

Comparative Quality Evaluation of the Output of Free Online Translation Machines between Arabic and Persian Based on the DQF-MQM Model

Shahryar Niazi 

Associate Professor, Department of Arabic
Language and Literature, University of Tehran,
Tehran, Iran

Mahmood Bijankhan 

Professor, Department of Linguistics, University
of Tehran, Tehran, Iran

Mazyar Pashaei* 

PhD Student in Arabic Translation Studies,
University of Tehran, Tehran, Iran

Abstract

Man's need to translate with more efficiency has made him endeavor to achieve advanced translation technologies. Most of the efforts in this field have been devoted to achieving machine (automatic) translation (without human intervention), which, although it does not have the quality of human translation, has other advantages such as speed and high availability and low cost. The peak of these benefits can be seen in free online translation machines. Some of these machines (i.e. Google, Bing, Yandex, Reverso, ModernMT, and NiuTrans) support Arabic to Persian translation and vice versa. The purpose of this research is to compare the quality of Arabic<>Persian translations provided by these machines with each other. In order to achieve this goal, first, two small Arabic and Persian corpuses, each

* Corresponding Author: mazyar.pashaei@ut.ac.ir

How to Cite: Niazi, Sh., Bijankhan, M., Pashaei, M. (2024). Comparative Quality Evaluation of Free Online Translation Machines between Arabic and Persian based on the DQF-MQM Model. *Translation Researches in the Arabic Language and Literature*, 14 (30), 63-90. doi: [10.22054/rctall.2024.78817.1724](https://doi.org/10.22054/rctall.2024.78817.1724)

containing 60 sentences with random types and topics, were selected from the sentences in the two Arabic and Persian frequency dictionaries published by Routledge, then these sentences were entered one by one into the aforementioned translation machines. and the received output was scrutinized by human evaluation method based on the DQF-MQM error classification and analysis model. The translation machines in order from highest to lowest output quality are: Google, Bing, Yandex, ModernMT, Reverso, and NiuTrans. This is not an absolute and constant result, but a statistical and probabilistic one; lower-ranked machines translate some sentences better than the higher-ranked machines.

Keywords: Translation Studies, Translation Technology, Machine Translation Evaluation, Google Translate, Bing Translator, Yandex Translate, Reverso, ModernMT, NiuTrans

Introduction

Translation technology has been an important branch of translation studies. In 1972, at the third conference of applied linguistics, James Holmes introduced the field of translation technology as a sub-branch of the "applied" branch of the emerging interdisciplinary science of "translation studies". He divided this field into three categories: theories of translation by humans, by machines, and by both (Machine-Aided Human Translation or Human-Aided Machine Translation).

Most of the efforts in the field of translation technology have been focused on making the machine able to translate without human intervention. This type of translation is called "automatic translation". Machine translation will not be able to beat professional human translation in the field of quality, but it has other advantages such as high speed, low cost and easy access.

The pinnacle of convenient access and low cost for translation services can be seen in free online translation machines. They can be accessed

and used for free through any system with a browser and connection to the Internet; Some also have a specific smartphone application that provides additional features such as offline translation.

The evaluation of the phenomenon of machine translation generally includes many topics; Different aspects of it can be examined in different ways in response to the different needs of the people involved (including: end user, developer, and investor). Our focus in this article is on evaluating the quality of the output or product of the translation machines. The questions of the research are:

- 1- Which free online translation machine do produce Arabic to Persian translation with better quality?
- 2- Which free online translation machine do produce Persian to Arabic translation with better quality?

A brief and widely used definition of “translated text quality” is as follows: “A quality translation demonstrates accuracy and fluency required for the audience and purpose and complies with all other specifications negotiated between the requester and provider, taking into account end-user needs”

Literature Review

Several scientific studies have dealt with the subject of comparative evaluation of machine translation for Arabic-English or Persian-English language pairs, but no research in this field has been published for Arabic-Persian language pairs. These researches have generally selected a test suite first, then translated it by several translation machines and studied the output using one or more special methods of machine translation evaluation. Here we present the summary of most recent researches. Ben Milad (2022), Almahasees (2020) and Al-Shalabi (2017) tested several machine translations between Arabic and English with different methods and all concluded that Google Translate produces better quality translations, just Abu-Ayyash (2017) concluded that Google Translate and Bing Translator produce similar quality outputs.

Research Methodology

There are various methods to evaluate machine translation quality. They are divided into two main subcategories of human and automatic evaluation. In this research we use a standard and up to date human evaluation model called DQF-MQM, and especially a subset of it that is appropriate for machine translation quality evaluation and is as follows: Four high-level error types of Accuracy, Fluency, Locale Convention, and Terminology. Accuracy type is further subdivided into four granular error types of Addition, Omission, Mistranslation, and Untranslated. Fluency type is further subdivided into three granular types of: Grammatical, Grammatical Register and Spelling. We used the excel template on the formal website of DQF-MQM to evaluate 60 Arabic-Persian and 60 Persian-Arabic translated sentences done by 6 online free translation machines that support Arabic<>Persian translation. The sentences were selected from two Arabic and Persian Frequency dictionaries and had random modes, genres and subjects.

Conclusion

In this research, the quality of Arabic<>Persian translation of six free online machine translation systems was evaluated using the DQF-MQM model. The ranks for Arabic to Persian translation quality are as follows: 1- Google 2- Bing 3- Yandex 4- ModernMT 5- Reverso 6- NiuTrans. The ranks for Persian to Arabic translation quality are as follows: 1- Google 2- Bing 3- Yandex and ModernMT 4- Reverso 5- NiuTrans. They can also be placed in three categories, that the machines in each category have similar qualities: 1- Google and Bing 2- Yandex and ModernMT 3- Reverso and NiuTrans. It should be noted that these rankings do not mean the absolute superiority of one of these translation machines over another. This possible superiority may also change with the passage of time, and in the future another translation machine would beat Google and Bing.



----- پژوهش‌های ترجمه در زبان و ادبیات عربی -----

دوره ۱۴، شماره ۳۰، بهار و تابستان ۱۴۰۳، ۶۳-۹۰

rctall.atu.ac.ir

DOI: [10.22054/rctall.2024.78817.1724](https://doi.org/10.22054/rctall.2024.78817.1724)

ارزیابی تطبیقی کیفیت خروجی ماشین‌های ترجمه برخط رایگان بین عربی و فارسی بر اساس مدل DQF-MQM

شهریار نیازی ^{ID} | دانشیار، گروه زبان و ادبیات عربی، دانشگاه تهران، تهران، ایران

محمود بی‌جن‌خان ^{ID} | استاد، گروه زبان‌شناسی، دانشگاه تهران، تهران، ایران

مازیار پاشائی* ^{ID} | دانشجوی دکتری مطالعات ترجمه عربی، دانشگاه تهران، تهران، ایران

چکیده

نیاز بشر به اجرای فرآیند ترجمه با بازدهی هر چه بیشتر موجب تلاش وی برای دست یافتن به فناوری‌های پیشرفته ترجمه بوده است. بخش اعظم تلاش‌ها در این میدان صرف رسیدن به ترجمه ماشینی یا خودکار (بدون دخالت انسان) شده است که کیفیت ترجمه انسانی را ندارد، اما دارای مزیت‌های دیگری مانند سرعت و دسترسی بالا و هزینه پایین است. اوج این مزایا را می‌توان در ماشین‌های ترجمه برخط رایگان دید. بعضی از این ماشین‌ها (یعنی گوگل، بینگ، یاندکس، رورسو، مادرن‌ام‌تی و نیوترنس) از ترجمه عربی به فارسی و برعکس پشتیبانی می‌کنند. هدف این پژوهش مقایسه کیفیت خروجی این ماشین‌های ترجمه با همدیگر و یافتن بهترین گزینه برای ترجمه خودکار بین زبان‌های عربی و فارسی است. برای رسیدن به این هدف، ابتدا دو پیکره کوچک عربی و فارسی هر کدام شامل ۶۰ جمله با انواع و موضوع‌های تصادفی از جملات موجود در دو کتاب فرهنگ بسامدی عربی و فارسی انتشارات راتلج انتخاب شد، سپس این جملات تک به تک در ماشین‌های ترجمه اشاره شده وارد شد و خروجی دریافت شده با روش ارزیابی انسانی بر اساس مدل تحلیل و طبقه‌بندی خطای DQF-MQM مورد بررسی قرار گرفت. ماشین‌های ترجمه به ترتیب از بیشترین به کمترین کیفیت خروجی از این قرار بودند: گوگل، بینگ، یاندکس، مادرن‌ام‌تی، رورسو، و نیوترنس. این نتیجه مطلق و همیشگی نیست، بلکه آماری و احتمالاتی است؛ ماشین‌های با رتبه پایین تر بعضی جملات را بهتر از ماشین‌های با رتبه بالاتر ترجمه می‌کنند.

کلیدواژه‌ها: مطالعات ترجمه، فناوری ترجمه، ارزیابی ترجمه ماشینی، گوگل ترنزلت، بینگ ترنزلتور، یاندکس ترنزلت، رورسو، مادرن‌ام‌تی، نیوترنس.

* نویسنده مسئول: mazyar.pashaei@ut.ac.ir

۱. مقدمه

فناوری ترجمه شاخه‌ای مهم از حوزه مطالعات ترجمه بوده است. جیمز هلمز^۱ در سال ۱۹۷۲ در سومین کنفرانس زبان‌شناسی کاربردی، حوزه فناوری ترجمه را به عنوان زیرشاخه‌ای از شاخه «کاربردی» از علم میان‌رشته‌ای نوظهور «مطالعات ترجمه» معرفی کرد (Malmkjær, 2013: 32-31). وی نظریه‌های ترجمه را با توجه به فاعل آن به سه گروه تقسیم می‌کند: نظریه‌های ترجمه توسط انسان، توسط ماشین و توسط هر دو (ترجمه انسان به کمک ماشین یا ترجمه ماشین به کمک انسان) (Holmes, 1988: 74).

با گذشت زمان و پیشرفت فناوری جایگاه فناوری ترجمه در مطالعات ترجمه اهمیت بیشتری یافت. ماری اسنل-هورنبی^۲ فرآیند جهانی‌سازی و پیشرفت چشمگیر فناوری اطلاعات و ارتباطات را عواملی خارجی می‌داند که صنعت‌های زبانی [مانند صنعت ترجمه] را همچون سایر جوانب زندگی مدرن دچار تحولات عمیقی کرد. این دگرگونی که در دهه ۱۹۹۰ در حوزه مطالعات ترجمه رخ داد «چرخش جهانی‌سازی»^۳ خوانده شده است و می‌تواند به اهمیت چرخش قبلی مطالعات ترجمه که در دهه ۱۹۸۰ رخ داد؛ یعنی «چرخش فرهنگی»^۴ باشد (Snell-Hornby, 2010: 367-368).

بخش عظیمی از تلاش‌ها در حوزه فناوری ترجمه بر این متمرکز بوده است که ماشین بتواند بدون دخالت انسان عمل ترجمه را انجام دهد. این نوع از ترجمه، ترجمه ماشینی نام گرفته است. ترجمه ماشینی نخواهد توانست ترجمه انسانی حرفه‌ای را در میدان کیفیت شکست دهد، اما از مزیت‌هایی دیگر مانند سرعت بالا، هزینه پایین و دسترسی راحت برخوردار است. متقاضیان همیشه به ترجمه با کیفیت بالای مناسب برای انتشار نیاز ندارند، بلکه گاهی نیاز دارند به درکی کلی از یک متن با زبان بیگانه برسند و ترجیح می‌دهند این کار با سرعت زیاد و هزینه کم انجام شود؛ هر چند کیفیت آن پایین باشد. همچنین گاهی نیاز است که متون با زبان بیگانه که به سرعت و با حجم بالا در فضای مجازی در حال تولید شدن هستند، ترجمه شوند. این نوع از تقاضا در بازار ترجمه نیز تنها با مزیت عرضه

1. Holmes, J.
2. Snell-Hornby, M.
3. Globalization Turn
4. Cultural Turn

شده توسط ماشین قابل پاسخگویی است. حتی در صورتی که نیاز به متن ترجمه شده با کیفیت بالا داشته باشیم، باز هم ماشین بلااستفاده نیست و می‌توان از آن برای تهیه سریع و راحت نسخه اولیه ترجمه استفاده کرد و آن را برای ویرایش به مترجمین حرفه‌ای سپرد تا آن را به نسخه ترجمه با کیفیت مبدل کنند. بنابراین در شرایط ویژه عصر اطلاعات، ترجمه ماشینی در کنار ترجمه انسانی نقشی مهم و حیاتی ایفا می‌کند (Hutchins, 2003: 5-24 & 7).

اوج دسترسی راحت و هزینه پایین برای خدمت ترجمه را می‌توان در ماشین‌های ترجمه برخط رایگان دید. آن‌ها از طریق هر سامانه دارای مرورگر و متصل به اینترنت به صورت رایگان قابل دسترسی و استفاده هستند؛ برخی نیز دارای برنامه کاربردی خاص تلفن هوشمند هستند که قابلیت‌های اضافه همچون ترجمه سیار برون‌خط (بدون اتصال به اینترنت) را فراهم می‌کنند. بعضی از این سامانه‌ها دارای امکانات مهم دیگری نیز هستند، از جمله: ترجمه صوت، تصویر و دستخط، و پخش صوتی متن مبدأ و مقصد ترجمه. بعضی از سامانه‌های ترجمه ماشینی برخط رایگان از زبان‌های عربی و فارسی پشتیبانی می‌کنند و می‌توانند متون را از عربی به فارسی و برعکس ترجمه کنند (رجوع شود به جدول (۱)).

ارزیابی پدیده ترجمه ماشینی به طور عام شامل مباحث بسیاری می‌شود؛ می‌توان با روش‌های مختلفی جنبه‌های مختلفی از آن را در پاسخ به نیازهای مختلف افراد دخیل (از جمله: کاربر نهایی، توسعه‌دهنده و سرمایه‌گذار) مورد بررسی قرار داد (White, 2003: 222).

تمرکز ما در این مقاله بر ارزیابی خروجی یا محصول ترجمه ماشینی خواهد بود. از آنجا که هدف از ماشین ترجمه، تولید همین محصول است، ارزیابی آن مورد توجه خاص تمامی افراد دخیل است و به عنوان یک معیار اصلی ارزیابی ترجمه ماشینی به طور عام شناخته می‌شود (Kit & Wong, 2023: 225).

این پژوهش قصد دارد کیفیت محصول ترجمه سامانه‌های ترجمه ماشینی برخط رایگان از عربی به فارسی و برعکس را مورد ارزیابی تطبیقی قرار دهد. سؤالات اصلی پژوهش عبارتند از:

- محصول ترجمه عربی به فارسی کدام ماشین ترجمه برخط رایگان کیفیت بهتری دارد؟

- محصول ترجمه فارسی به عربی کدام ماشین ترجمه برخط رایگان کیفیت بهتری دارد؟

۲. مبانی نظری

۱-۲. تعریف کیفیت محصول ترجمه ماشینی

یک تعریف موجز کیفیت متن ترجمه که جامع نظریات ترجمه از جمله اسکوپوس است و به طور گسترده در زمینه ترجمه ماشینی نیز استفاده می‌شود از این قرار است: «صحت^۱ و سلاست^۲ به مقتضای مخاطب و هدف و مطابقت با سایر مشخصات توافق شده بین درخواست‌کننده و ارائه‌دهنده خدمت ترجمه با در نظر گرفتن نیازهای کاربر نهایی». بنابراین کیفیت ترجمه سه معیار اصلی دارد: ۱- سلاست، فهم‌پذیری^۳ یا پذیرفتگی^۴ متن به زبان مقصد، ۲- صحت، بسندگی^۵ یا وفاداری^۶ به معنای تعادل معناشناسی و کاربردشناسی میان متن مبدأ و مقصد و ۳- مطابقت با مشخصات خاص درخواست‌دهنده؛ در صورت وجود (Chatzikoumi, 2020: 2).

۲-۲. روش‌های ارزیابی کیفیت ترجمه ماشینی

ارزیابی کیفیت ترجمه ماشینی معمولاً طی دو مرحله اساسی انجام می‌شود؛ ابتدا واحدهایی زبانی از زبان مبدأ انتخاب شده (مجموعه آزمایش^۷) و وارد ترجمه ماشینی می‌شوند، سپس خروجی سامانه به زبان مقصد دریافت شده و مورد بررسی و قضاوت قرار می‌گیرد. عده‌ای از این فرآیند می‌تواند به صورت دستی و مستقیماً توسط انسان انجام شود، یا به صورت اتوماتیک توسط ماشین انجام شود. بر این اساس ارزیابی ترجمه به دو نوع انسانی و خودکار (یا ماشینی) تقسیم می‌شود (Chatzikoumi, 2020: 3 and Kit & Wong, 2023: 227).

1. Accuracy
2. Fluency
3. Intelligibility
4. Acceptability
5. Adequacy
6. Fidelity

۷. تعریف اصطلاح «مجموعه آزمایش» (Test Suite/Set): مجموعه‌ای از متون، جمله‌ها یا عبارات که به یک سامانه ترجمه ماشینی وارد می‌شود تا خروجی یا محصول آن مورد ارزیابی قرار بگیرد. (Rothwell et al., 2023: xxi)

ارزیابی‌های خودکار و انسانی خود به انواع روش‌های دیگری تقسیم می‌شوند. ارزیابی ماشینی معمولاً به سه نوع تقسیم می‌شود:

۱- مبتنی بر ترجمه مرجع^۱: در این روش، میزان شباهت یا نزدیکی متون ترجمه شده توسط ماشین با ترجمه‌هایی انجام شده توسط مترجم‌های انسانی حرفه‌ای که ترجمه مرجع خوانده می‌شوند، سنجیده می‌شود.

۲- تخمین کیفیت^۲

۳- ارزیابی تشخیصی مبتنی بر نقاط بازرسی^۳

ارزیابی انسانی نیز معمولاً به شش نوع تقسیم می‌شود:

۱- برآورد کیفیت^۴: در این روش انسان به صورت فی‌البداهه نظر خود را درباره یک ترجمه بیان می‌کند؛ با تعیین سطح (برای مثال خوب، بد یا متوسط) یا نمره‌دهی (برای مثال از ۰ تا ۱۰۰).

۲- رتبه‌بندی: چینش چندین ترجمه به ترتیب کیفیت از بهترین تا بدترین

۳- تحلیل و طبقه‌بندی خطا^۵: بر خلاف دو روش قبل که بر قضاوت مستقیم ارزیاب متکی هستند و ممکن است تنها منعکس‌کننده نظرهایی شخصی باشند در این روش ارزیاب سعی می‌کند متن را تحلیل کند و انواع خطای رخ داده در ترجمه را شناسایی و سطح‌بندی کند. سپس نمره منفی تعلق گرفته بر این اساس محاسبه می‌شود؛ بنابراین، قابل اعتمادتر از سطح/نمره تعلق گرفته در روش‌های قبل است.

۴- استخراج اطلاعات^۶

۵- آزمون درک مطلب^۷

۶- پس‌ویرایش^۸ (Chatzikoumi, 2020: 4-9 and Kit & Wong, 2023: 227-237).

1. Reference Translation-based
2. Quality Estimation
3. Diagnostic Evaluation Based on Checkpoints
4. Quality Assessment
5. Error Analysis and categorization
6. Information Extraction
7. Comprehension Test
8. Post-editing

یک عامل بسیار مهم در فرآیند ارزیابی انسانی، قضاوت کنندگان^۱ هستند که به آن‌ها نشان‌گذار^۲ نیز گفته می‌شود که باید دارای ویژگی‌های خاصی باشند تا قضاوتشان قابل اعتماد باشد. بسته به نوع ارزیابی، قضاوت کنندگان می‌توانند تک زبانه یا دوزبانه باشند؛ یعنی افراد بومی یا شبه بومی زبان مقصد یا هر دو زبان مبدأ و مقصد. آموزش فرآیند به قضاوت کنندگان، دستورالعمل ارزشیابی به همراه مثال و همچنین آشنایی قضاوت کننده با حوزه موضوع متن از پیش نیازهای پروژه ارزیابی است. توصیه می‌شود که زبان مقصد زبان مادری قضاوت کننده باشد (Chatzikoumi, 2020: 10).

۲-۳. مدل DQF-MQM برای ارزیابی ترجمه ماشینی

در اینجا یکی از مدل‌های ارزیابی انسانی مبتنی بر تحلیل و طبقه‌بندی خطا به نام DQF-MQM که در این پژوهش از آن استفاده خواهد شد، توضیح داده می‌شود. این طبقه‌بندی تلفیقی است از دو مدل^۳ MQM و^۴ DQF که توسط دو گروه اروپایی متخصص بر اساس تلاش‌های گذشته در حوزه مطالعات و صنعت ترجمه طراحی شده‌اند. این مدل در سال ۲۰۱۴ پیشنهاد شد و در حال تبدیل شدن به یک استاندارد جهانی برای ارزیابی کیفیت ترجمه است (Lommel, 2018: 109- 110).

طبقه‌بندی DQF-MQM دارای هفت نوع خطای سطح بالا^۵ است و برای هر کدام از این نوع خطاها انواعی از خطاهای ریزدانه‌ای^۶ تعریف می‌کند. ذکر تمامی این موارد از حوصله این پژوهش خارج است و در اینجا تنها به بیان مواردی که برای طبقه‌بندی خطای ترجمه ماشینی مناسب هستند، اکتفا می‌شود (لومل و دیگران^۷، ۲۰۱۵: ۶؛ تاوس^۸، بی تا). این موارد از این قرار هستند:

1. Judge

۲. تعریف اصطلاح «نشانه گذاری» (annotation): به فرایند اضافه کردن اطلاعات زبانی به یک پیکره نشان گذاری گفته می‌شود. برای مثال ممکن است نقش‌های دستوری کلمات به پیکره اضافه شود. به فردی که این کار را می‌کند نشانه گذار (annotator) گفته می‌شود. (Sin-Wai, 2004: 12)

3. Multidimensional Quality Metrics

4. Dynamic Quality Framework

5. High-level error type

6. Granular error type

7. Lommel, A., et al.

8. TAUS

۲-۳-۱. خطای صحت

- * اضافه^۱: متن مقصد شامل متنی است که در متن مبدأ موجود نیست.
- * حذف^۲: محتوایی در متن مقصد از قلم افتاده است که در متن مبدأ وجود دارد.
- * سوءترجمه^۳: محتوای متن مقصد محتوای متن مبدأ را به درستی بازتاب نمی‌دهد.
- * بلاترجمه^۴: محتوایی که باید ترجمه^۵ می‌شد بدون ترجمه شدن و به زبان مبدأ به متن مقصد منتقل شده است.

۲-۳-۲. خطای سلاست

- اگر متن کاملاً غیر قابل فهم باشد، ما آن را به عنوان خطایی سطح بالا در سلاست در نظر می‌گیریم و اگر تا حدی قابل فهم باشد، اما دارای خطاهایی جزئی باشد یکی از انواع خطای ریزدانه ذیل را برای آن در نظر می‌گیریم:
- * دستور زبان^۶: خطاهای مربوط به دستور زبان یا نحو.
- * سیاق دستوری^۷: متن مقصد از سیاق دستوری اشتباه استفاده می‌کند؛ وقتی انتظار می‌رود از حالت‌های غیررسمی فعل یا ضمیر استفاده شود از حالت‌های رسمی استفاده شود؛ برای مثال انتظار رود از ضمیر محاوره‌ای «اون» استفاده شود، اما از حالت رسمی آن؛ یعنی «ایشان» استفاده شود.
- * املاء^۸: مسائل مربوط به غلط‌های املائی یا تایی در کلمات.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

-
1. Addition
 2. Omission
 3. Mistranslation
 4. Untranslated
 5. Lommel, A., et al.
 6. Grammar
 7. Grammatical Register
 8. Spelling

۲-۳-۳. خطای قرارداد محلی^۱

متن به قراردادهای ظاهری ویژه محل پایبند نیست و الزامات ارائه محتوا در منطقه زبان مقصد را نقض می‌کند. برای مثال، در یک متن فارسی به جای استفاده از گیومه (©) از علامت نقل قول غربی (””) استفاده شود.

۲-۳-۴. خطای اصطلاح شناسی^۲

یک اصطلاح (واژه مختص به حوزه‌ای خاص) با اصطلاح دیگری که دور از انتظار است، ترجمه شود.

برای ارزیابی یک سامانه معمولاً شمردن خطاها کافی نیست، بلکه ارزیابان باید بدانند هر کدام از این خطاها چقدر شدت^۳ دارند. شدت به طبیعت خطای یافته شده فی نفسه و اثر آن بر میزان مفید بودن ترجمه بستگی دارد. هر چه خطا شدیدتر باشد احتمال و شدت اثر منفی آن بر مخاطب بیشتر می‌شود. (Lommel, 2018: 120) مدل DQF-MQM دارای چهار نوع شدت خطا به این شرح (تاوس، بی تا) است:

* وخیم^۴: خطاهایی که ممکن است حامل پیامدهای ناگوار بهداشتی، ایمنی، حقوقی، یا مالی باشند، از دستورالعمل‌های استفاده ژئوپلتیکی تخطی کنند، به اعتبار شرکت آسیب بزنند، باعث اختلال در عملکرد نرم افزار، محصول، یا خدمات شوند، توهین آمیز تلقی شوند، یا ...

* عمده^۵: خطاهایی که ممکن است باعث سردرگمی یا گمراهی مخاطب شوند، یا مانع استفاده صحیح محصول / خدمت توسط کاربر شوند؛ چون تغییر قابل توجهی در معنای رخ داده است یا خطاها در بخشی قابل مشاهده یا مهم از محتوا رخ داده‌اند.

-
1. Locale Convention
 2. Terminology
 3. Severity
 4. Critical
 5. Major

* خرده^۱: خطاهایی که باعث از دست رفتن معنا و سردرگمی یا گمراهی مخاطب نمی‌شوند، اما مورد توجه واقع می‌شوند، از کیفیت اسلوبی، سلاست و وضوح یا جذابیت متن می‌کاهند.

* خنثی^۲: برای بایگانی اطلاعات اضافی، مسائل یا تغییراتی که باید انجام شوند، اما خطا شمرده نمی‌شوند؛ برای مثال آن‌ها صرفاً منعکس‌کننده انتخاب یا سبک ترجیحی بازبین هستند، اشتباهاتی هستند که تکرار شده‌اند، یا تغییراتی در دست‌والعمل یا واژه‌نامه هستند که هنوز پیاده‌سازی نشده‌اند و یا تغییری است که باید انجام شود، اما مترجم از آن بی‌خبر بوده است.

می‌توان از این سطوح شدت برای دادن نمره‌هایی منفی به هر کدام از خطاهای موجود در مجموعه‌ای از ترجمه‌ها استفاده کرد و از آن‌ها برای محاسبه سنجه‌ای برای ارزیابی^۳ کلی کیفیت مجموعه بهره برد. به ازای هر سطح شدت جریمه یا نمره منفی خاصی در نظر گرفته می‌شود؛ برای مثال برای هر خطای سطح وخیم ۱۰ نمره منفی، برای سطح عمده ۵ نمره منفی، برای سطح خرده ۱ نمره منفی و برای سطح خنثی ۰ نمره منفی در نظر گرفته می‌شود، سپس این مقادیر با هم جمع می‌شوند تا مقدار نمره منفی کل به دست بیاید. سپس می‌توان با این فرمول (Lommel, 2018: 121- 122) مقدار امتیاز کل کیفیت ترجمه‌های مورد نظر را به دست آورد:

$$\text{نمره منفی کل} = 1 - \frac{\text{تعداد کل کلمات}}{\text{امتیاز}}$$

1. Minor

2. Neutral

۳. تعریف اصطلاح «سنجه ارزیابی» (evaluation metric): معیار یا اندازه‌ای که برای سنجش و اندازه‌گیری کیفیت به کار می‌رود (Rothwell et al., 2023: xiii). تفاوت «سنجه» با مفهوم ساده «اندازه» در این است که سنجه با هدف خاصی اندازه‌گیری و محاسبه می‌شود؛ برای مثال، اینکه خانه‌ای ۶۰ متر مربع مساحت دارد صرفاً یک اندازه است، اما اینکه به ازای هر نفر از خانواده‌ای سه نفره ۲۰ متر مربع مساحت دارد، یک سنجه است که به هدف پی بردن به مناسب بودن یا نبودن این خانه برای این خانواده محاسبه شده است (Lommel & Melby, 2018: 7).

۳. پیشینه پژوهش

چند مقاله به ترجمه ماشینی بین عربی و فارسی پرداخته‌اند. نظری (۱۳۹۳) در مقاله «گونه‌شناسی چالش‌ها و جایگاه ترجمه ماشینی از عربی به فارسی» چالش‌های پیش رو در ترجمه ماشینی از عربی به فارسی از جمله مسائل ابهام نوشتاری، واژگانی و اعراب را به تفصیل بیان می‌کند. وی همچنین نمونه‌ای از ترجمه انجام شده توسط ماشین‌های ترجمه برخط رایگان را ارائه می‌دهد که نشان می‌دهد هنوز راه درازی تا رسیدن به ترجمه ماشینی عربی به فارسی قابل قبول پیش رو داریم (نظری، ۱۳۹۳: ۴۰ و ۵۴).

بشار (۲۰۲۱) در مقاله «اشکالیات الترجمة الآلية بين العربية والفارسية» به توضیحاتی در مورد ابهام‌های واژگانی مشکل‌آفرین برای ترجمه بین عربی و فارسی و همچنین به بررسی ترجمه متونی از انواع مختلف با مترجم گوگل می‌پردازد و نتیجه می‌گیرد که هنوز به کیفیت مطلوب نرسیده است. اگرچه برای انواعی از متون (فنی و غیر ادبی) بهتر از انواعی دیگر (شامل امثال و کنایات و اصطلاح‌ها) عمل می‌کند (بشار، ۲۰۲۱: ۱۹۲-۱۹۳).

عبدالعباس (۲۰۲۱) در مقاله «واکاوی ترجمه ماشینی تارنماهای خبری فارسی به عربی» ترجمه ماشینی برخی متون خبری را از فارسی به عربی مورد ارزیابی قرار می‌دهد و انواع اشتباهات موجود در آن را گزارش می‌دهد (عبدالعباس، ۲۰۲۱: ۵۱۴ و ۵۳۴).

چندین پژوهش علمی با موضوع ارزیابی تطبیقی ترجمه ماشینی به جفت‌زبان‌های عربی-انگلیسی یا فارسی-انگلیسی پرداخته‌اند، اما هنوز هیچ پژوهشی در این حوزه برای جفت زبان عربی-فارسی منتشر نشده است. این پژوهش‌ها عموماً ابتدا یک مجموعه آزمایش انتخاب کرده‌اند، سپس ترجمه آن را توسط چند ماشین ترجمه انجام داده و خروجی را با استفاده از یک یا چند یک از روش‌های خاص ارزیابی ترجمه ماشینی مورد مطالعه قرار داده‌اند. در ادامه خلاصه‌ای از جدیدترین مطالعات این حوزه آمده است.

بن میلاد (۲۰۲۲) در پژوهشی با عنوان «ارزیابی تطبیقی کیفیت ترجمه ماشینی عصبی در ترجمه انگلیسی عربی»^۱ متونی عربی و متونی انگلیسی هر کدام حدود ۱۰۰ جمله را از یک پیکره موازی متون خبری انتخاب کرد، و آن‌ها را توسط ماشین‌های ترجمه رایگان و

1. Comparative Evaluation of Neural Machine Translation Quality in Arabic English Translation

غیررایگان از جمله گوگل، بینگ و یاندکس به زبان مقابل ترجمه کرد. وی متون به دست آمده را با دو روش، ارزیابی انسانی برآورد کیفیت و ارزیابی خودکار با دو مدل مختلف، مورد بررسی قرار داد. ارزیابی انسانی گوگل را به برتری شناخت، اما مدل‌های ارزیابی خودکار مختلف نتایج متفاوتی به همراه داشت و بر برتری هیچ کدام از آن‌ها اتفاق نداشت (بن میلاد، ۲۰۲۲: ۱۴۲).

المحاسبیس (۲۰۲۰) در رساله دکتری خود با عنوان «ارزیابی در زمانی گوگل ترنزلت، مترجم مایکروسافت و صخر در ترجمه انگلیسی به عربی»^۱ پیکره‌ای حدوداً ۴۵۰۰ واژه‌ای از انواع مختلف متون عربی و انگلیسی می‌سازد و ترجمه آن به انگلیسی و عربی توسط سامانه‌هایی شامل گوگل و بینگ را به دو روش انسانی برآورد کیفیت و تحلیل خطا مورد ارزیابی قرار داد، و به این نتیجه رسید که گوگل ترجمه بهتری ارائه می‌دهد (المحاسبیس، ۲۰۲۰: ۳).

ابوعیاش (۲۰۱۷) در مقاله‌ای با عنوان «خطاها و عدم خطاها در ترجمه ماشینی انگلیسی به عربی ساختارهای جنسیتی در متون فنی»^۲ چند متن فنی انگلیسی با حدود ۴۰۰ واژه را به چند ماشین ترجمه شامل گوگل و بینگ داد و ترجمه عربی خروجی را به روش ارزیابی انسانی مبتنی بر نقاط بازرسی (تطابق جنس فعل - فاعل، صفت - موصوف، و ضمیر - مرجع) مورد مطالعه قرار داد و به این نتیجه رسید که گوگل و بینگ تعداد خطای برابری داشتند (ابوعیاش، ۲۰۱۷: ۷۳، ۷۹).

الشلبی و دیگران (۲۰۱۷) در پژوهشی با عنوان «ارزیابی ترجمه ماشینی از عربی به انگلیسی و بالعکس»^۳ پیکره موازی عربی - انگلیسی شامل ۶۰ جمله را از اینترنت جمع‌آوری کردند، سپس ترجمه هر جمله را به زبان دیگر توسط ماشین‌های ترجمه گوگل، بینگ، بابلون و سیستم مورد ارزیابی خودکار قرار دادند و به این نتیجه رسیدند که گوگل ترجمه عربی به انگلیسی و برعکس بهتری ارائه می‌دهد (الشلبی و دیگران، ۲۰۱۷: ۱).

-
1. Diachronic Evaluation of Google Translate, Microsoft Translator and Sakhr in English-Arabic Translation
 2. Errors and non-errors in English-Arabic machine translation of gender-bound constructs in technical texts
 3. Evaluating Machine Translations from Arabic into English and Vice Versa

۴. روش پژوهش

اولین گام برای اجرای ارزیابی ترجمه ماشینی، انتخاب مجموعه آزمایش است. همچنان که در سؤالات پژوهش منعکس شده است، این پژوهش سعی دارد به مقایسه کیفیت خروجی ماشین‌های ترجمه عربی به فارسی و برعکس بپردازد و به نوع خاصی از متون این دو زبان محدود نباشد. برای رسیدن به این هدف از جملات موجود در دو کتاب فرهنگ بسامدی عربی^۱ و فارسی^۲ انتشارات راتلج استفاده شد؛ چون این جملات از پیکره‌هایی با انواع مختلف متون نوشتاری و صوتی انتخاب شده‌اند و همچنانکه از عنوان کتاب‌ها نیز برمی‌آید سعی شده طوری انتخاب شوند که نماینده زبان‌های عربی و فارسی باشند و محدود به نوع خاصی از متون نباشند (Miller, et al., 2018: 1- 2 and Buckwalter & Parkinson, 2010: 3- 4). ۶۰ جمله اول از هر کدام از این دو کتاب به عنوان مجموعه آزمایش برای هر کدام از دو جهت ترجمه عربی به فارسی و برعکس انتخاب شدند. این جملات به حالت‌های^۳ مختلف نوشتاری و گفتاری، انواع مختلف خبری، داستانی، علمی و... و موضوعات مختلف سیاسی، دینی، تاریخی، جغرافیایی و... هستند. گام دوم انتخاب ماشین‌های ترجمه مورد آزمایش است. در اینجا همانطور که در مقدمه اشاره شد، سامانه‌های برخط رایگان پشتیبانی‌کننده از عربی و فارسی انتخاب شدند. با جست‌وجو در اینترنت شش سامانه با این مشخصات به دست آمدند: گوگل ترنزلت^۴، بینگ ترنزلتور^۵، یاندکس ترنزلت^۶، رورسو^۷، مدرن‌ام‌تی^۸، و نیوترنس^۹. در جدول (۱) معرفی اجمالی این سامانه‌ها و قابلیت‌های ادعایی‌شان آمده است.

1. A Frequency Dictionary of Arabic: Core Vocabulary for Learners (by Buckwalter & Parkinson)
2. A Frequency Dictionary of Persian: Core Vocabulary for Learners (by Miller et al.)
3. Mode
4. Google Translate: <https://translate.google.com/>
5. Microsoft Translator: <https://www.bing.com/translator>
6. Yandex Translate: <https://translate.yandex.com/>
7. Reverso: <https://www.reverso.net/text-translation>
8. ModernMT: <https://www.modernmt.com/translate>
9. NiuTrans: <https://translate.niutrans.com/>

جدول ۱. معرفی اجمالی مشخصات و قابلیت‌های ادعایی ترجمه‌های ماشینی برخط رایگان

نام ماشین ترجمه	گوگل	بینگ	یاندکس	رورسو	مادرنام‌تی	نیوترنس
کشور توسعه‌دهنده	آمریکا	آمریکا	روسیه	فرانسه	ایتالیا	چین
تعداد زبان پشتیبانی شده	۱۳۳	۸۸	۱۰۰	۲۶	۲۰۰	+۴۵۰
حداکثر تعداد نویسه	۵۰۰۰	۱۰۰۰	۱۰۰۰۰	۲۰۰۰	۵۰۰۰	۵۰۰۰
ترجمه صوت	✓	✓	✓	✗	✗	✗
پخش صوت	✓	✓	✓	✓	✗	✗
ترجمه تصویر	✓	✗	✓	✗	✗	✗
ترجمه سند	✓	✗	✓	✓	✗	✗
ترجمه وبسایت	✓	✗	✓	✓	✗	✗
ترجمه دستخط	✓	✓	✓	✗	✗	✗
ترجمه برونخط (آفلاین)	✓	✓	✓	✗	✗	✗

گام سوم انتخاب روشی برای ارزیابی خروجی ترجمه ماشینی است. در این پژوهش از ارزیابی انسانی استفاده خواهد شد که بر روش‌های ماشینی برتری کلی دارد (Kit & Wong, 2023: 227). روش‌های ارزیابی خودکار مبتنی بر ترجمه مرجع مخصوصاً برای زبان‌هایی مثل عربی و فارسی که از نظر صرفی غنی هستند یا از نظر منابع پیکره‌ای کمبود دارند، مناسب نیست (Beseiso, M., et al., 2022: 189).

از میان روش‌های انسانی روش مبتنی بر تحلیل و طبقه‌بندی خطا استفاده خواهد شد که بیش از همه از جانب‌گیری و نظر شخصی که از اصالت علمی پژوهش می‌کاهد به دور هستند. مدل‌های بسیاری برای این نوع از ارزیابی وجود دارد که در اینجا طبقه‌بندی موسوم به DQF-MQM که یک مورد به روز و استاندارد است (راتول و دیگران، ۲۰۲۳: ۱۰۷) اختیار شده است.

گام چهارم در ارزیابی انسانی انتخاب نشان‌گذار (ان) است. در اینجا نویسنده مسئول این مقاله خود نقش نشان‌گذار را نیز بازی کرده است. زبان فارسی زبان اصلی پژوهشگر است و به زبان‌های عربی و انگلیسی نیز تسلط نسبی دارد. ترجمه انگلیسی جملات مجموعه آزمایش در کتاب‌های مورد اشاره موجود هستند و می‌توانند به نحوی همچون ترجمه

مرجع عمل کنند. بنابراین، به نظر می‌رسد پژوهشگر شایستگی کافی برای ایفای نقش نشان‌گذاری این پیکره را دارا باشد.

گام پنجم انتخاب ابزاری برای تسهیل اجرای فرآیند ارزیابی است. در اینجا از قالب اکسل مخصوص مدل DQF-MQM، موجود روی سایت رسمی شرکت ابداع کننده آن، استفاده می‌شود.^۱ این قالب اکسل دارای هشت زبانه است؛ زبانه اول عنوان را دربر دارد. زبانه دوم مقدمه‌ای است که نحوه استفاده از قالب را توضیح داده است. زبانه سوم جدولی است که مشخصات کلی متن مبدأ همچون تعداد کلمات در آن وارد می‌شود. زبانه چهارم جدولی است که ارزیاب متن مبدأ را بخش به بخش^۲ در آن وارد می‌کند و خطا(های) موجود در هر بخش را با انتخاب گزینه مربوطه مشخص می‌کند. زبانه پنجم نمره منفی و امتیاز ترجمه را به صورت خودکار محاسبه کرده و نمایش می‌دهد. زبانه‌های بعدی به توضیح سطوح و شدت‌های مختلف خطاها پرداخته‌اند که در بخش ادبیات پژوهش آمد (TAUS, Undated).

گام ششم اجرای عملی فرآیند ارزیابی است. جملات مجموعه آزمایش تک تک وارد ماشین‌های ترجمه شدند و خروجی آن‌ها دریافت شد. هر کدام از این جملات به عنوان یک بخش محسوب می‌شوند که باید در خانه‌های مربوطه در قالب اکسل وارد شوند. این جملات وارد فایل اکسل شده و خطاها در خانه‌های مربوطه تعیین شدند. طبقه‌بندی خطای موجود در این قالب دارای شاخه‌ها و زیرشاخه‌های بسیاری است که همگی برای ارزیابی ترجمه ماشینی مناسب نیستند؛ نشان‌گذار خود را محدود می‌کند به زیر مجموعه‌ای از این طبقه‌بندی که برای ارزیابی ترجمه ماشینی توصیه شده و در بخش سوم مبانی نظری شرح آن آمد.

۱. آدرس صفحه دانلود:

<https://info.taus.net/dqf-mqf-error-typology-template-download>

۲. تعریف اصطلاح «بخش» (Segment): واحدهای کمینه‌ای که می‌توان متن مبدأ را به آن شکست و بخش متناظر با هر کدام را در متن مقصد (ترجمه شده توسط ماشین یا به کمک ماشین) یافت؛ برای مثال جمله، تیر و محتویات یک خانه از یک جدول. شبیه به مفهوم واحد ترجمه در مطالعات ترجمه است. (Rothwell et al., 2023: xix; Sin-) (Wai, 2004: 203)

۵. یافته‌ها

در این بخش گزارشی از داده‌های به دست آمده از فرآیند تحلیل و طبقه‌بندی خطای انجام شده، ارائه می‌شود. داده‌های مربوط به هر کدام از ماشین‌های ترجمه برای هر کدام از دو جهت عربی به فارسی و فارسی به عربی در قالب‌های اکسل جداگانه‌ای قرار دارد که همگی از طریق اینترنت قابل دسترسی است^۱. حجم این داده‌ها بیش از این است که در این مقابله قابل گنجاندن باشد. در ادامه توضیحات و مثال‌هایی دربارهٔ انواع و شدت‌های مختلف خطاهای مشاهده شده در ترجمه‌ها آورده می‌شود:

* گوگل جمله شماره ۲۸ عربی «اللی بیدخن وما بیعمل ریاضة نهائیا بیتعب بربع ساعة» را به «هر کس سیگار می‌کشد و اصلاً ورزش نمی‌کند در یک ربع خسته می‌شود» ترجمه کرده است که کلمه «اصلاً» در آن اضافه است؛ بنابراین به عنوان خطای صحت از نوع اضافه محسوب شد. با توجه به اینکه این خطا تغییر چندانی در معنا حاصل نمی‌کند، خطایی از شدت خرده دانسته شد.

* رورسو جمله شماره ۲۷ فارسی «افراد معمولی هم میتوانند برند.» را به «یمنکن للناس العادیین أيضاً» ترجمه کرده که معادل کلمه «برند» در آن وجود ندارد؛ بنابراین، خطای صحت از نوع حذف در آن رخ داده است. با توجه به اینکه این حذف موجب سردرگمی خواننده می‌شود، خطایی از شدت عمده رخ داده است.

* نیوترنس جمله شماره ۳۶ فارسی «من هم بهائی هستم.» را به «أنا أيضا أحمق.» ترجمه کرده است که خطای صحت از نوع سوء ترجمه رخ داده است، و با توجه به اینکه ممکن است، توهین‌آمیز تلقی شود، سطح شدت خطا وخیم تشخیص داده شد.

* یاندکس جمله شماره ۷ فارسی «چند روز در هفته ورزش میکنی؟» را به «کم یوما فی الأسبوع هل ممارسة؟» ترجمه کرده که نا مفهوم است؛ بنابراین، خطای از نوع سلاست رخ داده است. این نوع خطا در طبقه‌بندی MQM با اصطلاح نامفهوم^۲ شناخته می‌شود که در DQF-MQM معادلی ندارد؛ بنابراین، در قالب اکسل تنها با خطای سطح بالای

1. <https://github.com/mazyar1990/MTE-Spreadsheets>

2. Unintelligible

سلاست مشخص شده است و نوع آن تعیین نشده است (Lommel, et al., 2015: 16 (and TAUS, Undated).

* مادرنام‌تی جمله اخیر را به «کم عدد الأيام فی الأسبوع التي تمارس فيها الرياضة؟» که بهتر بود به جمله کوتاه‌تر «کم یوما فی الأسبوع تمارس الرياضة؟» ترجمه می‌شد، ترجمه کرده است؛ بنابراین، می‌توان آن را خطای سلاست از نوع دستوری دانست. با توجه به اینکه امکان دارد این مورد اصلاً خطا محسوب نشود از شدت ختی شناسایی شده است.

* نیوترنس جمله شماره ۱ عربی «أمضى البائع حياته في البحث عن الألباس» را به «اون فروشده تمام عمرش رو به دنبال الماس گذاشته بود» ترجمه کرد. کلمات «اون» و «رو» در حالت گفتاری یا عامیانه هستند و با حالت نوشتاری یا رسمی متن مبدأ هماهنگی ندارند؛ بنابراین، خطای سلاست از نوع سیاق دستوری رخ داده است.

* نیوترنس جمله شماره ۲۳ فارسی «رسید به سیصد تومن.» را به «وصل إلى ثلاثمائة تومين» ترجمه کرد که در کلمه «تومين» خطای سلاست از نوع املاتی دارد.

* یاندکس جمله شماره ۷ عربی «سقط الحاج محمود على الأرض من شدة وقع الخبر على نفسه» را به جمله‌ای فاقد نقطه ترجمه کرد؛ بنابراین، خطای سلاست از نوع علائم نگارشی رخ داد. این نوع از خطا در زیرمجموعه پیشنهاد شده در مبانی نظری نیامده بود، اما نشان‌گذار در عمل متوجه شد که برای ثبت کامل خطاها به آن نیاز دارد.

* رورسو جمله شماره ۸ عربی «هل هذا يعني أنكم ستدعمون موقف سوريا؟» را به «آیا این به این معنی است که شما از موضع سوریه حمایت می‌کنید؟» ترجمه کرده که در آن از علامت سؤال لاتین استفاده شده است؛ بنابراین، خطای قرارداد محلی رخ داده است.

* رورسو جمله شماره ۱۸ عربی «عين في قسم الهیستولوجی بدلا من الجراحة العامة التي كان يحلم بها» را به «او به جای جراحی عمومی که در خواب دید، به سمت دپارتمان هیدرولوژی منصوب شد» ترجمه کرده که در آن اصطلاح «الهیستولوجی» اشتباهاً به «هیدرولوژی» ترجمه شده است؛ بنابراین، خطای اصطلاح‌شناسی رخ داده است.

در زبانه پنجم هر کدام از فایل‌های اکسل مربوط به هر یک از ماشین‌های ترجمه برای هر یک از دو جهت ترجمه از عربی به فارسی و برعکس، تعداد انواع خطاها و امتیاز منفی

آن‌ها محاسبه و نمایش داده شده است که در دو جدول (۲) و (۳) به صورت خلاصه ارائه شده‌اند

جدول ۲. تعداد و نمره منفی برای هر نوع خطا برای ترجمه از عربی به فارسی

نام ماشین ترجمه	گوگل	بینگ	یاندکس	روسو	مادرن‌ام‌تی	نیوترنس
خطای صحت	۳۱	۹۹	۲۷	۷۵	۱۲۲	۳۳
خطای سلاست	۷	۷	۱۷	۳۰	۱۱	۵۲
قرارداد محلی	۰	۰	۱	۰	۴	۰
اصطلاح شناسی	۰	۰	۰	۰	۰	۰
تعداد/جریمه کل	۳۸	۱۰۶	۴۶	۱۱۱	۱۹۶	۴۸
امتیاز	۸۴٪	۸۳٪	۸۱٪	۷۰٪	۸۰٪	۶۷٪

جدول ۳. تعداد و نمره منفی برای هر نوع خطا برای ترجمه از فارسی به عربی

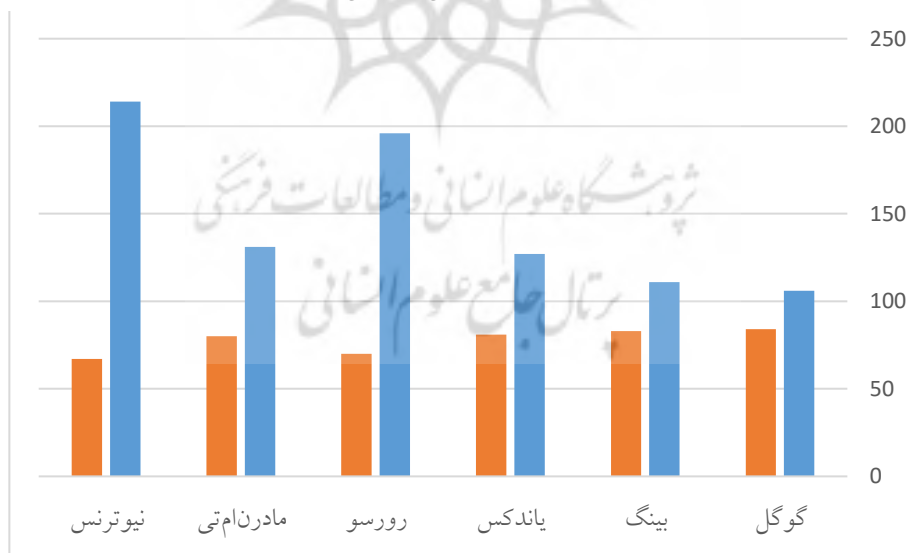
نام ماشین ترجمه	گوگل	بینگ	یاندکس	روسو	مادرن‌ام‌تی	نیوترنس
خطای صحت	۲۴	۶۰	۲۱	۷۳	۲۱	۲۶
خطای سلاست	۱	۱	۱	۵	۲	۶
قرارداد محلی	۰	۰	۰	۲	۰	۰
اصطلاح شناسی	۰	۰	۰	۰	۰	۰
تعداد/جریمه کل	۲۵	۶۱	۲۲	۷۸	۲۲	۳۲
امتیاز	۸۷٪	۸۳٪	۷۹٪	۷۶٪	۷۹٪	۷۲٪

برای محاسبه امتیاز کل هر کدام از ماشین‌های ترجمه باید تعداد کلمات کل جملات (برای عربی ۶۵۶ کلمه و برای فارسی ۴۵۴ کلمه) و میزان نمره منفی کل مورد اشاره در جدول‌های بالا را در فرمول بیان شده در مبانی نظری جایگذاری کنیم. امتیازهای موجود در سطر آخر جدول‌های بالا از این طریق محاسبه شدند. نمودارهای (۱) و (۲) امتیاز کل (رنگ نارنجی) و نمره منفی کل (رنگ آبی) ماشین‌های ترجمه از عربی به فارسی و برعکس را نمایش می‌دهند.

نمودار ۱. نمره منفی کل (آبی) و امتیاز کل (نارنجی) ماشین‌های ترجمه برخط رایگان برای ترجمه از عربی به فارسی



نمودار ۲. نمره منفی کل (آبی) و امتیاز کل (نارنجی) ماشین‌های ترجمه برخط رایگان برای ترجمه از فارسی به عربی



نشان‌گذار در عمل اجرای ارزیابی با مواردی مواجه شد که با زیرمجموعه DQF-MQM پیشنهادی برای ارزیابی ترجمه ماشینی ارائه شده در مبانی نظری، سازگاری نداشت. خطای بلاترجمه اصلاً مشاهده نشده است؛ بنابراین ظاهراً کاربرد چندانی برای ارزیابی ماشین‌های ترجمه امروزی ندارند. علاوه بر این، خطای سلاست علائم نگارشی در زیرمجموعه پیشنهادی موجود نبود، اما برای نشان‌گذاری کامل مورد نیاز بود.

نشان‌گذار در فرآیند نشان‌گذاری با مشکلات دیگری نیز مواجه شد. گاهی تعیین دقیق طبقه‌بندی امکان‌پذیر نبود؛ برای مثال، جمله «بسیاری از کردهای ترکیه به استانبول مهاجرت کرده‌اند» توسط مادرنام‌تی این گونه ترجمه شده «هاجر العديد من الأكراد الأتراک إلى اسطنبول» که از طرفی می‌توان گفت که «الأكراد الأتراک» یک سوءترجمه از «کردهای ترکیه» است و از طرفی دیگر می‌توان گفت که «الأكراد الأتراک» همان «أكراد ترکیا» است که صورت دستوری غلط‌اندازی به خود گرفته است.

همچنین تصمیم‌گیری در مورد شدت برخی از خطاها نیز سخت بود. برای مثال اسم خاص «عجلون» در جمله عربی شماره ۲۶ در ترجمه یاندکس به صورت «اجلون» آمده است؛ تصمیم‌گیری در مورد اینکه آیا این کلمه توسط خواننده درک خواهد شد و بنابراین شدت خرده داشته باشد، یا اینکه املاي این واژه به گونه‌ای تغییر کرده که موجب سردرگمی خواننده می‌شود و بنابراین شدت عمده داشته باشد، سخت است. این مشکل در حقیقت از مجهول بودن خواننده در این پژوهش نشأت می‌گیرد؛ اگر درک کافی از نوع مخاطب وجود داشته باشد، شاید تصمیم‌گیری در چنین مواردی راحت‌تر باشد.

مشکل دیگری که در تعیین شدت خطاها با آن مواجه شدیم، این بود که ممکن بود یک ماشین ترجمه، یک ترجمه نامفهوم از یک جمله ارائه دهد و بنابراین نمره منفی سطح عمده بگیرد؛ در حالی که یک ماشین دیگر تنها بخشی کوچک از آن جمله را اشتباه ترجمه کند و آن هم نمره منفی سطح عمده بگیرد. برای مثال جمله شماره ۲۹ عربی «قال إن المحافظة استعدت جيداً لهذه المناسبة» توسط رورسو به «او گفت که بخشدار به این مناسبت به خوبی آماده شده است» ترجمه شده که نامفهوم است و چون خواننده را سردرگم می‌کند باید نمره منفی سطح عمده بگیرد. علاوه بر این، این جمله توسط

مادر نام‌تی به «او گفت که فرمانداری به خوبی برای این مناسبت آماده شده است» ترجمه شده است که خطای سوء ترجمه فقط در کلمه «فرمانداری» رخ داده است و چون خواننده نمی‌تواند به واژه صحیح که «استان» است، پی ببرد؛ بنابراین، باید شدت از سطح عمده در نظر گرفته شود. در این صورت هر دو نمره منفی برابری خواهند گرفت، اما واضح است که باید رورسو نمره منفی بیشتری بگیرد، چون اشتباهات بیشتری مرتکب شده است که موجب نامفهوم شدن کل جمله شده است. برای جلوگیری از این مشکل، نشان‌گذار از سطح شدت و خیم برای ترجمه نامفهوم استفاده کرد.

همچنانکه در جدول‌های پیشینی مشاهده می‌شود، تعداد خطاهای اصطلاح‌شناسی و قرارداد محلی بسیار کم بود. کم بودن تعداد خطاهای اصطلاح‌شناسی ناشی از کم بودن تعداد اصطلاحات تخصصی در مجموعه آزمایش بود. اما کم بودن تعداد خطاهای قرارداد محلی بیشتر به طبیعت خود این نوع خطا برمی‌گردد؛ این نوع خطا منحصر است به عدم رعایت قالب محلی در مواردی معدود همچون علائم نگارشی، آدرس، تلفن، تاریخ، واحدهای اندازه‌گیری و ارز که تنوع یا کاربرد کمی دارند. جالب است که ماشین‌های ترجمه مواردی همچون تبدیل قالب تاریخ را به صورت هوشمند انجام می‌دهند؛ برای مثال، تمامی ماشین‌های ترجمه بجز نیوترنس در ترجمه جمله شماره ۵ عربی سال ۱۹۷۹ را به ۱۳۵۷ (اما در یک مورد به ۱۳۵۸) ترجمه کردند.

نکته دیگری که در جدول‌ها و نمودارهای بالا به چشم می‌خورد این است که عموماً کیفیت ترجمه در جهت فارسی به عربی بالاتر از جهت عربی به فارسی بود. این مورد می‌تواند به این دلیل باشد که در این پژوهش جملات مجموعه آزمایشی فارسی از جملات جملات مجموعه آزمایشی عربی ساده‌تر بودند و لزوماً به این معنا نیست که کیفیت ترجمه ماشین‌ها در جهت فارسی به عربی بهتر از جهت عربی به فارسی است.

بحث و نتیجه‌گیری

در این پژوهش کیفیت ترجمه عربی به فارسی و برعکس شش سامانه ترجمه ماشین‌ها برخط رایگان با استفاده از مدل DQF-MQM مورد ارزیابی واقع شد. همچنان که در جدول‌های (۲) و (۳) و نمودارهای (۱) و (۲) دیده می‌شود، این سامانه‌ها برای ترجمه عربی به فارسی به ترتیب کیفیت خروجی بالا به پایین از این قرار هستند: ۱- گوگل، ۲- بینگ، ۳-

یاندکس، ۴- مادرنام‌تی، ۵- رورسو و ۶- نیوترنس و برای ترجمه از فارسی به عربی از این قرار: ۱- گوگل، ۲- بینگ، ۳- یاندکس و مادرنام‌تی، ۴- رورسو و ۵- نیوترنس. همچنین می‌توان آن‌ها را در سه رده قرار داد که ماشین‌های هر رده امتیازهای نزدیک به هم دارند: ۱- گوگل و بینگ، ۲- یاندکس و مادرنام‌تی و ۳- رورسو و نیوترنس. این نتایج با آنچه در بخش پیشینه پژوهش آمد، مبنی بر اینکه گوگل غالباً توانسته بود ترجمه (بین عربی و انگلیسی) بهتری به نسبت سایر ماشین‌های ترجمه ارائه دهد، سازگاری دارد. لازم به ذکر است که این رتبه‌بندی به معنای برتری مطلق یکی از این ماشین‌های ترجمه بر دیگری نیست. برای مثال، نیوترنس رتبه آخر شده است، اما در ترجمه جمله فارسی شماره ۵ بهتر از گوگل عمل کرده است. بنابراین، می‌توان گفت که نتایج در بهترین حالت تنها بر یک برتری احتمالی دلالت دارند.

همین برتری احتمالی هم ممکن است، با گذر زمان تغییر کرده و در آینده ماشین ترجمه دیگری بر گوگل و بینگ فائق آید. بنابراین، توصیه می‌شود، چنین آزمایش‌هایی به صورت دوره‌ای تکرار شود. همچنین توصیه می‌شود، آزمایش‌های دیگر با تفکیک انواع متن صورت پذیرد تا به درک بهتری از ترجمه ماشینی عربی به فارسی و برعکس انواع مختلف متون برسیم. همچنین آزمایش‌های دیگری با روش‌های دیگر بیان شده در بخش مبانی نظری انجام شود یا به جای وارد کردن جمله به جمله که در این پژوهش انجام شد، متون به صورت کامل به ماشین ترجمه وارد شوند.

تعارض منافع

تعارض منافع ندارم.

ORCID

Shahryar Niazi



<https://orcid.org/0000-0002-5157-0009>

Mahmood Bijankhan



<https://orcid.org/0000-0002-4175-6854>

Mazyar Pashaei



<https://orcid.org/0000-0002-9633-5715>

منابع

- بشار، زینب محمد إبراهيم الدسوقي. (۲۰۲۱). إشکالیات الترجمة الآلية بين العربية والفارسية. *مجلة كلية الآداب والعلوم الإنسانية جامعة قناة السويس*، ۳۹ (۲)، ۱۵۰-۲۰۵.
- عبدالعباس، عمر أمين. (۲۰۲۱). واكاوی ترجمه ماشینی تارنماهای خبری فارسی به عربی. *مجلة الآداب*، ۱۳۸ (۱)، ۵۱۳-۵۳۶.
- نظری، علیرضا. (۱۳۹۳). گونه‌شناسی چالش‌ها و جایگاه ترجمه ماشینی از عربی به فارسی. *پژوهش‌های ترجمه در زبان و ادبیات عربی*، ۴ (۱۱)، ۲۹-۵۶. DOR: [20.1001.1.22519017.1393.4.11.2.6](https://doi.org/10.1001.1.22519017.1393.4.11.2.6)

English References

- Abu-Ayyash, E. A. S. (2017). Errors and non-errors in English-Arabic machine translation of gender-bound constructs in technical texts. *Procedia Computer Science*, 117, 73–80. <https://doi.org/10.1016/j.procs.2017.10.095>
- Almahasees, Z. (2020). Diachronic Evaluation of Google Translate, Microsoft Translator and Sakhr in English-Arabic Translation. The University of Western Australia.
- Al-Shalabi, R., Kanaan, G., Al-Sarhan, H., Drabsh, A., & Al-Husban, I. (2017). Evaluating Machine Translations from Arabic into English and Vice Versa. *International Research Journal of Electronics and Computer Engineering*, 3(2). <https://doi.org/10.24178/irjece.2017.3.2.01>
- Ben Milad, K. (2022). Comparative Evaluation of Neural Machine Translation Quality in Arabic English Translation. *New Trends in Translation and Technology 2022*, 142–151. DOI: [\[https://doi.org/10.59670/jns.v34i.1636\]](https://doi.org/10.59670/jns.v34i.1636)
- Beseiso, M., Tripathi, S., Al-Shboul, B., & Aljadid, R. (2022). Semantics based english-arabic machine translation evaluation. *Indonesian Journal of Electrical Engineering and Computer Science*, 27(1), 189–197. <http://doi.org/10.11591/ijeecs.v27.i1.pp189-197>
- Buckwalter, T., & Parkinson, D. B. (2010). *A Frequency Dictionary of Arabic: Core vocabulary for learners*. Routledge.
- Chatzikoumi, E. (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, 26(2), 137–161. <https://doi.org/10.1017/S1351324919000469>
- Holmes, J. (1988). *The Name and Nature of Translation Studies*. In *Translated!* (Broeck, Raymond van den, pp. 67–80). Brill. https://doi.org/10.1163/9789004486669_008

- Hutchins, W. J. (2003). *The development and use of machine translation systems and computer-based translation tools*. Bahri. Available online at: <https://aclanthology.org/www.mt-archive.info/00/IJT-2003-Hutchins.pdf> (Last Accessed March 2024)
- Kit, C., & Wong, B. T. M. (2023). *Evaluation in Machine Translation and Computer-Aided Translation*. In Routledge Encyclopedia of Translation Technology (pp. 219–244). Routledge. <https://doi.org/10.4324/9781003168348-13>
- Lommel, A. (2018). *Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies*. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment* Vol.1. Springer International Publishing. https://doi.org/10.1007/978-3-319-91241-7_6
- Lommel, A., & Melby, A. (2018). *Tutorial: MQM-DQF: A Good Marriage (Translation Quality for the 21st Century)*. In J. Campbell, A. Yanishevsky, J. Doyon, & D. Jones (Eds.), *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 2: User Track)*. Association for Machine Translation in the Americas. <https://aclanthology.org/W18-1925>
- Malmkjær, K. (2013). *Where are we? (From Holmes's map until now)*. In *The Routledge Handbook of Translation Studies*. Routledge.
- Miller, C., Aghajanian-Stewart, K., Bills, A., Corbett, R., Diaz, J., Golonka, E., Jones, E., Livingston, J., Mostafavi, S., Osthus, P., Ritch, J., Stewart, D., Strong, R., Triebwasser, T., Vinson, & Zajic, D. (2018). *A Frequency Dictionary of Persian: Core vocabulary for learners*. Routledge, Taylor & Francis Group.
- Rothwell, A., Moorkens, J., Fernández-Parra, M., Drugan, J., & Austermeuhl, F. (2023). *Translation Tools and Technologies*. Routledge.
- Sin-Wai, C. (2004). *A dictionary of translation technology*. Chinese University Press.
- Snell-Hornby, M. (2010). *The turns of Translation Studies*. In *Handbook of translation studies* (1, 366–370). John Benjamins
- TAUS. (n.d.). *DQF-MQM Error Typology Template Download*. Retrieved January 21, 2024, from <https://info.taus.net/dqf-mqf-error-typology-template-download> (Last Accessed May 2024)
- White, J. S. (2003). *How to evaluate machine translation*. In *Computers and Translation Studies* (pp. 111–244). <https://doi.org/10.1075/btl.35.16whi>

Translated References to English

- AbdulAbbas, O. A. (2021). Criticism and analysis of the machine translation of news websites from Persian into Arabic. *Al-Adab magazine*, 138 (1), 513-536. [In Persian]
- Bashar, Zainab Mohammad Ibrahim al-Dsoqi. (2021). Problems of machine translation between Arabic and Persian. *Journal of the Faculty of Arts and Human Sciences of Al Suez University*, 39 (2), 150-205. [In Arabic]
- Nazari, A. (2013). Typology of challenges and position of machine translation from Arabic to Persian. *Translation Researches in the Arabic Language and Literature*, 4(11), 29-56. DOR: [20.1001.1.22519017.1393.4.11.2.6](https://doi.org/20.1001.1.22519017.1393.4.11.2.6) [In Persian]



استناد به این مقاله: نیازی، شهریار، بی‌جن‌خان، محمود، پاشائی، مازیار. (۱۴۰۳). ارزیابی تطبیقی کیفیت خروجی ماشین‌های ترجمه برخط رایگان بین عربی و فارسی براساس مدل DQF-MQM. *دوفصلنامه پژوهش‌های ترجمه در زبان و ادبیات عربی*، ۱۴ (۳۰)، ۶۳-۹۰. doi: 10.22054/rctall.2024.78817.1724



Translation Researches in the Arabic Language and Literature is licensed under a Creative Commons Attribution-Noncommercial 4.0 International License.