385788/4180.pdf

Lee, S., Lee, J., Moon, H., Park, C., Seo, J., Eo, S., ... & Lim, H. (2023). A survey on evaluation metrics for machine translation. *Mathematics*, *11*(4), 1006. https://www.mdpi.com/2227-7390/11/4/1006

Lin, G. H. C., & Chien, P. S. C. (2009). Machine Translation for Academic Purposes. *Online Submission*. https://eric.ed.gov/?id=ED513879

Macías, L. P., Ramos, M. D. M. S., & Rico, C. (2020). Study on the usefulness of machine translation in the migratory context: Analysis of translators' perceptions. *Open Linguistics*, *6*(1), 68-76. https://www.degruyterbrill.com/document/doi/10.1515/opli-2020-0004/html

Melby, A. (2002). The translator workstation. In J. Newton (Ed.), *Computers in translation: A practical appraisal* (pp. 147–165). Routledge. https://doi.org/10.4324/9780203128978

Menezes, A. (2002). Better contextual translation using machine learning. In *Conference of the Association for Machine Translation in the Americas* (pp. 124-134). Berlin, Heidelberg: Springer Berlin Heidelberg. https://link.springer.com/chapter/10.1007/3-540-45820-4_13

Mercan, H., Akgün, Y., & Odacıoğlu, M. C. (2024). The evolution of machine translation: A review study. *International Journal of Language and Translation Studies*, *4*(1), 104-116. https://dergipark.org.tr/en/pub/lotus/issue/85399/1453321

Mirzaeian, V. R., & Oskoui, K. (2022). Investigating Iranian EFL student teachers' attitude toward the implementation of machine translation as an ICALL tool. *Journal of English Language Teaching and Learning*, *14*(30), 165-179. https://elt.tabrizu.ac.ir/article_14859_fc6ae91119fa2d187f07c069d36612a9.pdf

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

Munkova, D., Hajek, P., Munk, M., & Skalka, J. (2020). Evaluation of machine translation quality through the metrics of error rate and accuracy. *Procedia Computer Science*, *171*, 1327-1336. https://www.sciencedirect.com/science/article/pii/S1877050920311212

Okpor, M. D. (2014). Machine translation approaches: issues and challenges. *International Journal of Computer Science Issues (IJCSI)*, *11*(5), 159. https://www.academia.edu/download/41487988/IJCSI-_okpor.pdf

Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Language Resources and Evaluation*, *56*(2), 593-619. https://link.springer.com/article/10.1007/s10579-021-09537-5

Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Language Resources and Evaluation*, *56*(2), 593-619. https://link.springer.com/article/10.1007/s10579-021-09537-5

Rowling, J. K. (2023). *Harry Potter and the Sorcerer's Stone* (p. 864). Scholastic Incorporated. https://weobleyprimary.com/wp-content/uploads/2021/02/Jane-Ameghino_Harry-Potter-and-The-Sorcerers-Stone.pdf

Saffari, M., Pourhaji, M., Fathi, A. S., Sajjadi, S., & Mohammadi, M. (2023). Translating medical texts from Persian to English: Accuracy of machine translation. https://www.sid.ir/paper/1139965/en

Somers, H. (2014). Machine translation: Latest developments. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* (pp.512-528). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199573691.001.0001

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 competencies of some Iranian translators. International Journal of English Language & Translation Studies, 5(1), 78-87.* *https://www.researchgate.net/profile/Mahboubeh-Taghizadeh/publication/320212519_Exploring_Computer-Aided_Translation_Competences_of_Some_Iranian_Translators/links/59d504814585150177fc8f33/Exploring-Computer-Aided-Translation-Competences-of-Some-Iranian-Translators.pdf*

Taleghani, M., Pazouki, E., & Ghahraman, V. (2019). The correlation of machine translation evaluation metrics with human judgement on the Persian language. https://www.sid.ir/paper/320538/en

Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, M., & Liu, Y. (2020). Neural machine translation: A review of methods, resources, and tools. *AI Open*, *1*, 5-21. https://www.sciencedirect.com/science/article/pii/S2666651020300024

Vaezian, H., & Pakdaman, A. (2018). A Comparative Study of the Quality of Persian Translations by Google Translate and SDL Free Translation. *Iranian Journal of Translation Studies*, *15*(60), 61-78. https://journal.translationstudies.ir/index.php/ts/article/view/500

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 preprint arXiv:2406.13698*. https://arxiv.org/abs/2406.13698

Wang, W. (2023). A Review of Machine Translation Quality Assessment Methods. *Frontiers in Computing and Intelligent Systems*, *5*(2), 108-110. https://doi.org/10.54097/fcis.v5i2.13113

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

Yan, Y., Wang, T., Zhao, C., Huang, S., Chen, J., & Wang, M. (2023). BLEURT has universal translations: An analysis of automatic metrics by minimum risk training. *arXiv preprint arXiv:2307.03131*. https://arxiv.org/abs/2307.03131

Zappatore, M., & Ruggieri, G. (2024). Adopting machine translation in the healthcare sector: A methodological multi-criteria review. *Computer Speech & Language*, *84*, 101582. https://www.sciencedirect.com/science/article/pii/S0885230823001018

Zerva, C., & Martins, A. F. (2024). Conformalizing machine translation evaluation. *Transactions of the Association for Computational Linguistics*, *12*, 1460-1478. https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00711/125277