Global Journal of Information Technology: Emerging Technologies

Volume 15, Issue 1, (2025) 23-34

www.gjit.eu

A critical evaluation of translation quality and post-editing performance of fastdic translator

Hamidreza Abdi[a1], Universitat Pompeu Fabra, Spain, hooman_78h@yahoo.com https://researchid.co/hamid57

Suggested Citation:

Abdi, H. (2025). A critical evaluation of translation quality and post-editing performance of fastdic translator. *Global Journal of Information Technology: Emerging Technologies, 15*(1), 23-34. https://doi.org/10.18844/gjit.v15i1.9679

Received from; October 12, 2024, revised from; January 25, 2025 and accepted from March 22.

Selection and peer review under the responsibility of Assoc. Prof. Dr. Ezgi Pelin YILDIZ, Kafkas University, Turkey

iThenticate Similarity Rate: 3%

Abstract

This study aimed to critically assess the translation quality of Fastdic Translator, a domestic machine translation, using Wilss's (1982) matrix, which evaluates syntax, semantics, and pragmatics, to identify the system's strengths and weaknesses. Additionally, ChatGPT was utilized to determine if the translations generated by Fastdic Translator required post-editing and to assess whether ChatGPT could serve as a reliable AI tool for translators and end-users, potentially replacing human editors. The research involved a translation test of 80 sentences, chosen from Mistrík's (1997) text type classifications, including narrative, descriptive, and argumentative styles, with idiomatic expressions added by the researcher. Fastdic Translator first translated these sentences from English into Persian. Both the original English text and the Persian translations were provided to evaluators, who evaluated the translation quality using a Likert-scale questionnaire. The results indicated that while Fastdic Translator produced grammatically sound Persian translations, there were occasional issues with subject-verb agreement, especially in more complex sentences. Semantically, the translations were generally accurate. However, in the pragmatic evaluation, the system struggled, particularly with idiomatic expressions and implied meanings, producing inappropriate translations. The findings also revealed that post-editing with ChatGPT significantly improved the translations by correcting errors and better conveying the intended meaning while ensuring grammatical accuracy. Therefore, ChatGPT can be seen as a powerful and reliable AI editor for both end-users and translators.

Keywords: Machine Translation, Fastdic Translator, Wilss's Matrix, Translation Evaluation

***** ADDRESS FOR CORRESPONDENCE: **Hamidreza Abdi**, Universitat Pompeu Fabra, Spain.
E-mail address: hooman_78h@yahoo.com

1. INTRODUCTION

Machine translation (MT) is a type of natural language processing software that enables automatic translation between different languages on a computer. Essentially, it provides translators with the tools to use translation software effectively (Abdi, 2020). MT is not merely about replacing words by word; it operates based on linguistic rules derived from the Rule-Based Machine Translation approach. According to Zakaryia (2020), it also incorporates statistical models from the Statistical-Based Machine Translation approach, allowing for translations generated from established electronic corpora. This capability enables MT to offer users translations of a variety of texts.

Historically, access to MT was prohibitively expensive for many translators. Consequently, online MT services like Google Translate emerged, providing free access to translators. These online tools were embraced by the translation community for their ability to save time and lower translation costs, as noted by Abdi (2021a). Recently, Iranian developers have introduced domestic MT systems to fill this gap and attract Persian language users to local products. This development has resulted in the launch of several local MT options, including Targoman Translator, Abadis Translator, and Fastdic Translator, which are available for both free and paid use, depending on the number of words being translated. A critical question arises whether these local MTs produce high-quality translations to compete with famous MTs like Google Translator and Microsoft Bing Translator.

When considering translation quality, the main concern is how this crucial characteristic can be evaluated. House (1997) suggests that translation quality assessment (TQA) has been a significant concern throughout the history of translation studies. This may stem from the absence of universally accepted criteria for evaluating translations. The same issue persists in the realm of MT quality assessment, where no unique or general criteria exist for either automated or human evaluation. One straightforward approach to addressing this challenge is to assess a range of translations produced by various types of MT systems across different fields, such as economics and literature, using existing models and frameworks. This process can help establish general criteria for evaluating MT quality.

Wilss (1982) introduced a matrix for evaluating translations based on three critical dimensions: syntax, semantics, and pragmatics. While he regards this matrix as preliminary, it provides valuable insights for developing methodologies for human translation assessment. The matrix enables evaluators to analyze translations from various perspectives: syntactic analysis examines grammatical structure, semantic analysis investigates word meanings, and pragmatic analysis considers implied meanings within context. Therefore, Wilss's matrix serves as a useful tool for formulating practical methodologies in translation research, combining both descriptive and evaluative components.

In this study, Wilss's (1982) matrix is employed to assess Fastdic Translator by analyzing its syntactic, semantic, and pragmatic attributes to highlight its strengths and weaknesses. The aim is to assist developers in enhancing the tool by addressing its shortcomings and leveraging its strengths. Additionally, ChatGPT was used to determine whether the output from Fastdic Translator requires post-editing and to evaluate if ChatGPT could effectively serve as a substitute for human editors for both users and translators.

This study is part of a research project aimed at assessing the performance of Iranian software developers in creating local MT tools, such as Targoman Translator, Abadis Translator, and Fastdic Translator, with an emphasis on their quality and the necessity for post-editing. To achieve the objectives of the present study the following questions are raised:

1. What was the output quality of Fastdic 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 important because it focuses on developing localized tools that meet specific linguistic and cultural needs for Persian language MT systems. Persian-speaking users can benefit from MT tools that consider the unique aspects of their language instead of relying on global systems that may miss these nuances. The current study also explores whether ChatGPT is an effective tool for post-editing. This is crucial as it addresses the increasing use of AI in professional translation and assesses if these systems can reduce the need for human intervention.

If ChatGPT proves to be a reliable editor, it could change translation workflows by saving time and money in post-editing machine-generated translations. The results of this study could help Iranian MT developers improve the quality of their translations, especially in syntax, semantics, and pragmatics. The findings would also benefit Persian-speaking users who need translations for personal, academic, or business purposes by offering better localized MT tools. If local MTs perform well, users can trust these tools for reliable translations in everyday situations.

Furthermore, translators who use MT tools will gain a better understanding of how dependable local systems like Fastdic Translator are and how much post-editing they require. If the study shows that MT systems can produce high-quality translations with minimal edits, translators may find these tools useful for saving time and increasing efficiency in their work.

2. Review of the Related Literature

2.1. Methods of MT Evaluation

MT evaluation aims to assess the quality of MT output to identify its strengths and weaknesses for end users. Generally, there are two main types of MT evaluation: automated and human evaluations. Automated evaluation assesses the quality of MT output using metrics and references created by human translators, without human involvement in the assessment process. Common metrics like NIST and BLEU are used for this purpose, and Banerjee and Lavie (2005) suggest that the choice of metrics should depend on the specific goals of the evaluation.

Automated evaluation offers several benefits. Philipp (2007) outlines these as being fast and inexpensive, requiring no bilingual speakers, involving minimal human labor, and allowing continuous testing during system development. However, despite these advantages, automated evaluation has significant drawbacks that can impact the accuracy of quality assessment. Its primary limitation is that it focuses on word matching rather than meaning. Additionally, automated metrics are limited in evaluating the semantic and pragmatic aspects of translation, making human evaluation essential for obtaining reliable results, as Escribe (2019) highlights.

Human evaluation, on the other hand, involves professional translators assessing MT output. Pospelova and Rowda (2016) regard human evaluation as an essential method for MT quality assessment, even though there is no consensus on the best approach. Callison-Burch et al. (2007) argue that human evaluation is a more authoritative alternative to automated evaluation, noting the significant differences in results between the two methods and the higher inter-annotator agreement in human evaluations. For these reasons, the present study used human evaluation to assess the output quality of Fastdic Translator.

2.2. Fastdic Translator

Fastdic began its operations in 2006 with the launch of a website, later expanding by offering iOS (iPhone, iPad) and Android apps to create a more user-friendly experience. The dictionary includes over 250,000 words and thousands of sentence examples, with new words added daily by Fastdic's experienced editorial team. With more than two million active users across its website and apps, Fastdic continues to enhance the quality of its English and Persian content, as well as adding practical features and tools to provide a better user experience.

Available to end-users for over a decade, Fastdic has earned numerous awards, including the Best Website award at the 12th, 13th, and 14th Iran Web and Mobile Festival, Best Software in the Books and References category based on public votes, and Best iOS Software at the Iran Web Festival (iOS category) (https://fastdic.com).

2.3. Wilss' Matrix for MT Evaluation

Wilss (1982) developed a framework called trancism, a term coined by Abdi (2021a) and used as an alternative to Translation Criticism. This framework covers three key areas: syntax, semantics, and pragmatics, providing a structured way to evaluate translated texts. Wilss believed this approach was useful for MTs because it reflects the process of translating from the SL to the TL. This involves decoding the SL and then encoding it into the TL through technological and algorithmic methods. However, Wilss pointed out that this approach often falls short, as algorithmic methods tend to be incomplete and more of a theoretical promise than a fully realized solution.

Wilss (1982) discussed the early development of MT, which initially focused on a lexical approach supported by syntax rules. While this method initially seemed promising and raised high expectations, issues arose due to structural differences between languages. Unlike human translators, who consider both internal and external textual factors, computers are limited to processing only explicit, text-internal instructions. This limitation posed challenging, especially when computers encountered ambiguous or complex linguistic structures in source language texts (SLTs).

To overcome these issues, a combination of syntactic and semantic approaches was introduced, enabling more detailed structural analysis of language. This method provided greater flexibility across different language pairs, but still failed to fully address pragmatic challenges, such as interpreting participle constructions. Wilss (1982) argues that computers, unlike human translators, lack the cognitive ability to infer missing semantic elements necessary to accurately interpret such constructions. To improve MT, Wilss recommended an integrated approach that combines both syntactic and semantic analysis, allowing machines to better handle linguistic complexities and improve translation quality.

2.4. Recent Studies in the Field

Recent advancements in MT have attracted a lot of attention from researchers, leading to more studies and reviews. This research helps us better understand how MT systems work and explores new challenges and opportunities. For example, Zappatore and Ruggieri (2024) reviewed current MT research using a method commonly used in clinical trials, called PRISMA. Rivera-Trigueros (2022) looked at which MT systems are most popular, how they are built, and how their quality is evaluated. The study found that neural machine translation (NMT) is now the most common approach, with Google Translate being the most used system. Stahlberg (2020) explained that modern NMT systems evolved from earlier models like word and sentence embeddings. Although there are not many empirical studies on MT evaluation yet, some important work has been done. Hendy et al. (2023) evaluated Generative Pre-Trained Transformer (GPT) models for translation and found that these models perform well for languages with a lot of resources but struggle with less common languages. They also found that combining GPT models with other systems can improve translation quality. Ulitkin et al. (2021) used automatic metrics like BLEU and TER to compare Google and PROMT NMT systems, helping identify errors and improve future translations. Prathyusha (2022) created a dataset to test pronoun translation and ranked different MT systems based on how well they handled these errors. The study found that MT models don't perform consistently across languages. In Iran, research on local MT systems has been limited and most studies focusing on Google Translate. However, a few recent studies have explored CAT tools and information and communication technology (ICT) tools (e.g., Abdi, 2019, 2021b, 2022; Taghizadeh & Azizi, 2017). Aghai (2024) studied how well ChatGPT and Googl Translate handle literary translations from Persian into English, pointing out problems with accuracy and cultural understanding. Sartipi et al. (2023) found differences in how well different datasets performed in Persian-

English translations. Abdi (2021a) evaluated the ability of Google Translate to translate English into Persian. The study showed that while Google Translate did well in terms of meaning and understanding, it struggled with fluency. Similarly, Bonyadi (2020) found problems in Persian-to-English translations, including issues with tense, literal translations, and poor word choices. This study is different because it focuses on the local MT system, Fastdic Translator. It evaluates Fastdic's translation quality based on syntax, semantics, and pragmatics. The study also looks at whether ChatGPT can help improve Fastdic's translations and whether it can be a reliable tool for users and translators.

3. Method

Participants

The participants in this study were evaluators invited to evaluate the output quality of the domestic MT system, Fastdic Translator. To select the evaluators, the researcher initially reached out to 20 professional translators who were randomly chosen for their experience translating between English and Persian. These translators were selected from two websites: www.iacti.ir/members.html and www.proz.com. They were well-informed about the study's objectives, the research focus, and their important role in achieving the desired outcomes and helping improve the performance of the domestic MT system.

Instrumentation

A translation test comprising 80 statements was developed for data collection. Of these, 60 statements were selected following Mistrík's (1997) classification of text types: narrative, descriptive, and argumentative, with each category containing 20 statements from different sources. Narrative statements were taken from Rowling's (1997) Harry Potter and the Sorcerer's Stone, descriptive statements from Burke and Maxwell's (2012) Lonely Planet: Iran, which covers aspects of Iranian culture and sightseeing, and argumentative statements from scholarly articles, including Abdi (2021a, 2021b) and Dalaslan (2015). To challenge Fastdic Translator and evaluate its pragmatic accuracy, 20 idiomatic expressions with implied meanings were added, extracted from the Book of Idioms, which includes various American English idioms.

A panel of experts reviewed the test to ensure the statements were well-chosen and effectively assessed Fastdic Translator's output quality. These experts provided feedback, leading to revisions, including reordering and replacing some statements. To further validate the test, it was administered to 15 translators with similar qualifications to the evaluators. Their responses were correlated with the total test score, showing a strong correlation between ±0.50 and ±1 (ranging from 0.54 to 0.83, with an average of 0.76), demonstrating content validity. Additionally, the p-values for each item were below 0.05 ($p < 0.05$), confirming the overall validity of the test statements.

Data Collection and Analysis

For data collection, the translation test was first translated from English into Persian using Fastdic Translator. Both the Persian translations and the original 80 English statements were given to the evaluators. They evaluated Fastdic Translator's output based on three areas: syntax, semantics, and pragmatics. This was done according to Wilss's (1982) evaluation matrix.

The evaluators used a Likert-scale questionnaire. This questionnaire had five continuums: wrong, inappropriate, undecidable, correct, and appropriate. The scale helped evaluate how accurately the MT system produced translations in these three areas. ChatGPT was also used to improve the statements marked as wrong, inappropriate, or undecidable by the evaluators.

For data analysis, the percentage and mean score for each category were calculated. The results were then shown in tables. The Wilcoxon signed-rank test was used to check if the evaluators' assessments of the 80 Persian translations were consistent within each area of evaluation.

4. Results and Discussion

Syntactic Evaluation

Table 1 shows that the evaluators agreed with the syntactic correctness of the translations made by Fastdic Translator for both narrative and descriptive statements. Specifically, Fastdic Translator correctly translated 70% of narrative statements and 58% of descriptive statements. Furthermore, 24% of narrative statements and 37% of descriptive statements were deemed appropriate by the evaluators. The table also indicates that the evaluators agreed on the correct translation of a large majority (72%) of the argumentative statements. Overall, the evaluators found that 87% of the 60 Persian translations produced by Fastdic Translator were syntactically correct.

Table 1

Percentages of the Evaluators to Syntactic Evaluation of the 60 Statements

Type of Statements	Wrong	Inappropriate	Undecidable	Correct	Appropriate	*N*	
	%	%	%	%	%	%	*M*
Narrative	-	-	7.0	70.0	24.0	100.0	4.16
Descriptive	-	-	5.0	58.0	37.0	100.0	4.31
Argumentative	7.0	18.0	4.0	48.0	24.0	100.0	3.64
Total	2.0	6.0	5.0	59.0	28.0	100.0	4.03

A one-sample Wilcoxon signed-rank test was run to see if there was a significant relationship between the evaluators' assessments and the syntactic correctness of the Persian translations made by Fastdic Translator for each type of statement. The results, shown in Table 2, reveal that the p-values for all types of statements were below .05 ($p < .05$). This means the results are statistically significant.

The mean scores for each type of statement were all above the midpoint of 2.5. This shows strong agreement among evaluators about the syntactic correctness of the translations. The mean score for narrative statements was 4.16, while for descriptive statements, it was 4.31. This indicates that Fastdic Translator produced translations with high syntactic correctness. The mean score for argumentative statements was 3.64, which also shows above-average syntactic correctness.

Additionally, Table 2 shows that the overall p-value for the 60 Persian translations was 0.0, which is less than .05 ($p < .05$). The combined mean score for all statements was 4.03 out of 5, exceeding the midpoint (4.03 > 2.5). This indicates that the evaluators significantly agreed on the overall syntactic correctness of the Persian translations provided by Fastdic Translator.

Table 2

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

Statement Types	*N*	*MDN*	*p*
Narrative	20	4	.01
Descriptive	20	4	.002
Argumentative	20	4	.000

Total	60	4	.000

Semantic Evaluation

Table 3 shows that the evaluators did not agree on the semantic adequacy of most narrative statements generated by Fastdic Translator. They found that 64% of the narrative statements lacked semantic correctness. However, they agreed on the correctness of more than half of the Persian translations of descriptive statements, with 52% being accurate. For argumentative statements, the evaluators agreed that 60% of the translations were correct. Overall, the evaluators felt that the number of semantically correct translations produced by Fastdic Translator (49%) was nearly equal to the number of incorrect translations (48%).

Table 3

Percentages of the Evaluators to Semantic Evaluation of the 60 Statements

Type of Statements	Wrong	Inappropriate	Undecidable	Correct	Appropriate	*N*	
	%	*%*	*%*	*%*	*%*	*%*	*M*
Narrative	24.0	40.0	2.0	24.0	10.0	100.0	2.55
Descriptive	16.0	29.0	3.0	36.0	16.0	100.0	3.1
Argumentative	14.0	22.0	5.0	45.0	15.0	100.0	3.27
Total	18.0	30.0	3.0	35.0	14.0	100.0	2.93

A one-sample Wilcoxon test was conducted to assess how much the evaluators agreed with the semantic adequacy of each type of statement. The goal was to see if there was a significant relationship between their opinions and the semantic adequacy of the Persian translations. The results showed that the p-values for the descriptive and argumentative translations were below .05 ($p < .05$) (see Table 4). This indicates a significant relationship between the evaluators' assessments and the semantic adequacy of these translations. The mean scores for these translations were also above the average of 2.5, with descriptive translations scoring (MD = 3.1) and argumentative translations scoring (MA = 3.27), reflecting higher-than-average semantic adequacy.

On the other hand, for narrative statements, the p-value was above .05 ($p > .05$). This means the evaluators' opinions did not differ significantly from the average. The mean score for narrative statements was almost equal to the average (2.55 ≈ 2.5), suggesting that their semantic adequacy was average.

Additionally, Table 4 shows that the p-value for all 60 Persian translations was also above .05 ($p > .05$). The overall mean score of 2.93 was slightly higher than the average of 2.5. This finding indicates that, overall, there was a moderate level of agreement among evaluators regarding the semantic adequacy of more than half of the Persian translations produced by Fastdic Translator, suggesting an almost average level of semantic adequacy (MT = 2.93).

Table 4

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

Statement Types	*N*	*MDN*	*p*

Narrative	20	2	.07
Descriptive	20	4	.002
Argumentative	20	4	.000
Total	60	4	.001

Pragmatic Evaluation

According to Table 5, the evaluators did not agree with the pragmatic correctness of any of the 20 idiomatic statements (100%) produced by Fastdic Translator.

Table 5

Percentages of the Evaluators to Pragmatic Evaluation of the 20 Statements

Type of Statements	Wrong	Inappropriate	Undecidable	Correct	Appropriate	*N*	
	%	*%*	*%*	*%*	*%*	*%*	*M*
Idiomatic	37.0	63.0	-	-	-	100.0	1.63

A one-sample Wilcoxon signed-rank test was done to check if the evaluators agreed on the pragmatic adequacy of the 20 Persian translations made by Fastdic Translator. Based on Table 6, the p-value for the idiomatic statements was less than .05 ($p < .05$). Furthermore, the mean score for these statements was lower than the theoretical mean/median (1.63 < 2.5), indicating that the evaluators did not find the idiomatic translations to be pragmatically adequate.

Table 6

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

N	*MDN*	*p*
20	2	.01

5. Discussion

According to the results, Fastdic Translator performed well in translating the 80 statements into Persian. It mostly produced grammatically correct Persian translations that followed the SOV structure of Persian. However, there were some awkward translations, particularly concerning tense and subject-verb agreement, especially in complex sentences. For example, consider the English sentence the Mongol style, designed to overawe the viewer, was marked by towering entrance portals, colossal domes, and vaults reaching up into the skies. This sentence contains participle constructions: designed to overawe the viewer (a past participle phrase modifying the Mongol style) and reaching up into the skies (a present participle phrase modifying vaults).

While the participle constructions were translated correctly, the machine incorrectly used the past tense for the main verb was marked, translating it as مشخص شد. This error affects the semantic adequacy of the Persian translation. The correct tense should be the past continuous tense to enhance the understandability of the statement. Thus, was marked should be translated as مشخص می‌شد.

In this example, Fastdic Translator demonstrated its weakness in applying the subject-verb agreement rule when translating the English phrase the northern slopes of the Alborz Mountains are densely... and form the largest area of vegetation... into Persian. That is to say, it incorrectly used singular verbتشکیل می‌دهد (forms) for the plural subject دامنه‌های شمالی رشته‌کوه‌های البرز (the northern slopes of the Alborz Mountains), which is not grammatically correct in Persian. Fastdic Translator made a similar error when translating the sentence as the results indicate, translations Google Translate produced were acceptable to a certain degree into همانطور که نتایج نشان می‌دهد، ترجمه‌های انجام شده توسط گوگل ترنسلیت تا حدی قابل قبول بوده است.

In the Persian translation produced by Fastdic, the subject نتایج (results) is plural, but the verb نشان می‌دهد (indicates) is singular. In Persian, as in English, the verb must agree with the subject in number, so the correct verb form should have been plural نشان می‌دهند (form) fir plural subject نتایج (results). Similarly, the verb بوده است (has been) is singular, but the subject ترجمه‌ها (translations) is plural, requiring the verb form to match: بوده اند (have been).

These deficiencies could stem from Fastdic Translator's use of a less sophisticated translation algorithm compared to more advanced systems like Google Translate. As a result, it may apply simplified rules for subject-verb agreement, often defaulting to singular verbs, especially in complex sentences. Additionally, Fastdic Translator might lack the advanced contextual analysis tools necessary to ensure proper agreement between subjects and verbs, adjusting the verb form accordingly.

The results also showed that Fastdic Translator was moderately successful in producing semantically correct Persian translations. It is rather to say, the semantic quality of its Persian translations was rated slightly above average. For example, the English phrase Neville had been trying to catch his eye was translated as نویل سعی می‌کرد چشم او را جلب کند. This literal translation resulted in a loss of idiomatic meaning. The phrase catch his eye is idiomatic and implies trying to make eye contact to gain attention, but the Persian translation چشم او را جلب کند suggests something more superficial, like physically attracting someone's eye, which can be unclear or misleading in Persian.

After post-editing the original phrase with ChatGPT, a more contextually and semantically accurate translation was produced: نویل سعی می‌کرد نگاه/توجه او را جلب کند. In this version, the word چشم (eye) was replaced with نگاه (gaze) or توجه (attention) to convey the intended meaning of attracting someone's attention.

A similar issue occurred in the translation of Hermione hung her head. Harry was speechless into هرمیون سرش را آویزان کرد. هری لال بود. In the first sentence, the English idiom hung her head typically conveys feelings of shame, sadness, or defeat. While the Persian translation سرش را آویزان کرد is a literal rendering, it somewhat fails to capture the emotional nuance of the expression.

The translation of the second sentence, Harry was speechless, was particularly problematic in terms of both meaning and naturalness, because Fastdic Translator rendered it literally as هری لال بود, suggesting a permanent inability to speak. This translation misinterprets the English term speechless, which in this context means being temporarily stunned or shocked into silence, not physically unable to speak. After post-editing by ChatGPT, the sentence was translated more appropriately as هری زبانش بند آمده بود (Harry was at a loss for words), which accurately reflects the intended meaning.

Similarly, the first sentence, Hermione hung her head, was revised by ChatGPT to هرمیون سرش را پایین انداخت (Hermione lowered her head) to better capture the emotional tone, avoiding the overly literal translation that does not convey the nuance of sadness or defeat.

From a pragmatic standpoint, Fastdic Translator's performance was disappointing. All idiomatic expressions were translated inaccurately and with poor quality. For instance, the phrase there are different strokes for different folks, so not everyone will like this restaurant was translated by Fastdic as سکته مغزی های مختلف برای افراد مختلف وجود دارد، بنابراین همه این رستوران را دوست ندارند, containing significant pragmatic errors. The literal translation not only mismatches the original meaning but also fails to convey the idiom's intent. This error stems from Fastdic's inability to handle idiomatic language effectively, translating it word-for-word rather than capturing its true sense.

In a broader sense, pragmatically, the idiom different strokes for different folks carries a cultural and context-specific meaning that cannot be grasped by translating the words individually. In Persian, the phrase سکته مغزی refers to a medical condition (a stroke), which is entirely has nothing to do with the idiom's original meaning about personal preferences. As a result, the translation loses its intended pragmatic function and introduces a completely irrelevant medical concept. This literal translation also doesn't succeed in conveying the idea of the SL idiom different people have different tastes. After post-editing by ChatGPT, a more pragmatically appropriate translation was produced: هر کسی سلیقه خودش را دارد، بنابراین همه این رستوران را دوست نخواهند داشت (everyone has their own taste, so not everyone will like this restaurant), which preserves the intended meaning.

All in all, Fastdic Translator performed like a basic MT, lacking the advanced linguistic models required to handle subject-verb agreement and appropriate tense, especially in complex sentences. It also produced literal translations that didn't fully convey the meaning of the SL items. This issue extended to idiomatic phrases, where Fastdic Translator was unable to select suitable equivalents for SL items, leading to inappropriate translations. Its tendency to translate too literally is one of its major deficiencies, particularly impacting output quality in terms of pragmatics. In pragmatic translation, the goal is not only to convey the literal meaning but also to reflect cultural and contextual nuances. Fastdic's translations fall short of this goal as it often rendered unclear or inappropriate Persian translations.

6. Conclusion

The aim of this study was to critically assess the output quality of Fastdic Translator, a domestic MT) based on Wilss's matrix, including syntax, semantics, and pragmatics, to highlight the system's strengths and weaknesses. This research is part of a project focused on the achievements of Iranian software developers in creating domestic MT systems, namely Targoman Translator, Abadis Translator, and Fastdic Translator. The emphasis is on translation quality and post-editing processes.

The results indicated that Fastdic Translator generally produced grammatically correct Persian translations, although it displayed some deficiencies in subject-verb agreement, particularly in more complex sentences. Furthermore, the findings suggested that Fastdic Translator was somewhat successful in generating semantically accurate translations. However, in terms of pragmatic evaluation, the results revealed that it did not do well in a performance translation because it produced completely inappropriate translations for idiomatic expressions and phrases with implied meanings.

The primary weakness of Fastdic Translator lies in its difficulty managing implied meanings and selecting suitable equivalents for SL items, which require more advanced cognitive processing. Wilss (1982) emphasizes the importance of recognizing functional dependencies rather than simply focusing on formal structures for effective TL reproduction of such items.

In conclusion, while Fastdic Translator performs well regarding grammatical accuracy and somewhat in semantic accuracy in Persian translations, it exhibits weaknesses in subject-verb agreement in complex sentences and performs poorly in pragmatic translation, particularly with idiomatic expressions and implied meanings, where its translations are often inappropriate.

This issue arises from an over-reliance on literal translations, which prevents conveying deeper meanings that require cultural and contextual nuances. As a result, the quality of the output is significantly affected, making translations inappropriate. This limitation shows that post-editing is necessary for translations produced by Fastdic Translator.

Besides, using ChatGPT for post-editing greatly improves translations. It corrects errors and ensures grammatical standards are met while conveying the intended meaning more accurately. Although there were a few instances where it relied on literal translations and produced inappropriate results, these issues were fixed when asked to retranslate them. Nevertheless, this does not diminish ChatGPT's value; it remains a powerful and reliable AI tool for both end-users and translators.

The study offers helpful recommendations for Iranian MT developers, especially for the Fastdic Translator team. It is recommended that they create more advanced models for subject-verb agreement and tense handling, particularly for complex sentences, to improve grammatical accuracy.

Additionally, integrating contextual translation models can help move beyond literal translations and better convey intended meanings. To enhance the system's understanding of pragmatics, it would be effective to include features such as a contextual translation mode and a database of cultural idioms and phrases. These features would consider cultural and contextual nuances, ensuring translations are meaningful within their broader social or cultural contexts.

REFERENCES

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. (2021a). 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.v2i1.122

Abdi, H. (2021b). Familiarity of Iranian M.A. Translation Students with ICT Tools. *International Journal of Linguistics and Translation Studies, 2*(1), 17-26. https://doi.org/10.36892/ijlts.v2i1.118

Abdi, H. (2020). Translation and technology: Investigating the employment of computer-aided translation (CAT) tools among Iranian freelance translators. *Theory and Practice in Language Studies, 10*(7), 811-818. https://doi.org/10.17507/tpls.1007.13

Abdi, H. (2019). The Status of ICT employment among Iranian freelance translators. *International Journal of Innovation and Research in Educational Sciences, 6*(3), 339-349. https://www.ijires.org/index.php/issues?view=publication&task=show&id=501

Aghai, M. (2024). ChatGPT vs. google translate: Comparative analysis of translation quality. *Iranian Journal of Translation Studies, 22*(85), 85-100. https://dorl.net/dor/20.1001.1.17350212.1403.22.1.9.2

Banerjee, S., & Lavie, A. (2005). *Meteor: An automatic metric for MT evaluation with improved correlation with human judgments* [Conference session]. The ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, Prague, the Czech Republic. https://aclanthology.org/W05-0909

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*. Singapore: Lonely Planet Publications Pty Ltd. https://www.amazon.com/Lonely-Planet-Iran-Travel-Guide/dp/1786575418

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, the Czech Republic. https://aclanthology. org/W07-0718

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 thesis, Hacettepe University Graduate School of Social Sciences). http:// www.openaccess. hacettepe.edu.tr:8080/xmlui/handle/11655/1285?locale-attribute=en

Escribe, M. (2019). *Human evaluation of neural machine translation: The case of deep learning* [Conference session]. Second Workshop on Human-Informed Translation and Interpreting Technology, Varna, Bulgaria. https://doi.org/10.26615/issn.2683-0078.2019_005

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M., Awadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation. *arXiv:2302.09210v1 [cs.CL]*. https://doi.org/10.48550/ arXiv.2302.09210

House, J. (1997). *Translation quality assessment: A revisited model*. Narr.

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

Prathyusha, J. (2022). *Building generalizable models for discourse phenomena evaluation and machine translation* [Doctoral thesis, School of Computer Science & Engineering, Nanyang Technological University]. https://doi.org/10.32657/10356/165027

Philipp, K. (2007). Statistical and hybrid machine translation between all European languages. EUROMATRIX Consortium. https://cordis.europa.eu/docs/projects/ cnect/1/034291/080/publishing/readmore/EUROMATRIX-executive-summary.pdf

Pospelova, O. & Rowda, J (2016, June). Human evaluation of machine translation. *eBay*. https://www.tech.ebayinc.com/engineering /human-evaluation-of-machine-translation/

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.

Sartipi, A., Dehghan, M., & Fatemi, A. (2023, February 1). An evaluation of Persian-English machine translation datasets with transformers. *arXiv:2302.00321 [cs.CL].* https://doi.org/10.48550/arXiv.2302.00321

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. & Maryam, A. (2017). Exploring computer-aided translation competences of some Iranian translators. *International Journal of English Language & Translation Studies, 5*(1), 78-87. https://eltsjournal.org/archive/value5%20issue1/11-5-1-17.pdf

Ulitkin, I., Filippova, I., Ivanova1, N., & Poroykov, A. (2021). Automatic evaluation of the quality of machine translation of a scientific text: the results of a five-year-long experiment. *TPACEE, 284*, 1-12. https://doi.org/10.1051/e3sconf/202128408001

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

Zakaryia, M. (2020). *Diachronic evaluation of Google Translate, Microsoft Translator and Sakhr in English-Arabic translation* (Doctoral dissertation, the University of Western Australia). Western Australia Digital Archive. https://api.research-repository.uwa.edu .au

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