
J. New Trends Engl. Lang. Learn. 2025; 4 (Special Issue) Article si-06

Journal of New Trends
in
English Language Learning (JNTELL)

https://doi.org/10.57647/JNTELL.2025.si-06

# A Comparative Assessment of Online Machine Translation Software in Translating English Medical Terms into Persian

Masoud Ghavidast[1], Mahnaz Azad[2*]

[1]*Department of Translation Studies, NT.C., Islamic Azad University, Tehran, Iran*

[2]*Department of English Language Teaching, ET.C., Islamic Azad University, Tehran, Iran*

***Corresponding author:** mahnaz.azad@yahoo.com

**Original Research**

Received:
1 September 2025

Revised:
27 September 2025

Accepted:
30 September 2025

Published in Issue:
19 October 2025



**Abstract**

With the ever-increasing use of online translators, it is more crucial than ever to evaluate their efficiency in various fields, including the medical field, where terminology can pose significant challenges in translation due to the technical and specialized nature of these terms. Regarding the significance of the issue, this study compared the performance of four free online machine translation software (Google Translate, Yandex, Aryanpour, and Faraazin) in translating English medical terms into Persian. The study employed a comparative corpus-based, quantitative approach to evaluate the accuracy, efficiency, and precision of each machine translation tool. Three hundred medical phrases were randomly selected from the Dorland's Medical Dictionary to be machine translated, analyzed, and compared. The Dorland medical dictionary was used as the criterion in this study. The results identified Google Translate as the most effective and precise tool, providing more accurate translations compared to the other translation tools. The study also demonstrated that although the proportions of 'proper' translation were similar across all four MT tools, the proportions of 'improper' translation and 'untranslated' items were significantly different across the tools. Hence, the findings have some implications for those involved in medical translation indicating that while MT tools can provide a reliable translation of most medical terms, it is imperative to exercise caution when using these tools for critical and high-stakes medical contexts.

**Keywords:** Aryanpour, Faraazin, Google Translate, Online machine translation, Medical terminology, Yandex Translate

**Cite this article:** Ghavidast, A., & Azad, M. (2025). A comparative assessment of online machine translation software in translating english medical terms into persian, *Journal of New Trends in English Language Learning, 4* (Special Issue), Article si-06. https://doi.org/10.57647/JNTELL.2025.si-06

## 1. Introduction

Machine translation (MT), as a subcategory of computational linguistics, has been developed to translate texts. MT is a computerized process for translating texts that differs from computer-aided translation (CAT) tools, which rely on online dictionaries and terminology databases. In fact, MT

aims to automate the translation procedure. In recent years, online translators have become one of the most frequently used MT tools, and even non-academic users benefit from their remarkable capacities in the translation field (Arnold et al., 1994; Khadem Olhosseiny & Askari, 2024; Taleghani & Pazouki, 2018). Furthermore, systematic research has revealed that learners prefer using online translation gadgets and applications (Merschel et al., 2013) while translation experts consider it unproductive and, in most cases, a form of academic dishonesty (Hellmich, 2019).

Alternatively, medical translation is a cognitive process requiring specialized skill and expertise to make sure that the translation accurately conveys the exact meaning of medical terminologies and avoids any potential errors or misunderstandings (Montero-Martínez & de Quesada, 2003; Munassar Awadh, 2024; Sedaghat Rostami & Bigdeli, 2016; Williams & Chesterman, 2002). Medical terminology is a fundamental resource to any sort of health care service, function, and task that encompasses a diverse range of specialized terms used in various contexts (Buzarna-Tihenea, 2015a & b). As Nădrag (2016) noted, medical translation is challenging for several reasons, including the use of jargon and idiosyncratic phrases that may sound unfamiliar in everyday conversations. Furthermore, specific medical terms have been developed to describe diseases, medications, dosages, and examination methods, which can add to the problem (Karwacka, 2015; Nord, 1997; Pekcoşkun Güner, 2020). Thus, a translator in the medical area must be an expert, because any mistake made in the translation process could have a significant and irreparable impact on clinical outcomes (Betancourt & Jacobs, 2000). Also, as stated by MOGI's translation team (2011), while translating medical texts, a translator must possess the ability to effectively convey messages to both professional healthcare staff and non-professional audiences in the field, which necessitates a high level of accuracy and consistency (Gharsa, 2015; Nădrag, 2016). The popularity of online machine translation software, the substantial role they play in the translation field, and the challenge they might pose in the medical field indicate that evaluating the quality of their performance in translating this type of text is absolutely vital, which was a motivation to conduct this study. Therefore, this investigation was an attempt to compare the efficiency and precision of the performances of two foreign-based (i.e., Google Translate and Yandex Translate) and two Iranian-based (i.e., Faraazin and Aryanpour) online translation software in comparison to a reliable medical reference book (the Dorland's Medical Dictionary) regarding the translation of medical terms.

### 1.1. Research Questions

Following the intention of this inquiry, these research questions were posed:

**RQ1:** To what extent is online machine translation software accurate in the translation of English medical terms into Persian?

**RQ2:** Which of the online machine translation software is more efficient in translating English medical terms into Persian?

**RQ3:** Are there any statistically significant differences between the degrees of precision of the MT software examined in this study?

**RQ4:** To what extent are the online MT software examined in this study reliable tools for translating English medical terms into Persian?

## 2. Literature Reviews

### 2.1. Machine Translation

Machine Translation (MT) is a term developed over time and has been interpreted differently. Originally, MT referred to a fully computerized process without human contribution, while the European Association of Machine Translation broadened the definition, incorporating the use of computers for various types of translation. The International Association of Machine Translation (IAMT) described MT as an input receiver that produces the corresponding output with no guarantee of its reliable quality (Wang et al., 2022).

Neither of the definitions presented earlier includes any mention of human intervention in MT. Although some scholars, such as Costa-jussà et al. (2012), do acknowledge the possibility of some form of human involvement in the translation process of different languages via MT, others have conflicting opinions on what role human intervention should play in MT (Aiken & Balan, 2011; Archer, 2002; Randhawa et al., 2013). Despite this lack of consensus, the term MT is still commonly used to refer to automated systems as well as those that involve some level of human involvement, since no other term has replaced it (Somers, 2003). The idea of using digital computers for translation was first proposed by Booth in 1946. Since its inception, MT has undergone two major strategies for development: *Rule-Based* and *Corpus-Based* (Statistical) strategies. Rule-based MT is mostly used for providing dictionaries and grammatical programs, and it requires more information about the linguistic rules of both intended languages (Kaliyadan & Pillai, 2010) while Corpus-based or Statistical method was an approach that uses data to automatically generate translation knowledge without relying on human skills (The Google Translate is the most recognized instance

of the Statistical system) (Deilen et al., 2024; Zappatore & Ruggieri, 2024).

The triumph of SMT models has inspired many scholars to suggest fresh models such as factored SMT models, hierarchical SMT models, and SMT models that can boost the efficiency of SMT techniques (Chiang, 2007; Koehn et al., 2007; Wang, 2006). In 2003, *Phrase-Based* MT was introduced, which improved the translation quality (Koehn & Knowles, 2017; Och & Ney, 2003).

In light of the significant advancements in deep learning techniques in other fields, the *Neural Translation Models* (NMT) were introduced (Bahdanau et al., 2015; Sutskever et al., 2014) whose fundamental concept involved converting the SL into a compact, meaningful representation, which was then utilized to create the translation through the application of an attention mechanism (Dong et al., 2015; Zhang & Zang, 2020). In 2016, Google released an NMT system, soon followed by similar systems from other institutions. Therefore, it only took roughly one year from the initial proposal of NMT in 2014 for it to be implemented online, whereas it took approximately 16 years for SMT systems to be adopted for online services (Wu et al., 2016). Subsequently, other models were introduced, leading to further substantial enhancements in translation quality.

This remarkable progress prompted extensive debates on whether MT can match the quality of human translation (Koehn & Knowles, 2017; Vaswani et al., 2017). In light of the impressive achievements of NMT, numerous techniques were developed to improve both the quality and efficiency of multilingual translation (Gehring et al., 2017; Wang et al., 2022).

### 2.2. Medical Field and Machine Translation

Healthcare professionals resort to using online MT tools as a final option when alternative language services are not accessible, and there is a mounting body of proof to support this claim (Vieira et al., 2020). According to a survey conducted by Turner et al. (2013), nearly one-third of local health departments resorted to MT to translate written materials as they did not have sufficient funds for professional translations. In critical situations, where all other options are exhausted, healthcare professionals have effectively utilized MT to communicate directly with their patients (Kaliyadan & Pillai, 2010; Khadem Olhosseiny & Askari, 2024; Pekcoşkun Güner, 2020).

However, it is uncertain how healthcare professionals can use these tools efficiently without any adverse outcomes except for refraining from their usage (Moberly, 2018). Some medical practitioners have expressed their inclination to use translation tools since open-ended MT systems are more likely to produce erroneous translation (Khoong et al., 2019; Randhawa et al., 2013). Phrase-based translation apps limit the user's communication options, but are more dependable since all the phrases available have been professionally translated (Dew et al., 2018; Spechbach et al., 2017).

Several studies have assessed online translators' performances in medical communication. Khanna et al. (2011) studied the degree of Google Translate (GT) accuracy in translating patients' information and found low fluency errors. Patil and Davies (2014) worked on GT to analyze its accuracy in translating medical phrases.

The result showed that GT has 57.7% accuracy, and it is not trustworthy for critical medical communication, but it can be used along with human translation services. Similarly, Chen et al. (2016) evaluated the use of GT in Chinese-English translation and vice versa. The findings indicated that the association between the level of difficulty and the accuracy of the translations provided by GT was negative.

Studies conducted in Iran on translating medical terms have exclusively investigated the models of translation employed. For example, Yassi (2018) investigated the translation of English-Persian surgical-medical terms based on Newmark's model. The results highlighted a lack of proper Persian equivalence for the medical terms. Also, Rahmanian (2018) compared the human translation of the medical terminology based on Reiss's Model and found that neither translation students nor medical students could independently produce high-quality translation without the use of MT.

Honari (2018) examined the effectiveness of MT in the field of technical and scientific documents (abstracts) from English to Persian and vice versa, concluding that in translating scientific and technical terminology, MT is neither suitable nor effective.

Franca Daniele (2019) conducted a comprehensive study to analyze the efficiency of GT in English-Italian translation of medical content and found that the average percentage of translation errors was 15%, which suggests that this online translation tool performs satisfactorily when translating complex medical essays and writings.

Furthermore, Khoong et al. (2019) assessed GT capacity to translate medical guidelines. The study found that its new algorithm translated the instructions with increased precision and reduced errors compared to the previous ones. Also, Birkenbeuel et al. (2021) studied the effectiveness of GT in the accurate English-Spanish translation of sentences commonly used in healthcare situations. The findings demonstrated that the precision of the translation varied depending on the text length.

As the literature revealed, although some studies have evaluated the translation of medical terms, they have typically concentrated on a general evaluation of online MT software based on different models and no study has compared the accuracy and efficiency of MT tools as investigated in this inquiry so far.

Therefore, this research sought to fill the gap by examining the accuracy, efficiency, and precision of the performances of two foreign-based (i.e., Google Translate and Yandex Translate) and two Iranian-based (i.e., Faraazin and Aryanpour) online translation software in comparison to the Dorland's Medical Dictionary (as the most reliable medical reference book) regarding the translation of medical terms.

## 3. Method

### 3.1. Design

This study employed a quantitative comparative approach in which the performances of four online translation software were compared with a reliable source, considering the translation of medical terminologies to Persian.

### 3.2. The Corpus

Three hundred medical phrases concerning diseases and disorders formed the corpus of this study, selected from the Dorland's Medical Dictionary as the main resource for both selecting the phrases and checking the accuracy of the translations (Newman Dorland, 1890).

### 3.3. Instruments

Four different online MT software, namely, Google Translate (GT), Yandex Translate (YT), Faraazin (F), and Aryanpour (A), were used to translate the medical terms into Persian, and their accuracy, efficiency, and precision were evaluated based on the Dorland's medical dictionary as a reliable criterion.

#### 3.3.1. *Google Translate*

Google Translate (GT) is believed to be the most extensively employed online translation service. According to Kroulek (2016), more than 500 million individuals avail themselves of this facility to translate approximately 100 billion words every day, which encompasses 103 languages. Google introduced GT in 2006, utilizing a statistical, phrase-based MT model which was later updated to incorporate a neural system (McGuire, 2018; Turner, 2016).

#### 3.3.2. *Yandex Translate*

Yandex Translate is an online tool that enables users to convert text or web pages into different languages (Yandex, 2017). It uses statistical MT, analyzes thousands of translated texts, and creates a dictionary of one-word translations. To translate a piece of text, the computer consults a word database and language models to determine its meaning in the context (Yandex, 2023). In 2017, YT combined both statistical and neural MT models, with which words, sentences, or web pages can be translated in 98 languages.

#### 3.3.3. *Faraazin Translator*

The Faraazin Smart Translator project started in 2009 in the Text Processing and Natural Languages Laboratory of Tehran University and continued with the support of the Communication and Information Technology Research Center. FT uses the joint neural translation engine developed by two companies, Vera Azar Adan and Targoman Smart Processing, and by using this translation engine, which is published as an open source, it provides higher translation quality than all existing translation engines (Faraazin, 2023).

#### 3.3.4. *Aryanpour*

Aryanpour Dictionary is free to use and can translate texts and documents from one language to another using powerful multilingual tools. It is powered by the Machine Learning Translation API and employs the most up-to-date machine learning techniques for optimum results (Aryanpour, 2023).

### 3.4. Data Collection Procedure

Before choosing the corpus, the researchers consulted with some experts in the medical field, considering the category of terms to be used. Considering the importance of the diseases and health problems suggested by the experts, the source terms and translated terms were extracted from the Dorland's medical dictionary based on a random sampling method. In the next stage, the selected terms were inserted into each translation software separately, and the translations were obtained. Next, the definitions provided by different translation software were recorded and compared to determine their accuracy, efficiency, and precision based on the Dorland's medical dictionary. The performance quality and reliability of the MT software were assessed and compared using SPSS software version 24. For data analysis, In order to assess the extent to which each online MT software is accurate in translating English medical terms into Persian, the percentages of correct translations were calculated. Then, to compare the efficiency of these translation tools, pairwise comparisons were conducted via using the McNemar-Bowker Test. Moreover, to determine if there were any statistical differences between the precisions of the MT software examined in this study, a Chi-Square test for

independence was performed. Finally, a comprehensive analysis was conducted via using Fleiss' Kappa coefficient to evaluate the reliability of the online MT software utilized in the translation of English medical terms to Persian.

## 4. Results

The translation tools analyzed in this study included Google Translate, Yandex Translate, Aryanpour, and Faraazin. To probe the research questions, translation levels were categorized as: E (Eponyms: Here, the name of a disease or disorder that is also the name of the person who first discovered it), F (Wrong translation), L (Loan Translation), NT (Not translated), S (Correct translation with similar equivalence), and T (Correct translation). Then, various statistical analyses were performed on the data collected which are explained below.

### 4.1. Data Analysis

4.1.1. *The First Research Question*

*To what extent is online machine translation software accurate in the English to Persian translation of medical terms?*

To evaluate the extent to which each online MT software is accurate in translating English medical terms into Persian, the percentages of correct translations were calculated for each tool based on the provided data. In this study, 'accuracy' is defined as the percentage of the terms translated correctly measured via the following formula:

$$\text{Accuracy (\%)} = \frac{\text{Number of Correct Translated Terms}}{\text{Total Number of Terms)} * 100.}$$

Table 1 provides both the absolute counts (N) and the corresponding percentages (%) of each translation tool and its translation level. Table 1 presents a comprehensive comparison of translation accuracy for the four translation tools across different translation levels. In this step, the focus was on assessing the percentage of correct translation denoted as 'T'. The results indicate distinct performance variations among the examined tools. According to the Table, *GT* achieved an accuracy of 37.00%, demonstrating its with an accuracy rate of 33.30%, followed by *AT* with the accuracy of 33.00% showcased their reliability in the context of medical terminology, while *FT* achieved an accuracy rate of 17.00%. Figure 1 displays the schematic presentation of the findings.

**Table 1.** Comparison of Translation Accuracy of Different MT at Various Translation Levels

| Levels | GT | | YT | | AT | | FT | |
|---|---|---|---|---|---|---|---|---|
| | N | % | N | % | N | % | N | % |
| E | 25 | 8.30 | 30 | 10.00 | 22 | 7.30 | 5 | 1.70 |
| F | 98 | 32.70 | 109 | 36.30 | 99 | 33.00 | 37 | 12.30 |
| L | 19 | 6.30 | 19 | 6.30 | 21 | 7.00 | NA | NA |
| NT | 6 | 2.00 | 2 | 0.70 | 15 | 5.00 | 193 | 64.30 |
| S | 41 | 13.70 | 40 | 13.30 | 44 | 14.70 | 14 | 4.70 |
| T | 111 | 37.00 | 100 | 33.30 | 99 | 33.00 | 51 | 17.00 |

* E: Eponyms; F: Wrong translation; L: Loan Translation; NT: Not translated; S: Correct translation with similar equivalence; T: Correct translation

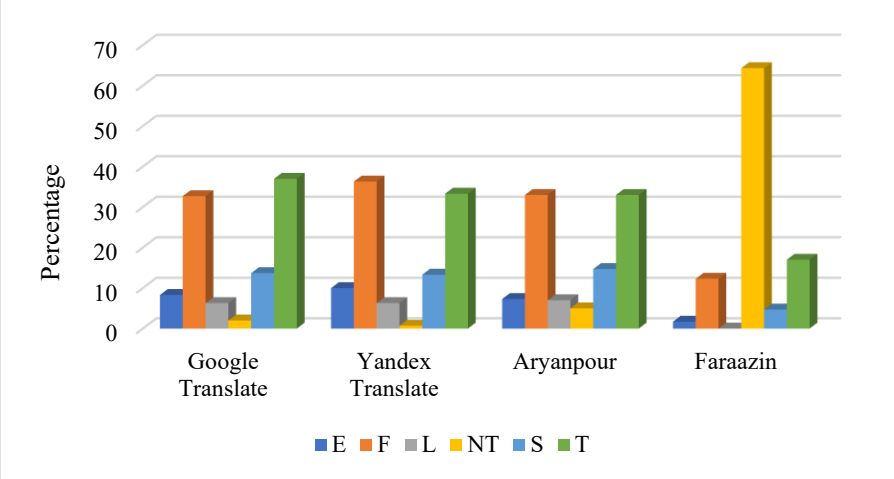


**Figure 1.** Comparison of Translation Accuracy of Different MT at Various Translation Levels

Moreover, to assess the accuracy of online MT software in Persian translation of medical terminologies, the translated terms were categorized into 'proper' and 'improper' categories. In the 'proper translation' category, both 'T' (Correct translation) and 'S' (Correct translation with similar equivalence) were included as they represent accurate translation. Conversely, in the 'improper translation' category included 'E' (Eponyms), 'L' (Loan Translation), and 'F' (Wrong translation) as they represent various forms of inaccuracy in translation. Also, the category 'NT' (Not translated) was included, which represented the terms the translation tools were unable to handle or for which there were no suitable equivalents in Persian. Table 2 provides a comparative analysis of the performances of the four online MT tools in question. The analysis in Table 2 categorizes translation outcomes into three main groups: 'Proper translation', 'Improper translation', and 'Not translated'. Table 2 displays both the absolute counts (n) and the corresponding percentages (%) for each translation tool. Considering 'Proper Translation', *GT* demonstrated the highest percentage of 'Proper translation' at 50.70%, indicating that it successfully provided accurate translations for a substantial portion of the medical terms. *AT* closely followed with an accuracy rate of 47.7% in the 'Proper translation' category; *YT* displayed a similar high accuracy of 46.2% for 'Proper translation', while *FT* showed the lowest percentage in this category at 21.70%.

**Table 2.** Comparative Analysis of the Four Translation Tools' Performances

| | Translation Tool | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | GT | | YT | | AT | | FT | |
| | N | % | N | % | N | % | N | % |
| Proper translation | 152 | 50.70 | 140 | 46.20 | 143 | 47.70 | 65 | 21.70 |
| Improper translation | 142 | 43.30 | 158 | 52.70 | 142 | 47.30 | 42 | 14.00 |
| Not translated | 6 | 2.00 | 2 | 0.7 | 15 | 5.00 | 193 | 64.30 |

'Improper translation' encompasses various forms of inaccuracy, including loan translations, eponyms, and incorrect translations. *YT* had the highest percentage of 'Improper translation' at 52.70%. *AT* closely followed with 47.30% in the 'Improper translation' category. *GT* displayed a percentage of 'Improper translation' at 43.30%, and *FT* exhibited the lowest percentage in this category, at 14.00%. The 'Not translated' category represents terms that the translation tools were unable to handle or for which there were no suitable equivalents in Persian. *FT* had the highest percentage in this category at 64.30% and *AT* showed a percentage of 5.00%. *GT* and *YT* had relatively low percentages in this category, at 2.00% and .70% respectively.

4.1.2. *The Second Research Question*

*Which of the online machine translation software is more efficient in Persian translation of English medical terms?*

To compare the efficiency of different translation tools, pairwise comparisons were conducted via McNemar's test, through which each pair of translation tools was compared to determine which one is significantly better than the other in terms of accuracy. The cross-tabulation analysis of *GT* and *YT* reveals noteworthy differences in their performances in translating medical terms. While both tools excel in categorizing translations as 'Proper translation' (Google vs. Yandex: 50.7% vs. 46.7%), there are significant discrepancies in their handling of 'Improper translation' categories (Google vs. Yandex: 43.3% vs. 52.7%). *GT* produced more untranslated results than *YT* (2.0% vs. 0.7%). The McNemar-Bowker Test indicates a highly statistically significant difference between the two tools (P = .035), favoring Google translate in all terms of translation. Findings of this analysis suggest that *GT* and *AT* both offer viable options for translating English medical terms into Persian, while *AT* exhibited more improper translation (47.3%). *GT* demonstrated a higher accuracy rate in providing proper translations compared to *AT* (50.7% vs. 47.7%). Additionally, *AT* had a higher incidence of terms that could not be translated (5.0% vs.2.0%). The McNemar-Bowker Test confirmed the statistical significance of these differences (p =.021). *GT* appears to offer more efficient Persian translation of English medical terms due to its higher rate of proper translations and lower rate of improper and untranslated terms compared to that of *AT*.

**Table 3**. Comparative Analysis of Translation Outcomes across the Four Translation Tools

| | | Proper Translation | Improper Translation | Not Translated | P-value |
|---|---|---|---|---|---|
| | | YT | | | |
| Google Translate | Proper translation | 131 (43.70) | 21 (7.00) | 0 (0.00) | 0.035 |
| | Improper translation | 8 (2.70) | 133 (43.30) | 1 (0.30) | |
| | Not translated | 1 (0.30) | 4 (1.30) | 1 (0.30) | |
| | | AT | | | |
| Google Translate | Proper translation | 135 (45.00) | 16 (5.30) | 1 (0.30) | 0.021 |
| | Improper translation | 7 (2.30) | 124 (47.30) | 11 (3.70) | |
| | Not translated | 1 (0.30) | 2 (0.70) | 3 (1.00) | |
| | | FT | | | |
| Google Translate | Proper translation | 64 (21.30) | 8 (2.70) | 80 (26.70) | 0.001 |
| | Improper translation | 1 (0.30) | 33 (11.00) | 108 (36.00) | |
| | Not translated | 0 (0.00) | 1 (0.30) | 5 (1.70) | |
| | | AT | | | |
| Yandex Translate | Proper translation | 126 (42.00) | 13 (4.30) | 1 (0.30) | 0.008 |
| | Improper translation | 17 (5.70) | 128 (42.70) | 13 (4.30) | |
| | Not translated | 0 (0.00) | 1 (0.30) | 1 (0.30) | |
| | | FT | | | |
| Yandex Translate | Proper translation | 61 (20.30) | 7 (2.30) | 72 (24.00) | 0.001 |
| | Improper translation | 4 (1.30) | 35 (11.70) | 119 (39.70) | |
| | Not translated | 0 (0.00) | 0 (0.00) | 2 (0.70) | |
| | | FT | | | |
| Aryanpour | Proper translation | 64 (21.30) | 4 (1.30) | 75 (25.00) | 0.001 |
| | Improper translation | 1 (0.30) | 37 (12.30) | 104 (34.70) | |
| | Not translated | 0 (0.00) | 1 (0.30) | 14 (4.70) | |

Note: The McNemar-Bowker Test was used to analyze the relationship between two distinct translation tools.

Moreover, the findings indicate that *GT* outperformed *FT* in terms of both proper translations (50.7% vs. 21.7%) and improper translations (43.3% vs. 14.0%). *GT* demonstrated a higher rate of accurate and contextually appropriate translations, but also exhibited a higher incidence of improper translations. In contrast, a statistically significant proportion of terms that could not be translated by *FT* compared to *GT* (64.3% vs. 2.0%) (P< .001). *GT* clearly performs better in terms of both proper translation and minimizing untranslated results, even though it exhibits a higher rate of improper translation. Overall, *GT* is considered more efficient in this context. When *AT* was compared to *YT*, we found significant differences in translation outcomes. *AT* yielded more 'Proper translation' results, indicating its efficiency in accurate translations (47.7% vs. 46.2%). Furthermore, *AT* had fewer 'Improper translations', suggesting it performed better in this aspect (47.3% vs. 52.7%). However, *AT* exhibited a higher proportion of terms could not be translated in comparison to *YT* (5.0% vs. 0.7%). Considering the statistical significance of the differences (p=.008), if proper translation and the reduction of improper translation are highly valued, *AT* is to be preferred.

However, if minimizing untranslated results is the primary concern, then *YT* is regarded as more accurate in this regard.

The comparison between *FT* and *YT* revealed significant disparities in translation outcomes. *YT* had more 'Proper translation' results, indicating its efficiency in providing more accurate translations (46.2% vs. 21.7%). However, *FT* had fewer 'Improper translations', suggesting it performed better in avoiding mistranslation (14.0% vs. 52.7%). Moreover, a significant difference was observed between the percentages of the terms *FT* could not translate compared to those of *YT* (64.3% vs. 0.7%; p<.001).

Furthermore, *AT* outperformed *FT* in 'Proper translation' (41.7% vs. 21.7%). *FT* had a lower rate of improper translation compared to *AT* (14.0% vs. 47.3%), while *it* produced more untranslated results than *AT* (64.3% vs. 5.0%). The findings indicated a statistically significant difference between the two tools (P < .001), favoring *AT* in terms of proper translation.

4.1.3. *The Third Research Question*

*Are there any statistically meaningful differences between the degrees of precision of the machine translation software examined in this study?*

In order to determine whether a statistically meaningful difference existed between the translation precision of the tools (Proper, Improper, or Not translated), a Chi-square test for independence was performed.

**Table 4.** Comparative Precision Assessment of the MT Software

| | | Outcome | | | | | | P |
|---|---|---|---|---|---|---|---|---|
| | | Proper translation | | Improper translation | | Not translated | | |
| | | N | % | N | % | N | % | |
| Translation Tool | Google Translate | $152_a$ | 30.4% | $142_a$ | 29.3% | $6_b$ | 2.8% | 0.001 |
| | Yandex Translate | $140_a$ | 28.0% | $158_a$ | 32.6% | $2_b$ | 0.9% | |
| | Aryanpour | $143_a$ | 28.6% | $142_a$ | 29.3% | $15_b$ | 6.9% | |
| | Faraazin | $65_a$ | 13.0% | $42_a$ | 8.7% | $193_b$ | 89.4% | |

Note: all proportions are reported at the .05 significance level.

As presented in the Table, the Chi-Square test results presented meaningful variations in the precision of the MT tools examined in this study when translating medical terms (p<.001). Nevertheless, it is noteworthy that both 'proper' and 'improper' translation categories displayed notably higher percentages when compared to the category of terms that were not translated.

4.1.4. *The Fourth Research Question*

*To what extent are the online machine translation software examined in this study reliable tools for English-Persian translation of medical terms?*

To evaluate the reliability of the online MT software utilized in the Persian translation of English medical terms, a comprehensive analysis was conducted using Fleiss' Kappa coefficients to assess the agreement among multiple raters, in this case, the various translation tools employed.

**Table 5.** The Overall Kappa for Evaluating Agreement among Multiple Raters

| | Kappa | Asymptotic Standard Error | Z | P | Lower 95% Asy.CI Bound | Upper 95% Asy.CI Bound |
|---|---|---|---|---|---|---|
| Overall with three categories | 0.38 | 0.02 | 21.92 | 0.001 | 0.35 | 0.42 |
| Overall with six categories | 0.47 | 0.01 | 38.53 | 0 | 0.44 | 0.49 |

The results demonstrated an overall Kappa coefficient of .38 for three distinct categories (proper, improper, and not translated), with a significant p-value (P<.001). This finding indicates a substantial level of agreement among the translation tools regarding the translation of medical terms. Additionally, when considering a more

detailed breakdown into six categories (T: Correct translation, F: Wrong translation, L: Loan Translation, E: Eponyms, NT: Not translated, S: Correct translation with similar equivalence), the Fleiss' Kappa coefficient yielded a value of .47 (P<.001). This result signifies a noteworthy level of agreement among the translation tools concerning the translation of medical terms across these specific categories.

**Table 6.** Kappa for Individual Categories

| Classification | Rating Category | Conditional Probability | Kappa | SE | Z | P | 95% CI | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Three Groups | Proper translation | 0.78 | 0.61 | 0.02 | 26.04 | <0.001 | 0.57 | 0.66 |
| | Improper Translation | 0.68 | 0.36 | 0.02 | 19.31 | <0.001 | 0.41 | 0.50 |
| | Not translated | 0.08 | -0.12 | 0.02 | -5.16 | <0.001 | -0.17 | -0.08 |
| Six Groups | E | 0.65 | 0.63 | 0.02 | 26.51 | <0.001 | 0.58 | 0.67 |
| | F | 0.65 | 0.51 | 0.02 | 21.59 | <0.001 | 0.46 | 0.56 |
| | L | 0.59 | 0.57 | 0.02 | 24.02 | <0.001 | 0.52 | 0.61 |
| | NT | 0.08 | -0.12 | 0.02 | -5.16 | <0.001 | -0.17 | -0.08 |
| | S | 0.64 | 0.59 | 0.02 | 25.05 | <0.001 | 0.54 | 0.64 |
| | T | 0.80 | 0.71 | 0.02 | 30.16 | <0.001 | 0.67 | 0.76 |

E: Eponyms; F: Wrong translation; L: Loan Translation; NT: Not translated; S: Correct translation with similar equivalence; T: Correct translation

Additionally, Kappa coefficients were computed for individual categories of translation ratings, as shown in Table 6. *Proper translation*, which represents translations deemed highly reliable, exhibited a Kappa coefficient of .61 ($p < .001$), signifying strong agreement among the translation tools. *Improper translation*, denoting moderately reliable translations, yielded a Kappa coefficient of .36 ($p < .001$), indicating a moderate level of agreement among the translation tools. However, *Not translated* category, representing unreliable translations, displayed a negative Kappa coefficient of -.12 ($p < .001$), suggesting a significant disagreement among the translation tools.

Furthermore, the full categories of translation levels were also considered. The first category, 'Eponyms', yielded a Kappa coefficient of .63 ($p < .001$), indicating strong agreement among the translation tools regarding the translated terms categorized as eponyms. The second category, 'Wrong translation', exhibiting a Kappa coefficient of .51 ($p < .001$), suggests a substantial level of agreement among the translation tools for wrong translations. The 'Loan Translation' category, displaying a conditional probability of .59 and yielding a Kappa coefficient of .57 ($p < .001$), indicates a noteworthy level of agreement among the translation tools for loan translations.

Moreover, the 'Not translated' category, exhibiting a low conditional probability of .08 and yielding a negative Kappa coefficient of -.12 ($p < .001$), implies substantial disagreement among the translation tools for 'not translated' terms. Also, the fifth category, displaying a Kappa coefficient of .59 ($p < .001$), indicates a strong level of agreement among the translation tools. Finally, the sixth category, with a Kappa coefficient of .71 ($p < .001$), signifies a substantial level of agreement among the translation tools for correct translation.

## 5. Discussion

The focus of this inquiry was to compare the performances of four online MT software in Persian translation of English medical terms, assessed based on the Dorland Medical Dictionary as the criterion. The findings signified that *Google Translate* appeared as the most accurate among the evaluated MT tools. It performed well in terms of proper translation and minimum untranslated results, despite having a higher rate of improper translation. *Yandex Translate* followed as the second most accurate option and performed competitively in terms of proper translation, but had a slightly higher rate of improper translation compared to *Google Translate*. However, it excelled in minimizing untranslated results. *Aryanpour* demonstrated moderate accuracy in providing proper translation and had a higher incidence of untranslated terms. Finally, *Faraazin Translate* ranked the lowest in terms of accuracy. It had the lowest rate of proper translation and a moderate rate of improper translation. Moreover, it had a significantly higher percentage of untranslated terms, making it less efficient in medical contexts. Ultimately, the choice between *Faraazin* and *Yandex Translate* is contingent upon specific needs and priorities. If accurate translation is prioritized and a higher rate of untranslated terms is acceptable, *Yandex Translate* may be deemed superior. Conversely, if the

goal is to minimize mistranslation and a higher rate of untranslated terms is acceptable, *Faraazin* could be considered the preferable option. The findings are in line with those of Zappatore and Ruggieri (2024), Randhawa et al. (2013), Aiken and Balan (2011), and Kaliyadan and Pillai (2010) asserting that when used in the healthcare sector, MT should be exclusively used for translating non-sensitive texts. Furthermore, the results showed that there were significant differences in the precision and effectiveness rates of the MT software in translating medical terms. The tools showed notably higher percentages of 'proper' translation compared to the categories of ‘improper’ and ‘not translated’ items. This finding suggests that while online MT software is capable of translating the medical terms, their accuracy can vary significantly depending on the tool used. This result is in agreement with that reported by Deilen et al. (2024) and Zhao et al. (2021), concluding that although MT translations seemed easier to read, they contained a considerable rate of complex structures than the human translations with various types of mistakes.With regard to the reliability of the software, a high level of inter-rater reliability was observed, indicating that the assessed translation tools were in substantial agreement in their *proper* translation of medical terms. The high levels of reliability suggest that online MT software can provide a consistent and reliable translation of some medical terminologies despite their varying degrees of accuracy, efficiency, and precision. However, the results revealed that the ‘improper’ and ‘untranslated’ categories exhibited statistically significant disparities across the translation tools examined, which are in line with the studies by Xie et al. (2021), Weng et al. (2019), Daniele (2019), and Patil and Davies (2014). They found that *Google Translate* is a practical tool for translating medical content; however, its accuracy may vary depending on the type of content being translated. Furthermore, Costa-jussà et al. (2024) concluded that MT tools are not still efficient enough to be applied in the medical field where a translation with the highest rate of adequacy is required. Therefore, it is important to exercise caution when using online translation tools and to consider their limitations. Additionally, human translators should still be relied on for critical and high-stakes medical communication. Nonetheless, the results of this study are not consistent with those of Honari (2018) and Zhu et al. (2021), who claimed that MT systems are not proper when it comes to translating scientific or technical expressions from English to Persian. Regarding the use of MT tools in clinical settings, Mehandru et al. (2022) recommended several strategies for designing appropriate MT systems, such as accounting for dialect differences, combining neural MT with human-translated texts, encouraging feedback systems, considering the patients' literacy levels, and setting collaborative suggestions for practitioners. It shows that using MT tools could be effective, particularly in those with a higher literacy level. Also, despite the fluctuating levels of correctness, Birkenbeuel et al. (2021) found that *Google Translate* can be conveniently employed without any financial expenses, providing almost instantaneous interpreting services. Owing to its attributes, it has the capacity to consistently aid in effective one-way verbal communication between medical professionals and patients who have a restricted comprehension of English. All in all, the findings point to the fact that professionally trained human medical translators still remain the most reliable criterion for translation in clinical practices.

## 6. Conclusion

Medical professionals, researchers, and students increasingly rely on MT tools for communication and collaboration across linguistic boundaries while the accuracy and reliability of their translation output are still a matter of investigation. Translating medical terminology using online MT tools can be difficult due to the technical and specialized nature of the terms and some potential limitations of the tools. This study compared the extent to which online MT software can accurately translate English medical terms into Persian and the findings illustrated that while translating the medical terms, healthcare professionals, researchers, and medical students should be aware of these limitations when using the MT tools. Pairwise comparisons also revealed that although these tools seem the most viable options for English-Persian translation of medical terms, they have variations in accuracy, efficiency, and precision and they showed considerable differences while dealing with ‘the items they could not translate’. Thus, it is crucial to exercise caution when using online MT tools for medical translations, particularly in critical and high-stakes medical communication, where the accuracy and reliability of the translation output are critical. Human translators should still be the preferred option in such situations. Like all other studies, this research had some limitations that should be taken into account while generalizing the results. In this research, the medical terms were chosen by three medical experts; thus, it might affect the nature of the terms selected. Also, the reference book in this study was ‘The Dorland’s Medical Dictionary’, since no other more reliable source was available. Finally, this study only focused on Google Translate, Aryanpour, Faraazin, and Yandex

Translate; thus, the outcomes should not be generalized to other translation software.The research highlights both positive and negative aspects of using online MT software to translate medical terms. While such tools are helpful for medical professionals, researchers, and students in overcoming language barriers, they cause challenges when it comes to translating specialized and technical medical terminology. Medical professionals can use the findings of the study to make informed decisions about using online MT tools when communicating or collaborating with others. Moreover, researchers in the field of medical translation can use the results as a reference to better understand the reliability and efficiency of online MT tools for medical translations. Furthermore, medical students who may use online translation tools, while translating medical terminology and concepts, should take their strengths and limitations into account. Finally, developers and designers of online MT tools may use the insights gained from this research to detect errors and improve the algorithms of the MT tools. Future research may examine the accuracy and precision of online MT tools in translating a more extensive range of medical terms. Additional research could assess the impact of context on online MT output quality, including the styles and lengths of medical texts or medical specialties (e.g., radiology, surgery, and cardiology).

Moreover, future studies should explore the impact of cultural variations and idiomatic expressions on the accuracy and reliability of online MT output. Overall, this study may contribute to the already existing findings on evaluating the performance of online MT software in the medical field and provide practical implications to enhance the quality of medical translations.

**Authors Contribution**
All authors have contributed equally to prepare the paper.

**Availability of data and materials**
The data that support the findings of this study are available from the corresponding author, upon reasonable request.

**Conflict of interests**
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] Aiken, M., & Balan, S. (2011). An analysis of Google Translate accuracy. *Translation J.*, 16(2). Available from: www.bokorlang.com/journal/56google.htm.

[2] Archer, J. (2002). Internationalisation, technology and translation. Perspectives: *Studies in Translatology*, 10(2), 87–117. https://doi.org/10.1080/0907676X.2002.9961437.

[3] Arnold, D. L., Balkan, S., Meijer, R.L., Humphreys, L., & Sadler, L. (1994). Machine translation: An introductory guide. *NCC Blackwell*.

[4] Aryanpour (2023). English to Persian Dictionary: Aryanpour English to Farsi Translator. *Retrieved February* 20, 2023, from https://aryanpour.com

[5] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *In Proceedings* of the 3rd International Conference on Learning Representations (May 7–9), San Diego, USA.

[6] Betancourt, J. R., & Jacobs, E. A. (2000). Language barriers to informed consent and confidentiality: the impact on women's health. *Journal of American Medical Women's Association*, 55(5), 294-295. https://www.researchgate.net/publication/12256281.

[7] Birkenbeuel, J., Joyce, H. J., Sahyouni, R., Cheung, D., Maducdoc, M., Mostaghni, N., … Lin, H. W. (2021). Google translate in healthcare: preliminary evaluation of transcription, translation, and speech synthesis accuracy. *BMJ Innovations*, 7(2), 422–429. https://doi.org/10.1136/bmjinnov-2019-000347.

[8] Brown, P., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Mercer, R., & Roossin, P. (1990). A statistical approach to language translation. *Computational Linguistics*, 16(2), 71-76. http://dx.doi.org/10.3115/991635.991651.

[9] Buzarna-Tihenea, G.A. (2015a). Several aspects regarding the specialized translation of medical texts. From theory to practice. *In Ovidius University Annals, Economic Sciences Series (pp. 17-21). Ovidius University Press.*

[10] Buzarna-Tihenea, G. A. (2015b). Procedures and difficulties in the translation of medical texts. An English-Romanian case study. In Discourse as a Form of Multiculturalism in Literature and Communication, Section*: Language and Discourse* (pp. 164-175). Arhipelag XXI Press.

[11] Chen, X., Acosta, S., & Barry, A. E. (2016). Evaluating the accuracy of Google Translate for diabetes education material. JMIR *Diabetes*, 1(1), e3. https://doi.org/10.2196/diabetes.5848.

[12] Chiang, D. (2007). Hierarchical phrase-based translation. *Computational Linguistics*, 33(2), 201-228. https://doi.org/10.1162/coli.2007.33.2.201.

[13] Church, K.W., & Mercer, R.L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, 19(1), 1-24. MIT Press. https://aclanthology.org/J93-1001/

[14] Merschel, L. M., Reisinger, D., & Clifford, J. (2013). Meeting the challenges of machine translation. *The Language Educator*, 8, 44–47. https://www.researchgate.net/publication/362177070

[15] Costa-jussà, M.R., Farrús, M., & Serrano Pons, J. (2012). Machine translation in medicine: A quality analysis of statistical machine translation in the medical domain. *Advanced Research in Scientific Areas*, 3(7), 1995-1998. https://www.upf.edu/documents/2981382/2985826/

[16] Costa-jussà, M.R., Andrews, P., Basta, Ch., Ciro, J., Falenska, A., Goldfarb-Tarrant, S., … Sánchez, E. (2024). Overview of the shared task on machine translation gender bias evaluation with multilingual holistic bias. In Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP) (pp. 399–404), Bangkok, Thailand. *Association for Computational Linguistics*.

[17] Daniele, F. (2019). Performance of an automatic translator in translating medical abstracts. *Heliyon*, 5(10). https://doi.org/10.1016/j.heliyon.2019.e02687

[18] Deilen, S., Lapshinova-Koltunski, E., Hernández Garrido, S., Maaß, Ch., Hörner, J., Theel, V., & Ziemer, S. (2024). Towards AI-supported health communication in plain language: Evaluating intralingual machine translation of medical texts. *In Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024 (pages 44–53), Torino, Italia. ELRA and ICCL.*

[19] Dew, K. N., Turner, A. M., Choi, Y. K., Bosold, A., & Kirchhoff, K. (2018). Development of machine translation technology for assisting health communication: A systematic review. *Journal of Biomedical Informatics*, 85, 56-67. https://doi.org/10.1016/j.jbi.2018.07.018.

[20] Dong, D., Wu, H., He, W., Yu, D., & Wang, H. (2015). Multi-task learning for multiple language translation. *In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Jul 26–31), Beijing, China.*

[21] Faraazin. (2023). Online translation from English to Persian and vice versa. *Retrieved February* 23, 2023, from https://www.faraazin.ir

[22] Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y.N.(2017). *Convolutional sequence to sequence learning. In Proceedings of the 34th International Conference on Machine Learning*; 2017 Aug 6–11; Sydney, NSW, Australia. https://doi.org/10.48550/arXiv.1705.03122

[23] Gharsa, M. A. (2015). The problem of translating medical terms from English into Arabic. Unpublished doctoral dissertation, Durham University. *Durham, United Kingdom. Retrieved May* 2024, from https://etheses.dur.ac.uk/11166.html.

[24] Hellmich, E. A. (2019). Language teacher beliefs about technology: expanding the ecology. *International Journal of Computer-Assisted Language Learning and Teaching*, 9, 1–17. https://doi.org/10.4018/IJCALLT.2019100101

[25] Honari, M. (2018). The problems of machine translation in English-Persian and Persian-English translation of technical and scientific texts based on Waddington's Method A. *Unpublished master's thesis, Islamic Azad University of Bandar Abbas, Iran*.

[26] Hutchins, W.J. (2000). The IAMT certification initiative and defining translation system categories. In Proceedings of 5th EAMT Workshop: Harvesting Existing Resources, Ljubljana, Slovenia. *European Association for Machine Translation*. https://aclanthology.org/2000.eamt-1.8/

[27] Kaliyadan, F., & Pillai, S. G. (2010). The use of Google language tools as an interpretation aid in cross-cultural doctor-patient interaction: A pilot study. *Informatics in Primary Care*, 18(2), 141-143. http://dx.doi.org/10.14236/jhi.v18i2.764.

[28] Karwacka, W. (2015). Medical translation. In Ł. Bogucki, S. Goźdź-Roszkowski, P. Stalmaszczyk (Eds.), *Ways to translation (pp 271-298). Wydawnictwo Uniwersytetu Łódzkiego.*

[29] Khadem Olhosseiny, Z. S., & Askari, M.I. (2024). CAT tools in translation proficiency development: EFL students' perceptions in focus. *JNTELL*, 3(2), 35-68. DOI: 10.71528/2024.24011091

[30] Khanna, R. R., Karliner, L. S., Eck, M., Vittinghoff, E., Koenig, C. J., & Fang, M. C. (2011). Performance of an online translation tool when applied to patient educational material. *Journal of Hospital Medicine*, 6(9), 519-25. https://doi.org/10.1002/jhm.898.

[31] Khoong, E. C., Steinbrook, E., Brown, C., & Fernandez, A. (2019). Assessing the use of Google Translate for Spanish and Chinese translations of emergency department discharge instructions. *JAMA Internal Medicine*, 179(4), 580–582. https://doi.org/10.1001/jamainternmed.2018.7653.

[32] Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. *In Proceedings of the First Workshop on Neural Machine Translation, 2017 Aug 4, Vancouver, BC, Canada.*

[33] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., … Herbst, E. (2007). Moses: open-source toolkit for statistical machine translation. *In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (Jun 25–27), Prague, Czech Republic.*

[34] Kroulek, A. (2016). 11 Google Translate facts you should know. *Retrieved April* 8, 2023, from https://www.k-international.com/blog/google-translate-facts/.

[35] McGuire, N. (2018). How accurate is Google Translate in 2018? Argo Translation. *Retrieved April* 8, 2024, from https://www.argotrans.com/blog/accurate-google-translate-2018/.

[36] Mehandru, N., Robertson, S., & Salehi, N. (2022). Reliable and safe use of machine translation in medical settings. ACM Conference on Fairness, Accountability, and Transparency, 2022. *Retrieved from* https://doi.org/10.1145/3531146.3533244.

[37] Moberly, T. (2018). Doctors are cautioned against using Google Translate in consultations. *British Medical Journal*, 363. https://doi.org/10.1136/bmj.k4546.

[38] MOGI's Translation Team. (2011). Technical translation. *Retrieved April* 2023, from https://www.moi.translations.com/technical-documentation-Translation.html.

[39] Montero-Martínez, S., & de Quesada, M. G. (2003). *Terminological analysis for translation. Perspectives*, 11(4), 293–314. https://doi.org/10.1080/0907676X.2003.9961482

[40] Munassar Awadh, A.N. (2024). Challenges and strategies of translating scientific texts: A comparative study of human translation and artificial intelligence. *Educational Administration: Theory and Practice*, 30(4), 9898–9909. https://doi.org/10.53555/kuey.v30i5.4147

[41] Nădrag, L. (2016). Issues raised by the English-Romanian translation of cardiovascular terminology. In I. Boldea & D.M. Buda (Eds.), Convergent discourses: *Exploring the contexts of communication* (pp. 272-278). Arhipelag XXI Press.

[42] Nagao, M. (1984). A framework for a mechanical translation between Japanese and English by analogy principle. In A. Elithorn & R. Banerji (eds.), *Artificial and human intelligence* (pp.173-180).

Elsevier North-Holland, Inc.

[43] Newman Dorland, W. (1890). *Dorland's Medical Dictionary.* [Translated by H. Namavar & M. Zarabi, 2020, 32nd edition]. http://elsevierhealth.com/product.jsp?isbn=9781455709854&dmnum=12449

[44] National Research Council. (1966). Language and machines: Computers in translation and linguistics. *The National Academies Press*. https://doi.org/10.17226/9547.

[45] Nord, C. (1997). Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis. *Rodopi*.

[46] Och, F. J., & Ney H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1), 19-51. http://dx.doi.org/10.1162/089120103321337421.

[47] Pekcoşkun Güner, S. (2020). Translation of medical texts: challenges and solutions. *The Journal of Social Sciences*, 340-351. http://dx.doi.org/10.29228/SOBIDER.43857

[48] Patil, S., & Davies, P. (2014). Use of Google Translate in medical communication: evaluation of accuracy. *British Medical Journal*, 349. DOI: https://doi.org/10.1136/bmj.g7392

[49] Rahmanian, A. (2018). The problems with medical translation terminology based on Reiss's Model. *Unpublished master's thesis, Islamic Azad University of Bandar Abbas, Iran*.

[50] Randhawa, G., Ferreyra, M., Ahmed, R., Ezzat, O., & Pottie, K. (2013). Using machine translation in clinical practice. *Canadian Family Physician*, *59(4), 382-383. PMID: 23585608; PMCID: PMC3625087*.

[51] Sager, J.C. (1994). Language engineering and translation: Consequences of automation. *John Benjamins Publishing Company*. https://doi.org/10.1075/target.8.1.17sch

[52] Sedaghat Rostami, M., & Bigdeli, A. (2016). Translation of social sciences texts: problems and solutions. *Journal of Iranian Social Studies*, 10(2.3), 57-76. http://www.jss-isa.ir/article_31590_en.html

[53] Somers, H.L. (2003). Introduction. In H.L. Somers (Eds.), Computers and translation: *A translator's guide* (pp. 1-11). John Benjamins.

[54] Spechbach, H., Mallem, I. S. H., Gerlach, J., Tsourakis, N., & Bouillon, P. (2017). Comparison of the quality of two speech translators in emergency settings: A case study with standardized Arabic-speaking patients with abdominal pain. *In Proceedings of European Congress of Emergency Medicine (EUSEM 2017). Athens, Greece.*

[55] Sutskever, I., Vinyals, O., & Le, Q.V. (2014). Sequence-to-sequence learning with neural networks. In Proceedings of the 27th *International Conference on Neural Information Processing Systems* (8–13), Montreal, QC, Canada.

[56] Taleghani, M., & Pazouki, E. (2018). Free online translators: A comparative assessment in terms of idioms and phrasal verbs. *International Journal of English Language & Translation Studies*, 6(1), 15-19. http://www.eltsjournal.org

[57] Turner, A. M., Mandel, H., & Capurro, D. (2013). Local health department translation processes: potential of machine translation technologies to help meet needs. *In AMIA annual symposium proceedings, 2013, American Medical Informatics Association.*

[58] Turner, K. (2016). Google Translate is getting really, really accurate. Washington Post, October 3. *Retrieved May* 8, 2024, from https://www.washingtonpost.com/news/innovations/wp/2016/10/03/google-translate-is-getting-really-really-accurate.

[59] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. … Polosukhinet, I. (2017). Attention is all you need. *In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017 Dec 4–9, Long Beach, CA, USA.*

[60] Vieira, L. N., O'Hagan, M., & O'Sullivan, C. (2020). Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society, 11, 1515-1532*. https://doi.org/10.1080/1369118X.2020.1776370.

[61] Wang, H. (2006). Multi-strategy machine translation. In Y.Q. Cao & M.S. Sun (Eds.), *Frontiers of Chinese information processing* (pp. 45-52). Tsinghua University Press.

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

[63] Weng, W.H., Chung, Y. A., & Szolovits, P. (2019). Unsupervised clinical language translation. *In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '19, Association for Computing Machinery (pp. 3121-3131). New York, NY, USA.* 10.1145/3292500.3330710

[64] Williams, J., & Chesterman, A. (2002). The map: A beginner's guide to doing research in translation studies. *Saint Jerome Publication.*

[65] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., … Dean, J. (2016). Google's neural machine translation system: bridging the gap between human and machine translation. https://doi.org/10.48550/arXiv.1810.08398.

[66] Xie, W., Ji, M., Zhao, M., Zhou, T., Yang, F., Qian, X., Chow, C.-Y., Lam, K.-Y., & Hao, T. (2021). Detecting symptom errors in neural machine translation of patient health information on depressive disorders: Developing interpretable Bayesian machine learning classifiers. *Front. Psychiatry,* 12 (2021). https://doi.org/10.3389/fpsyt.2021.771562

[67] Yandex. (2017, September 14). One model is better than two. Yandex Translate launches a hybrid machine translation system. *Retrieved February* 23, 2024, from https://yandex.com/company/blog/one-model-is-better-than-two-yu-yandex-translate-launches-a-hybrid-machine-translation-system/

[68] Yandex. (2023). About Yandex Translate. *Retrieved February* 20, 2024, from https://yandex.com/support/translate/

[69] Yassi, M. (2018). A study of equivalence in English-Persian translation of surgical medical terms based on Newmark's Model. *Unpublished master's thesis, Islamic Azad University of Marvdasht Branch, Iran.*

[70] Zappatore, M., & Ruggieri, G. (2024). Adopting machine translation in the healthcare sector: A methodological multi-criteria review. *Computer Speech and Language*, 84, 101582.

https://doi.org/10.1016/j.csl.2023.101582

[71] Zhang, J.J., & Zong, C.Q. (2020). Neural machine translation: challenges, progress, and future. *Science China Technological Sciences*, 63(10), 2028-2050. https://doi.org/10.48550/arXiv.2004.05809

[72] Zhao, L., Gao, W., & Fang, J. (2021). High-performance English–Chinese machine translation based on GPU-enabled deep neural networks with domain corpus. *Appl. Sci*., 11 (22), 10915. https://doi.org/10.3390/app112210915

[73] Zhu, Y., Sha, Y., Wu, H., Li, M., Hoffman, R.A., & Wang, M.D. (2022). Proposing causal sequence of death by neural machine translation in public health informatics. *IEEE J. Biomed. Health Inf.,* 26 (2022), 1422-1431. DOI: 10.1109/JBHI.2022.3163013