

Efficiency of Machine Translation in the Language Processing Process; Using Context Clues in Finding the [Exact] Meaning of Quranic Words [In Persian]

Zaynab Shams^{1*}, Sepideh Chehreh²

1 PhD student of Qur'anic and Hadith Sciences, Faculty of Theology, Kashan University, Iran

2 Master of Artificial Intelligence, Islamic Azad University Science and Research Branch, Iran

*Corresponding author: Z.Shams@qom.ac.ir

Received: 24 Sep, 2021 Revised: 27 Nov, 2021 Accepted: 06 Aug, 2022

ABSTRACT

Translation is the transfer of the content of a text from the source language in to the target language, which is done by finding semantic equivalents between the two languages. The most important problems facing translation are the ambiguities in vocabulary and sentence structure. In a division, there are five important types of lexical ambiguity (categorical ambiguities, homophones, homographs, polysemy and transitive ambiguity), and two important types of structural ambiguity (real structural ambiguities and systemic ambiguities). Machine translation (MT), which is a part of the computer-based field of natural language processing (NLP) in computational linguistics and artificial intelligence, is considered as one of the automatic techniques that that convert unstructured text into structured data, and by converting text into information, it has been able to apply further analysis to the data to extract useful information. In this article, which was compiled in a library method, a theoretical plan has been proposed to resolve the issues surrounding the meaning of words in the machine translation of the Quran, the purpose of which is to help better understand the meaning of the words of the Quran, by taking advantage of the context clues and styles of the expressions. In the proposed method, a more suitable equivalent word is chosen in the target language by taking advantage of the context rule and text mining techniques, and referring to it. In this plan, the context is considered in the scale of words, which can be developed to other types if the conditions are met. In short, this plan has two steps: prioritizing (weighting) the adjacent words next to each other (any word within the range of verses where there is a consensus about their simultaneous descent) and then, comparing with the homonyms words (polysemous), and also comparing the equivalents of a word with the equivalents of other words (synonymization). In order to make the results more accurate, more specifications of the words can be prepared manually, tables that include

things such as whether the verses are Meccan or Medinan, the order of revelation of the Surahs, the concepts and interpretations that are mentioned in the meaning of the words of the Qur'an in dictionaries such as Lisan al-Arab by Ibn Manzur and The Book of Vocabulary in the Strange Qur'an by Al-Ragheb Al-Isfahani and so on. Indexing techniques are used to obtain input data. In the pre-processing stage, the data that is less important (Stop Words) (such as “al-lazi (which)”, “al-lati (that is)”, “lam (not)”, “k'ana (was)”, “kaannama (as if)”, etc.) should be removed to get a better output. To change the shape of the data, the diacritic can be removed to make coding easier, and to reduce the sample size, the infix of the words can be used. In order to prepare a record of specifications for each word that is processed as input, based on the rule of context clues, at first, it is necessary to create a tokenizer, to prepare it in the primary data, and in the entire collection of input verses, a weight should be assigned to each word based on the two criteria of spatial proximity and frequency of repetition. The closer the words are to the desired word or the more it is repeated, the more weight is assigned to it, which represents their stronger semantic connection, and vice versa. Naturally, the words that are in the same verse (have the same number of the verse) have a greater influence than the words that are in other verses and at a further distance. In measuring the frequency criterion, weighted frequency (TF/IDF Weight) is used to show the importance of the word in the surah, the value (TF/IDF value) increases proportionally to the number of times a word appears in each surah or set of input verses, and is balanced by the number of verses that are in the Surah and contain the word. Finally, it was concluded that by using the contiguity of words and the semantic relations between them, and with the help of text mining techniques, a greater understanding of the vocabulary was obtained, which leads to a more appropriate selection of the equivalent word in the target language.

Keywords: Computational Linguistics, Sociology of Language, Machine Translation, Qur'an, Context Correlation, Finding Equivalents for the Words.

اثربخشی ترجمه ماشینی در فرآیند پردازش زبان؛ بهره گیری از قرینه سیاق در معنایابی واژگان قرآن

زینب شمس^{۱*}، سپیده چهره^۲

۱. دانشجوی دکتری علوم قرآن و حدیث، دانشکده الهیات، دانشگاه کاشان، ایران

۲. کارشناس ارشد هوش مصنوعی دانشگاه علوم تحقیقات تهران، ایران

*نویسنده مسئول مقاله Email: Z.Shams@qom.ac.ir

پذیرش: ۱۴۰۱/۰۵/۱۵

اصلاح: ۱۴۰۰/۰۹/۰۶

دریافت: ۱۴۰۰/۰۷/۰۲

چکیده

به فرآیند برگرداندن مطلبی از زبان مبدأ به زبان مقصد که با یافتن هم ارزهای معناشناختی میان دو زبان صورت می‌گیرد، ترجمه می‌گویند. مهم‌ترین مشکلات ترجمه، ابهاماتی است که در واژگان و ساختار جملات وجود دارند. در یک تقسیم‌بندی، پنج نوع مهم ابهام واژگانی (ابهام‌های مقوله‌ای، واژه‌های هم‌آوا، واژه‌های هم‌نویسه، چند معنایی و ابهام انتقالی) و دو نوع مهم ابهام ساختاری (ابهام‌های ساختاری واقعی و ابهام‌های سیستمی) وجود دارد. ترجمه ماشینی (Machine translation: MT) که بخشی از حوزه پردازش زبان طبیعی (Natural Language Processing: NLP) مبتنی بر کامپیوتر در زبان‌شناسی رایانه‌ای و هوش مصنوعی بوده به عنوان یکی از تکنیک‌های خودکاری است که متن بدون ساختار را به داده‌های ساختاری تبدیل می‌کند با تبدیل متن به اطلاعات، توانسته است تحلیل‌های بیشتری را به داده‌ها اعمال کرده تا اطلاعات مفیدی استخراج شود. در این نوشتار که به روش کتابخانه‌ای تدوین شده، جهت رفع مسائل پیرامون معنای واژگان در ترجمه ماشینی قرآن، طرحی به صورت نظری پیشنهاد شده که هدف آن کمک به فهم بهتر معنای واژگان قرآن، با بهرمندی از قرینه سیاق و بافت عبارت است. در روش پیشنهادی با بهره‌مندی از قاعده سیاق و تکنیک‌های متن کاوی، و با استناد به آن، واژه معادل مناسب‌تری در زبان مقصد برگزیند. در این طرح، سیاق را در مقیاس کلمات دانسته که می‌توان آن را به شرط احراز شرایط به انواع دیگر توسعه داد. به طور خلاصه این طرح دو مرحله دارد: اولویت‌بندی (وزن‌دهی) واژگان هم‌جوار هم ورودی (هر واژه در محدوده آیاتی که در مورد نزول یک‌بار

اثربخشی ترجمه ماشینی در فرآیند پردازش زبان؛ بهره‌گیری از قرینه سیاق در معنایابی...

آن‌ها اتفاق نظر وجود دارد) و سپس مقایسه با کلماتی که اشتراک لفظی (چندمعنا) دارند و نیز مقایسه هم‌نظیران یک واژه با هم‌نظیران سایر واژگان (مترادف‌یابی). می‌باشد. برای دقیق‌تر شدن نتایج می‌توان مشخصات بیشتری از کلمات را به صورت دستی تهیه نمود، جداولی شامل مواردی چون مکئی یا مدنی بود آیات، ترتیب نزول سوره، مفاهیم و تعبیری که در معنای کلمات قرآن در فرهنگ لغاتی چون لسان العرب ابن منظور و فرهنگ لغت راغب اصفهانی آمده است و غیره. برای بدست آوردن داده‌های ورودی از تکنیک‌های نمایه‌سازی استفاده می‌شود. در مرحله پیش پردازش باید داده‌هایی که دارای اهمیت کمتری است (Stop Words) (مانند "الذی"، "التی"، "لم"، "کان"، "کانما" و غیره) حذف شود تا خروجی بهتری بدست آید. برای تغییر شکل داده می‌توان اعراب را حذف کرد تا کدنویسی راحت‌تر انجام شود، برای کاهش نمونه نیز می‌توان از ریشه میانوندی کلمات استفاده نمود. برای اینکه با استناد به قاعده سیاق، برای یکایک کلماتی که به عنوان ورودی مورد پردازش قرار می‌گیرند، رکوردی از مشخصات تهیه نمود، لازم است ابتدا واحد سازی (Tokenizer) صورت گیرد، در داده‌های اولیه تهیه شده، در کل مجموعه آیات ورودی، بر اساس دو معیار قرابت مکانی و فراوانی تکرار، به هر کلمه وزنی اختصاص یابد. هر چه کلمات به کلمه مورد نظر نزدیک‌تر و یا بیشتر تکرار شده باشد، وزن بیشتری به آن اختصاص داده می‌شود که معرف ارتباط معنایی قوی‌تر آنان است و برعکس. طبیعتاً کلماتی که در یک آیه قرار دارند (شماره آیه یکسانی دارند) نسب به کلماتی که در آیات دیگر و فاصله دورتر قرار دارند از ظریب تأثیر بیشتری برخوردار هستند. در سنجش معیار فراوانی: برای نشان دادن اهمیت کلمه در سوره از فراوانی وزنی (TF/IDF Weight) استفاده می‌شود، مقدار (TF/IDF) به تناسب تعداد تکرار کلمه در هر سوره یا مجموعه آیات ورودی، افزایش می‌یابد و توسط تعداد آیاتی که در سوره هستند و شامل کلمه نیز می‌باشند متعادل می‌شود. در نهایت این نتیجه حاصل آمد که از هم‌جواری کلمات و روابط معنایی بین آن‌ها و با کمک تکنیک‌های متن کاوی، فهم بیشتری از واژگان حاصل شده که این مهم‌ترین گزینش مناسب‌تر واژه معادل در زبان مقصد را منجر می‌شود.

واژگان کلیدی: زبان شناسی رایانه‌ای، جامعه شناسی زبان، ترجمه ماشینی، قرآن، قرینه سیاق، معادل یابی واژگان.

۱ مقدمه

ترجمه؛ نقش مهمی در تحوّل و ارتقا فرهنگی انسان‌ها ایفا می‌کند، مردمان اقصی نقاط جهان با زبان‌های گوناگون، از طریق ترجمه با زندگی، تاریخ و فرهنگ یکدیگر آشنا می‌شوند و به تبادل اطلاعات می‌پردازند. ترجمه توسط انسان همواره با مشکلاتی (از قبیل دقت پایین، هزینه و زمان بالا و غیره) مواجه بوده است که این عوامل موجب استقبال از ترجمه‌های ماشینی به رغم کیفیت پایینی که نسبت به ترجمه انسانی دارند شده است. قرآن، این سرچشمه بی‌کران حقایق و معارف، تا کنون توسط انسان‌ها به زبان‌های مختلف ترجمه شده است. ترجمه قرآن توسط انسان، علاوه بر مشکلاتی که ترجمه هر متن با آن مواجه است، به عنوان یک متن مقدس با چالش‌هایی مواجه است، که باید به آن‌ها مسائل خاص ترجمه قرآن نیز افزوده شود. از طریق ترجمه ماشینی قرآن، پیام خداوند که خطاب به تمامی انسان‌هاست، به گونه‌ای علمی به گوش آنان می‌رسد. در سال‌های اخیر که محققین به کاوش متون با استفاده از تکنیک‌های هوش مصنوعی پرداخته‌اند، ترجمه ماشینی قرآن نیز مورد توجه واقع شده است. این حقایق که در حال حاضر قرآن به بسیاری از زبان‌ها ترجمه نشده، بسیاری از ترجمه‌ها مغرضانه بوده، در مواردی متعدّد تشتت عمیقی بین آراء مترجمان وجود دارد و غیره، نیاز به ترجمه ماشینی قرآن را بیش از پیش نمایان می‌سازد، از اینرو با پیش‌بینی آینده روشن در زمینه ترجمه ماشینی قرآن، در این تحقیق ترجمه ماشینی قرآن مد نظر قرار گرفته است.

عبارت "ترجمه ماشینی قرآن" برای بسیاری ناشناخته و غیرممکن است، اینکه منظور از ترجمه ماشینی قرآن چیست؟ و چگونه می‌توان از امتیازاتی که متن قرآن دارد، در فرآیند ترجمه ماشینی قرآن بهره برد؟ سوالاتی است که سعی شده به آن پاسخ داده شود.

در این نوشتار که مطالب آن به روش کتابخانه‌ای تهیه شده، ضمن بیان مطالب اجمالی در مورد ترجمه، ترجمه ماشینی و ترجمه قرآن، به ترجمه ماشینی قرآن پرداخته و با تحلیل یکی از چالش‌های مهم ترجمه یعنی "فهم صحیح‌تر معنای واژگان در زبان مقصد"، که نه تنها در قرآن و نه تنها در ترجمه ماشینی بلکه در ترجمه هر متنی، مترجم با آن مواجه است، طرح پیشنهادی برای بهبود انتخاب واژه هم ارز واژگان قرآنی در زبان مقصد، به صورت نظری ارائه شده است.

در دهه اخیر موضوع ترجمه ماشینی قرآن مورد توجه بسیاری از محققان که اغلب مسلمان هستند قرار گرفته است. در یک نگاه کلی تحقیقات انجام شده در این زمینه به دو دسته قابل تقسیم است: یک) تحقیقاتی که متوجه زبان عربی است و خاص متن قرآن نیست: مانند مقاله "گونه-شناسی چالش‌ها و جایگاه ترجمه ماشینی از عربی به فارسی"، که در سال ۱۳۹۳ش نوشته شده است. نویسنده؛ آقای علی رضا نظری ابتدا با نگاه تحلیلی به جایگاه و میزان رواج ترجمه ماشینی بین دو زبان اشاره داشته و سپس گونه‌ای متعدد چالش ترجمه ماشینی از عربی به فارسی را ارائه داده و نشان می‌دهد که ترجمه ماشینی با نگاه مطلق به آن، چالش ویژه‌ای در زبان عربی دارد که بسیاری از زبان‌های دیگر فاقد آن‌ها هستند از جمله مسئله اعراب، لذا ترجمه ترجمه ماشینی از زبانی چون عربی به زبانی همچون فارسی، شاید از دشوارترین مراحل در تکامل یک ماشین ترجمه جهانی باشد. (دو) تحقیقاتی که خاص ترجمه ماشینی قرآن است: چند نمونه از این تحقیقات که اغلب با بهره‌مندی از تکنیک‌های متن کاوی سعی در بهبود فرآیند ترجمه ماشینی قرآن دارند، به صورت زیر است:

- در مقاله "Computable Difference Matrix for Synonyms in the Holy Quran" که در سال ۲۰۱۴ توسط محمد الشعری و خالد الفیطوری نوشته شده است. نویسندگان برای تشخیص معانی حقیقی واژگان دشوار قرآن، مفاهیم نهفته در هر یک از واژگان مترادف (خواهر) را از فرهنگ لغت‌ها و سایر کتب مرتبط استخراج نموده و بر اساس روابط معنایی که بین آن‌ها وجود دارد، در ماتریسی به نام ماتریس اختلاف (Difference Matrix: DM)، مجموعه‌ای از کلمات کلیدی که در مفهوم آن‌ها موثر بوده‌اند را به عنوان مشخصات طبقه‌بندی، مورد بررسی قرار داده‌اند و به دقت ابعاد اشتراکات و اختلافات آن واژگان را برای یافتن واژه معادل مناسب در ترجمه، بدست آورده‌اند.

- مقاله "An Automatic Evaluation for Online Machine Translation: Holy Quran Case Study" که السوخانی، الکبی و الصمدی، در سال ۲۰۱۶ نوشته‌اند، یک روش ارزیابی اتوماتیک برای ترجمه ماشینی آنلاین رایگان (FOMT) متن قرآن به زبان انگلیسی ارائه شده است. در این روش معیاری به نام ارزیابی مشخصات ضروری متن (Assessment of Text Essential Characteristics: ATEC) سنجیده می‌شود که معیار همبستگی بین

خروجی FOMT و ترجمه انسانی را محاسبه می‌نماید، نتایج نشان می‌دهد مترجم آنلاین گوگل عملکرد بهتری نسبت به مترجم آنلاین بینگ در ترجمه قرآن به انگلیسی دارد.

- "یافتن الگوهای مکرر در قرآن کریم به کمک روش‌های متن کاوی" عنوان مقاله‌ای است که در سال ۱۳۹۷ توسط اصلانی و اسماعیلی نوشته شده است. این مقاله، با دو تغییر در واقع بهبودی بر مقاله "Frequent Pattern Extraction in The Tafseer of al-quran" است که در سال ۲۰۱۴ نوشته شده است، یک تغییر آنکه بر روی زبان اصلی قرآن انجام شده، نه ترجمه مالایی آن، دیگر آنکه بر روی کل آیات قرآن قابل اجراست، نه فقط شش سوره مورد آزمایش. این پژوهش تلاشی برای تهیه الگوهای مکرر قرآن با استفاده از الگوریتم Apriori است، نتایج نشان می‌دهد با اطمینان بالایی می‌توان با استفاده از متن کاوی و یافتن الگوهای پرتکرار قرآن، کلمات و آیات را از حیث معنی، مفهوم و تفسیر بیشتر مورد بررسی قرار داد و از یافته‌های آن در تحقیقات بهره برد.

- مقاله‌ای که در سال ۱۳۹۸ توسط آذری، شکر الهی فر و لسانی فشاری با عنوان "الگوپردازی مفهوم‌یابی از آیات قرآن کریم با استفاده از دانش متن کاوی رایانشی" نوشته شده است. این مقاله با بهره‌مندی فناوریانه از گزاره‌هایی که مستند به ظاهر آیات است، به مفهوم‌یابی آیاتی پرداخته که به موضوع تحقیق و کلید واژه‌های آن مرتبط است، اساس آن بر پایه دو ویژگی: یکی محدود نمودن دایره آیات (اصلی و همجوار) در موضوع، و دیگری محوریت یافتن کلیدواژه‌ها در برداشت تمامی مفاهیم شکل گرفته است.

این پژوهش‌ها تنها پاسخگوی بخش اندکی از چالش‌های پیش‌رو در ترجمه ماشینی قرآن است، توسعه و پیاده‌سازی راهکارهای پیشنهاد شده توسط محققین داخلی و خارجی در بهبود و حل مشکلات ترجمه ماشینی بر روی متن قرآن، عرصه‌ای است که جای رشد بسیار دارد و همت مضاعف متخصصین را می‌طلبد.

۲ ترجمه

به فرآیند یا محصول فرآیند برگرداندن مطلبی از زبان مبدأ به زبان مقصد که با یافتن هم‌ارزهای معناشناختی میان دو زبان صورت می‌گیرد، ترجمه می‌گویند. تاریخ ترجمه به کهنگی تاریخ زبان-

های متنوع، و تاریخ ترجمه مکتوب به کهنگی تاریخ کتابت است (پاکتچی، ۱۳۹۲: ۱۴). در چند سده اخیر مردم و دولت مردان هر سرزمینی بیش از پیش بر حفظ فرهنگ و زبان خودشان تاکید دارند. مردمان سرزمین‌های مختلف سعی دارند در برابر از بین رفتن زبان‌ها و فرهنگ‌ها مقاومت کنند، در چنین دنیایی ترجمه کمک می‌کند تا از دیوارهای بسته اطراف ذهن و محیط عبور کرد و با موضوعات جدید آشنا شد. در واقع ترجمه دربی را باز می‌کند به سمت دنیای افکار و اعتقادات انسان‌های متفاوت با نژادهای رنگارنگ. در یک تقسیم‌بندی کلی، ترجمه به دو صورت مستقیم (که نمودی از تعهد به متن است) (۱) و آزاد (که نمودی از تعهد به معنا و وضوح مطلب است) (۲)، انجام می‌شود (پاکتچی، ۱۳۹۲: ۵۵-۵۶). در زمینه ترجمه تحقیقات بسیاری صورت گرفته و دانشمندانی چون راجر بیکن، مارتین لوتر، جان دریدن و ساموئل جانسون نظریات متفاوتی در مباحث مختلف از جمله غالیتت مولف یا خواننده، دانش و وظایف مترجمان و غیره مطرح نموده‌اند. دسته‌بندی‌های متفاوتی از مشکلات ترجمه شده است. لاینز مشکلاتی که در زمینه متن معادل ترجمه‌ای برای کلمات زبان مبدأ وجود دارد را به سه دسته تقسیم می‌کند:

۱- ممکن است در یک زبان، دو یا چند معنا به کلمات هم‌آوایی مربوط شوند ولی در زبان دیگر چنین نباشد. (مثل کلمه SOUND، که به معانی صدا و موسیقی، حق، قانونی و منطقی، و نیز به طور کامل آمده است).

۲- ممکن است در یک زبان برای مفهوم و معنای به خصوصی، کلمه جود داشته باشد، در حالی که در زبان دیگری فاقد کلمه برای همان مفهوم باشد. مترجم می‌تواند برای این مفاهیم در زبان مقصد، کلماتی ابداع کند و یا کلمات و اصطلاحات زبان مبدأ را به زبان مقصد انتقال دهد، ولی مفاهیمی وجود دارد که مختص فرهنگ زبان مبدأ می‌باشد و حتی با انتقال کلمه‌ی زبان مبدأ، معنا و مفهوم مربوطه، قابل درک نیستند. مانند مفهوم "وضو"، "صیغه" و غیره که قابل ترجمه به زبان انگلیسی نیستند.

۳- ممکن است معانی و مفاهیم کلماتی که از نظر معنانشناختی در زبان‌های مختلف معادل به نظر می‌رسند، دارای مرزهای مشخص نبوده و با هم متجانس و سازگار نباشد (LYONS, 1970: 235).

اساساً مهم‌ترین مشکلات ترجمه، ابهاماتی است که در واژگان و ساختار جملات وجود دارند. در یک تقسیم‌بندی، پنج نوع مهم ابهام واژگانی (ابهام‌های مقوله‌ای، واژه‌های هم‌آوا هم‌نویسه،

واژه‌های هم‌نویسه، چند معنایی و ابهام انتقالی) و دو نوع مهم ابهام ساختاری (ابهام‌های ساختاری واقعی و ابهام‌های سیستمی) وجود دارد. رفع ابهام از معنای کلمات و جملات یک زبان، از سه راه کلی می‌تواند انجام پذیرد که هر راه خود روش‌های متعددی دارد: روش اول ابهام با مرئی یا سرپرست است که مبنای آن مجموعه آموزشی برجسب‌دار است، روش دوم: رفع ابهام بر مبنای منابع لغوی مانند فرهنگ لغت یا فرهنگ تساروس است. و بالاخره روش سوم رفع ابهام بدون مرئی یا سرپرست است که در این حالت تنها مجموعه لغات و متون بدون برجسب در دسترس است (فلاحتی، ۱۳۸۵).

۳ ترجمه ماشینی

توانایی ماشین برای ترجمه یک متن از زبان مبدأ به زبان مقصد را ترجمه ماشینی (Machine translation: MT) می‌گویند که بخشی از حوزه پردازش زبان طبیعی (Natural Language Processing: NLP) مبتنی بر کامپیوتر در زبان‌شناسی رایانه‌ای و هوش مصنوعی است. ترجمه ماشینی یکی از تکنیک‌های خودکار مورد استفاده است که متن بدون ساختار را به داده‌های ساختاری تبدیل می‌کند که یک کامپیوتر می‌تواند پردازش و درک کند. با تبدیل متن به اطلاعات، می‌توانیم تحلیل‌های بیشتری را به داده‌ها اعمال کنیم تا اطلاعات مفیدی را استخراج نماییم. برای کشف اطلاعات مفید از فرآیند داده کاوی استفاده می‌کنیم. داده کاوی به فرایند کلی کشف دانش مفید از داده‌ها گفته می‌شود اصطلاح فرایند کشف دانش در پایگاه داده‌ها شامل آماده کردن داده‌ها، انتخاب داده‌ها، تمیز کردن داده‌ها و درک درست از فرایند داده کاوی است که موجب می‌شود تا اطلاعاتی که برای ما مفید هستند از داده‌ها استخراج شوند. داده کاوی معمولاً به کشف الگوهای مفید از میان داده‌ها اشاره می‌نماید، منظور از الگوی مفید، مدلی در داده‌ها است که ارتباط میان یک زیر مجموعه از داده‌ها را توصیف می‌کند و معتبر، ساده، قابل فهم و جدید است.

اولین بار ویور در سال ۱۹۴۹ میلادی با ارائه روش‌هایی از قبیل استفاده از فنون رمز نویسی در جنگ، تجزیه آماری و غیره، ایده ترجمه ماشینی را مطرح نمود. اولین پروژه در این زمینه نیز در سال ۱۹۵۴ میلادی توسط لنون دوسترت با همکاری شرکت IBM به ثمر رسید، ترجمه‌های ماشینی در این دوره که اولین نسل از ماشین‌های ترجمه محسوب می‌شود، اغلب فرهنگ لغت و تحلیل‌گر

آماری بوده‌اند که گرامری در سطح پایین داشته‌اند. در دهه بعد (نسل دوم)، ماشین‌های مبتنی بر قواعد نوشتاری توسعه یافتند که قادر به ارائه خلاصه‌ای از متن تبدیل یافته نیز بودند. در دهه بعد تحقیقات زیادی صورت گرفت که منجر به ساخت سیستم‌های مبتنی بر موتورال برای ترجمه گزارشات آب و هوایی در رسانه عمومی در سال ۱۹۷۶ میلادی شد و در اواخر این دهه، ترجمه ماشینی وارد بازار تجارت شد. در خلال سال‌های ۱۹۸۰ میلادی طرح مبتنی بر انتقال، با روش‌های جدید نظریه چند زبانی در هم آمیخت و تا اواخر این دهه سیستم‌هایی با قابلیت ترجمه حجم زیادی از متون با کیفیت پایین عرضه شد که اغلب نیازمند پس ویرایش‌های انسانی بودند. دهه ۱۹۹۰ میلادی با پیشرفت‌های چشمگیری در ترجمه ماشینی که همراه با تغییرات اساسی در استراتژی ترجمه بود آغاز شد. در این دهه دیگر زبان؛ به عنوان یک پدیده ساکن با قوانین ثابت تلقی نشد بلکه پیکره‌ای فعال که بر طبق واقعیت‌های فرهنگی اجتماعی تغییر می‌کند، در نظر گرفته شد و نسل سوم ماشین‌های ترجمه به وجود آمد، در این نسل متدهای مبتنی بر نمونه و روش‌های آماری بکار گرفته شدند، همچنین دانش جمع آوری شده محدود به حوزه مشخصی گردید و از بازخورد کاربران ماشینی هم در جمع آوری آماری اطلاعات استفاده می‌شد. در دوره حاضر بهره‌برداری تجاری از ترجمه ماشینی نسل دوم، همچنان مشهود است در حالی که نیاز مبرم به متدهای ارزیابی وجود دارد، در این دوره محققان در پی اهدافی چون ترکیب ماشین‌های ترجمه با سیستم‌های حافظه ترجمه، ماشین‌های ترجمه سخن‌گو، بالابردن کیفیت ترجمه و غیره هستند. امروزه بیشتر شرکت‌های بزرگ از ترجمه ماشینی استفاده می‌نمایند و فروش نرم‌افزارهای آن در بازار رو به افزایش است. این موقعیت سبب ایجاد سرویس‌های ترجمه ماشینی شبکه‌ای متصل به اینترنت مانند آلتاویستا که سرویس‌های پست الکترونیکی سریع، صفحات وب و غیره را به زبان دلخواه ارائه می‌دهند شده است (تدین، ۱۳۹۳: ۲۰-۲۵). روش‌های ترجمه ماشینی در دو دسته کلی روش‌های مبتنی بر متن (۳) و روش‌های مبتنی بر قانون (۴) تقسیم می‌شود (تدین، ۱۳۹۳: ۸-۱۵). در معماری ماشین‌های ترجمه سه جز اصلی که هر جز به نوبه خود از مجموعه اجزایی تشکیل شده‌اند وجود دارد، این اجزا در سه سطح: ۱- واژگانی (تجزیه و تحلیل مرفولوژیکی واژه‌های واحد تحت ترجمه)، ۲- صرفی (وقتی متن به واحدهای پردازش تقطیع شد، تحلیل دستوری آغاز و قواعد صرفی طی چند مرحله اعمال می‌شود) و نحوی (برای یافتن قرینه‌های دستوری، باید

به اطلاعات در تناظر ساخت‌های نحوی دو زبان مبدأ و مقصد دست یافت، شاید مهم‌ترین رکن یک ماشین مترجم تجزیه و تحلیلگر نحوی یا پارسر باشد، و ۳- معنایی، قرار دارد.

در حال حاضر ترجمه‌های ماشینی آنلاین متفاوتی (مانند Google Translate، Bing، Microsoft، Yandex Translate، Online Reverser، و Sistranet) وجود دارند که متن زبان مبدأ را به زبان مقصد بر می‌گردانند و البته هر کدام مزایا و معایب خود را دارند. از مزایای ترجمه-های ماشینی می‌توان این موارد را ذکر نمود: عملکرد آسان و سریع، تشخیص خودکار زبان مبدأ، برگرداندن به زبان‌های مختلف، با صدا و تلفظ‌های مختلف خواندن متون مبدأ و ترجمه شده، ورود متن مبدأ به صورت صوت یا با استفاده از دوربین‌ها خود سایت ترجمه‌گر (گوگل ترانزلیت)، بررسی برخی ترجمه‌ها توسط انجمن ترجمه، امکان کپی برداری، ایمیل و به اشتراک گذاشتن متن ترجمه شده، امکان مقایسه بین ترجمه‌ها، پیشنهاددهی و تصحیح ترجمه‌های نادرست (یاندکس)، ترجمه تصاویری که آپلود شده‌اند (یاندکس)، اسپلینگ یا چک کردن هجای کلمات (ریورسر)، رتبه‌بندی اعتبار ترجمه‌ها به منظور اصلاح (ریورسر) و غیره. از جمله معایب این ترجمه ماشینی نیز تخصصی نبودن، وجود اشتباهات بسیار، روان نبودن جملات، محدودیت تعداد کاراکترها در باکس ورودی، عدم پشتیبانی برخی از زبان‌ها، نیاز به ساخت اکانت (حساب کاربری) جهت ترجمه (سیسترانت) و غیره. روش‌های متفاوتی برای ترجمه ماشینی وجود دارد از جمله: شیوه مبتنی بر معماری مستقیم یا ترانسفورمر، شیوه مبتنی بر انتقال، شیوه میان‌زبانی، شیوه مبتنی بر پیکره‌بندی، یوه مبتنی بر پیکره‌زبانی، روش‌های آماری ترجمه ماشینی، روش‌های مبتنی بر مثال ترجمه ماشینی، سیستم‌های زبان کنترل شده، سیستم‌های تمام خودکار ترجمه ماشینی، ترجمه ماشینی پیوندی (جوادی پور، ۱۳۹۴). برخی از مشکلاتی که در فرآیند ترجمه بیان شد، گلوگاه-های اصلی در ساخت یک ماشین مترجم محسوب می‌شود. چالش‌های پیش رو در ترجمه ماشینی عمدتاً متوجه موانع محاسباتی و کامپیوتری نیست، بلکه زبان شناختی است. مشکلات موجود در این مورد را می‌توان به دو دسته تقسیم کرد: دسته اول مشکلاتی هستند که در هنگام پردازش هر یک از زبان‌های مبدأ و مقصد ممکن است پیش آید، در حالی که مشکلات دسته دوم در هنگام انتقال و ترجمه از زبان مبدأ به مقصد رخ می‌دهند. مشکلاتی مثل تعیین و استخراج ساختار گرامری هر یک از زبان‌های مبدأ و مقصد، تشخیص ریشه کلمات تغییر شکل یافته، تشخیص نقش

یک کلمه در جمله در موارد مبهم، تشخیص زمان فعل در جملاتی که فاقد قید زمان هستند، تشخیص افعال چند قسمتی به خصوص افعال جدایی‌پذیر از جمله مشکلاتی هستند که می‌توان آن‌ها را در دسته اول قرار داد. از مهم‌ترین موارد موجود در دسته دوم می‌توان به متفاوت بودن ساختار گرامری زبان‌های متعدد اشاره کرد، عموماً هر مقدار ساختارهای گرامری زبان‌های مبدأ و مقصد قرابت بیشتری داشته باشند ممکن است نتایج مناسب‌تر و با کیفیت‌تری بدست آیند، علاوه بر مشکلات اشاره شده، گروهی از مشکلات، چه در هنگام پردازش هر یک از زبان‌های مبدأ و مقصد و چه در هنگام انتقال و ترجمه از زبان مبدأ به مقصد وجود دارند، این گروه از مشکلات مربوط به منابع دانشی هستند که وجود آن‌ها برای هر مترجم ماشینی حیاتی است ولی فراهم کردن آن‌ها امری بسیار مشکل است، مثل وجود فرهنگ لغتی غنی برای هر یک از زبان‌ها که ماشین قادر به استخراج اطلاعات مورد نیاز از آن باشد و یا وجود مجموعه متون گسترده به منظور استخراج استنتاج اطلاعات لغوی به خصوص در مترجمه‌های ماشینی مبتنی بر مجموعه متون و یا تجزیه-گرهای مناسب و بسیاری موارد دیگر (تدین، ۱۳۹۳: ۲۹-۳۰). تکنیک‌های موجود در رفع ابهامات در ماشین ترجمه به چهار دسته نظارتی (از تکنیک‌ها و الگوریتم‌های موجود در یادگیری ماشین و متن کاوی استفاده می‌نماید) (۵)، غیرنظارتی (۶)، مبتنی بر فرهنگ لغت (مبنای کار استفاده از منابع لغوی است که توسط ماشین قابل خواندن باشند مانند تساروس، فرهنگ لغات و هستی‌شناسی)، روش‌های ترکیبی و خلاقانه (هر روشی غیر از موارد قبل) تقسیم می‌شوند.

به رغم همه مشکلات، ترجمه ماشینی بنا به دلایلی از جمله سرعت بالا، بدون خستگی و وقفه بودن، مزیت‌های اقتصادی، قدرت ماشین در ذخیره اطلاعات زیاد در حافظه‌اش و تصمیم‌گیری همه‌جانبه بر اساس آن اطلاعات، مورد استقبال و بهره‌برداری روزافزون قرار دارد. گفتنی است سنجش کیفیت ترجمه ماشینی همواره یکی از مسائلی بوده که توجه محققین و توسعه‌دهندگان امر ترجمه ماشینی را به خود واداشته است، چرا که در ترجمه ماشینی، زبان‌های انسانی وجود دارد و با زبان طبیعی نمی‌توان مانند سایر مدل‌های ریاضی و یا تئوری‌های فیزیکی برخورد کرد، زیرا ممکن است برای یک جمله زبان مبدأ چندین ترجمه کامل و قوی در زبان مقصد وجود داشته باشد. واضح است که بهترین داور برای ارزیابی کیفیت یک مترجم ماشینی خود انسان‌ها هستند. جنبه-هایی مثل رعایت شدن قواعد زبانی، رساندن مفهوم درست، کیفیت، سلیسی و روان بودن ترجمه و بسیاری از موارد دیگر تنها توسط انسان مورد سنجش قرار می‌گیرد، اما مشکلات متعدّد سبب شد

متخصصان امر به دنبال روش ارزیابی اتوماتیکی باشند که علاوه بر سریع بودن، مستقل از زبان مبدأ و مقصد بوده و همچنین وابستگی تنگاتنگی با ارزیابی انسانی داشته باشد. از معروف‌ترین این روش‌ها، روشی به نام بلو (BLEU) است (۷).

۴ ترجمه قرآن

از همان صدر اسلام و همزمان با فتوحات و ورود مسلمانان به سرزمین‌هایی که مردم تازه مسلمان غیرعرب داشتند، ترجمه قرآن به عنوان یک ضرورت خودنمایی کرد، با گسترش اسلام، این ضرورت بیشتر و بیشتر شد. البته از همان ابتدا مخالفانی نیز داشت، در اواخر قرن سوم قمری، طبری برگرداندن واژه به واژه قرآن را مصداقی از ترجمه به رأی دانست، ولی همو ترجمه‌ای که همراه تفسیر موجز باشد را پذیرفت و به عنوان راه حل مطرح ساخت (طبری، ۱۴۰۵: ۲۵۳)، که زمینه‌ساز ترجمه در دوره امیر منصور سامانی (مشهور به تفسیر طبری) شد (پاکتچی، ۱۳۹۲: ۱۷). ترجمه قرآن دارای مبانی وجودی، فقهی، قرآنی، ادبی، تفسیری، کلامی، زبان‌شناختی و روش‌شناختی است، ضوابط و قواعدی موثری در اصل اعتبار، استواری و کمال ترجمه قرآن وجود دارد. برای مترجم قرآن نیز شرایط عام و خاصی برشمرده‌اند و طبیعی است که دانش او در فقه، علوم تجربی، ادبیات عرب، تفسیر و واژه‌شناسی در کیفیت ترجمه اثر دارد. سه روش ترجمه قرآن: کلمه به کلمه (تحت اللفظی/حرفی/همگون)، آزاد (تفسیری/خلاصه التفسیر) و جمله به جمله (هسته به هسته/معنوی/امین/محتوا به محتوا) است. سبک‌های ترجمه قرآن نیز بر اساس هدف مترجم، مخاطبین، زبان ترجمه، داشتن یا نداشتن اطلاعات اضافه، گرایش‌ها، اعتبار مترجم یا ترجمه، عناصر هنری، متفاوت است (رضایی اصفهانی، ۱۳۹۱). در ترجمه قرآن چالش‌های متفاوتی وجود دارد که برخی از آن‌ها اختصاصی به متون مقدس ندارد، مانند چند معنایی، برخی دیگر در ترجمه قرآن و سایر کتب مقدس مشترک و برخی خاص قرآن است. از چالش‌های مشترک در ترجمه کتب مقدس: وجود همزمان دو وجه (یکی همگانی بودن پیام برای عموم انسان‌ها و دیگری جوهی که خاص فرهنگ مردم عصر نزول است)، مبنای پذیرش و مشروعیت ترجمه، چالش روزمرگی و متمایز بودن (به منظور نوآوری و شکستن فضای تکرار و خودکارشدگی)، سختی تعهد، رعایت حدتسامح در هم ارزی است (پاکتچی، ۱۳۹۲: ۸۹-۱۲۶). اما برخی مباحث به دلیل ویژگی‌های

تمایزی که باور غالب مسلمانان در طول تاریخ بوده (همچون وحیانی دانستن الفاظ قرآن، اعجاز بیانی، عدم تحریف و غیره)، خاص قرآن است که موجب شده برخلاف کتب مقدس پیشین، قرآن کریم فاقد هرگونه ترجمه رسمی باشد، تا به حال کسی دعوی قدسی بودن ترجمه قرآن را نداشته و هیچ ترجمه‌ای نتوانسته به جای متن عربی، از کاربرد مناسبی برخوردار باشد (پاکتچی، ۱۳۹۲: ۱۳۷). از جمله ابهامات ترجمه قرآن، مسائل هم‌ارزی و روابط معنایی بین واژگان است، که در این بخش به برخی از آن‌ها بیان شده است.

۱-۴. مسائل هم‌ارزی و روابط معنایی واژگان قرآنی

روابط معنایی، کوشش دارد در شبکه معناهای یک زبان، رابطه بین واژه‌ها راه به صورت دو به دو مشخص نماید (Saeed, 2003: 63). در اینجا فرا بردن مبحث روابط معنایی از فضای دورن زبانی به فضای بین‌زبانی موردنظر است که افرادی چون یاکوبسن به آن پرداخته‌اند. از مسائل روابط معنایی در ترجمه قرآن عبارتند از:

- شمول معنایی: این مسئله در معارف اسلامی رابطه عموم و خصوص مطلق خوانده می‌شود. گاه برای واژه قرآنی که زبان مبدأ است به عنوان "شامل"، چند واژه "زیرشمول" در زبان مقصد، وجود دارد. در این حالت مترجم بسته به قرآنی چون بافت سخن یکی از گزینه‌ها را انتخاب می‌کند. برای نمونه واژه "انهار" (شامل) در "وَإِنَّ مِنَ الْجِبَارَةِ لِمَا يُتَّقَرُّ مِنْهُ الْأَنْهَارُ" (بقره/۷۴) یا "وَسَخَّرَ لَكُمْ الْأَنْهَارَ" (ابراهیم/۳۲)، که در فارسی به رود/رودخانه، جوی/جویبار و یا یه وام واژه نهر، برگردانده شده است، یا واژه "شجر" (شامل) در "كَانَهَا كَوْكَبٌ دُرِّيٌّ يُوقَدُ مِنْ شَجَرَةٍ مُبَارَكَةٍ زَيْتُونَةٍ" (نور/۳۵) و "وَشَجَرَةٍ تَخْرُجُ مِنْ طُورِ سَيْنَاءَ تَنْبُتُ بِالذَّهْنِ" (مومنون/۲۰)، که در فارسی به عنوان وام‌واژه وارد نشده و در ترجمه فارسی به زیرشمول درخت یا بوته برگردانده شده است. اما گاه واژه در زبان مقصد، نسبت به واژه عربی شمول معنایی دارد. در این موارد، مترجمان با مشکل انتخاب از میان دو یا چند معنای زیر شمول مواجه نیست، اما در خود متن ترجمه با انتقال از معنای خاص به معنای عام، یک یا چند مولفه از مولفه‌های معنایی واژه مبدأ از دست خواهد رفت و با نوعی کاستی معنایی رخ می‌دهد، به عنوان نمونه واژه‌های "ثور" و "بقرة" که ناظر به جنس مذکر و مونث گاو است و در فارسی واژه گاو، تمایزی از حیث جنسیت ندارد، در این شرایط معمولاً مترجمان با یک صفت (ماده گاو)، این کاستی را جبران می‌نمایند.

- چند معنایی (Polysemy): یعنی یک واژه در معانی متعدد بکار برده شود. در چند معنایی، مترجم در یک زبان با معانی متعدد و لفظ واحد، و در طرف دیگر معانی متعدد با الفاظ متعدد، مواجه می‌شود، به گونه‌ای که در روابط بین‌زبانی، گزینش یک واژه لازم است. چند معنایی اختصاص به متون مقدس ندارد ولی در متون مقدس بر پایه مسائل اعتقادی، چند معنایی از محدوده زبان شناختی فراتر می‌رود و به نوعی شگرد در ارتباط میان خدا و انسان تبدیل می‌شود. چند معنایی در قرآن از دیرباز مطرح بود و برخی مفسران تعبیر آیات مشابه در "هُنَّ أُمُّ الْكِتَابِ وَأُخْرُ مُتَشَابِهَاتٌ" (آل عمران/۷) را همین چند معنایی می‌دانند. در زبان قرآن، "لسان" هم به معنای عضوی از بدن بکاررفته "و لِسَانًا وَ شَفَتَيْنِ" (بلد/۹) معادل Tongue و هم به معنای نظامی از ترکیب واژگان "وَمَا أَرْسَلْنَا مِنْ رَّسُولٍ إِلَّا بِلِسَانٍ قَوْمِهِ" (ابراهیم/۴) معادل Language، "حرام" هم به معنای مقدس "قَوْلٍ وَجْهَكَ شَطْرَ الْمَسْجِدِ الْحَرَامِ" (بقره/۱۴۴) و هم معنای ممنوع "هَذَا حَلَالٌ وَ هَذَا حَرَامٌ لِيَتَفَتَّرُوا عَلَى اللَّهِ الْكُذِبَ" (نحل/۱۱۶) را دارد که مترجمین را در موضع انتخاب قرار می‌دهد. با متمرکز شدن بر روی رابطه میان چند معنایی با اصل زبان و موقعیت تکلم، می‌توان چند معنایی را لحاظ همین دو معیار، بر دو دسته قسم دانست: چندمعنایی برآمده از محدودیت زبانی (مانند واژه "نجم" به عنوان مشترک لفظی که در عربی هم به معنای "ستاره" بوده و هم به معنای "گیاه رویدنی بدون ساقه/بوته"، بکار می‌رود، در "وَالْتَجِمُّ وَالشَّجَرُ يَسْجُدَانِ" (الرحمان/۶)، برخی مانند عاملی آن را به "ستاره" و برخی مانند فولادوند آن را به "بوته" برگردانده‌اند) (۸)، و چند معنایی برآمده از اغراض خداوند (در سطح‌بندی مخاطب یا هر غرض حکیمانه دیگر): (مانند تعلیم اسماء در "وَعَلَّمَ آدَمَ الْأَسْمَاءَ كُلَّهَا" (بقره/۳۱) (۹) (پاکتچی، ۱۳۹۲: ۷۹-۸۲).

- واژگان مترادف: حالت عکس اشتراک لفظی است، دو گونه مترادف در قرآن وجود دارد:

۱- مترادف تام (در درجه بالایی از انطباق با قابیت جابه جایی) (۱۰)، مانند واژه "الفلک" در "فَكَذَّبُوهُ فَأَنْجَيْنَاهُ وَالَّذِينَ مَعَهُ فِي الْفُلْكِ" (اعراف/۶۴) و "السفینه" در "فَأَنْجَيْنَاهُ وَأَصْحَابَ السَّفِينَةِ وَجَعَلْنَاهَا آيَةً لِلْعَالَمِينَ" (عنکبوت/۱۵) و ۲- مترادف متقارب (در بخش مهمی از معنا مشترک هستند اما مولفه متمایز کننده از یکدیگر دارند)، مثلاً واژه "قلب" در "خَتَمَ اللَّهُ عَلَى قُلُوبِهِمْ وَعَلَى سَمْعِهِمْ" (بقره/۷) و "فواد" در "وَ لَا تَقْفُ مَا لَيْسَ لَكَ بِهِ عِلْمٌ إِنَّ السَّمْعَ وَ الْبَصَرَ وَ الْفُؤَادَ" (اسرا/۳۶)، قرآن در مورد "فواد"، تعبیر ثبت (هود/۱۲۰، فرقان/۳۲) و فراغ (قصص/۱۰) و در مورد "قلب" تعبیر اطمینان را

اثربخشی ترجمه ماشینی در فرآیند پردازش زبان؛ بهره‌گیری از قرینه سیاق در معنایابی...

بکار بسته، در حالی که در "قلب" سخن از فراغ نیست. مسائل دیگر چون هم‌ارزی و هویت تاریخی واژه، استعاره و مجاز (۱۱)، تسامح در الزامات زبانی (۱۲)، جابه‌جایی مفهوم و مصداق در هم‌ارزی، واژه‌های کم‌شناخته شده در روابط هم‌ارزی، بستر فرهنگی واژگان، کاهش دامنه معنایی، افزایش دامنه معنایی، برابر انگاری از سر تسامح و غیره از جمله مواردی است که در هم‌ارزی واژگان در ترجمه قرآن مطرح است (۱۳).

۵ ترجمه ماشینی قرآن (QMT)

منظور از ترجمه ماشینی قرآن، ترجمه سیستمی (غیر انسانی) قرآن که توسط موتورهای جستجو و یا نرم‌افزار انجام می‌شود است. ترجمه ماشینی قرآن همانند ترجمه ماشینی بسیاری از متون با مسائلی مواجه هست که بهره‌مندی از تکنیک‌های متن‌کاوی در مرتفع ساختن آن‌ها بسیار کارساز است، متن‌کاوی قرآن نسبت به سایر متون امتیازاتی دارد که این هدف را تسهیل می‌بخشد. مانند روشن بودن ساختار زبان عربی از لحاظ صرف و نحو، وجود ترجمه‌های انسانی به زبان‌های مختلف، منبع داده جامع واحد (نیازی به جمع‌آوری داده‌ها از منابع گوناگون که ممکن است شامل چند پایگاه داده ناهمگن باشد و مشکلاتی چون افزونگی و ناسازگاری داده را به همراه داشته باشد نیست) (۱۴)، بدون تغییر است (تحریف نشده)، لذا کلیه صفات خاصه، کامل است (مقدار دارد) و در مرحله پیش پرداز، مقادیر ناقص وجود ندارد.

از فواید متن‌کاوی قرآن این است که در مجادله‌های علمی بین دانشمندان علوم قرآنی که طرفین با دلایل عقلی و نقلی سعی در اثبات نظر و دفاع از موضع خود دارند، ارائه ادله‌ی علمی و غیرفردی، مورد وفاق همگان خواهد بود، لذا با بهره‌مندی از فناوری‌های نوین در متن‌کاوی قرآن، می‌توان به بسیاری از اختلاف‌ناشی از خطاهای انسانی و یا سوگیری‌های مغرضانه خاتمه بخشید و ترجمه‌های به‌رأی را شناخت و این عوامل موجبات وحدت بیشتر را فراهم خواهد آورد. علاوه بر این با توجه به ماهیت میان‌رشته‌ای این قبیل تحقیقات، بیان شدن آموزه‌های اسلام، از زبان علمی (غیر از زبان دین)، با پذیرش بیشتری مواجه خواهد شد.

۶ طرح پیشنهادی

در مسائل ذکر شده در ترجمه قرآن، مترجم با تکیه بر توانش زبانی خود، بافت و موضوع متن، می‌تواند معادل واژه را برگزیند، اما یک رایانه فاقد چنین توانی است. بنابراین طراح ماشین ترجمه باید خلأ موجود را ترمیم کند. این تمهیدات باید از نخستین مراحل پردازش یعنی از مرحله واژگان و نحو و ساختار آغاز گردد و تا آخرین مرحله یعنی مرحله تجزیه و تحلیل معناشناختی متن تداوم یابد (فلاحی، ۱۳۸۵: ۴۰-۴۲). در طرح پیشنهادی این نوشتار، ضمن باور و حیانی دانستن و عدم تحریف قرآن، با قبول این عقیده که ترتیب آیات در سور قرآن کریم توقیفی (۱۵) است، چینش واژگان در آیات قرآن را مورد بهره‌برداری قرار گرفته است. یکی از اصطلاحات رایج در علوم قرآنی "سیاق" است. سیاق عبارت است کیفیت قرار گرفتن یک واژه در یک جمله، در واقع پیوند خاص میان مفردات یک کلمه و کلمات قبل و بعد، به گونه‌ای که بتوان از آن معنایی کشف کرد که از منطوق و مفهوم آیه آشکارا به دست نمی‌آید، بلکه از لوازم عقلی آن است. سابقه استفاده از سیاق آیات به عنوان یکی از قرائین لفظی (۱۶)، در تفسیر قرآن به زمان صحابه و تابعین برمی‌گردد. استفاده از دلالت سیاق، کمبود لفظ در بیان معانی را تا اندازه‌ای جبران می‌کند. در تقسیم‌بندی سیاق به انواع مختلفی اشاره شده است، که در یک جمع‌بندی می‌توان در پنج دسته قرار گرفته شوند (ایزدی و زندیه، ۱۳۹۱: ۱۹-۲۲):

- ۱- سیاق عام یا اسلوبی (استخراج یک اسلوب از ادبیات قرآن) (رجبی، ۱۳۸۷: ۱۱۳)
- ۲- سیاق سوره (رابطه بین انتهای سوره با ابتدای سوره بعدی)
- ۳- سیاق سوره (مجموعه آیات یک سوره که بر مبنای وحدت موضوعی هر سوره شکل می‌گیرد)
- ۴- سیاق آیات در سوره (مجموعه آیات یک سوره و بر مبنای وحدت موضوعی هر سوره)
- ۵- سیاق کلمات (ساختار معنایی کلی حاکم بر مجموعه‌ای از کلمات درون یک جمله یا آیه) (رجبی، ۱۳۸۷: ۱۱۳).

اساساً دلالت سیاق از مقوله دلالت عقلی و بر اساس التزامات عقلی الفاظ است، بنابراین استفاده از قرائین خارجی مانند شأن نزول و آیات دیگر قرآن که در کنار آیه مورد بحث قرار ندارد و همچنین قواعد شرعی و مانند آن‌ها، دلالت سیاق به شمار نمی‌آید. در زبان‌شناسی جدید مفاهیم

اثربخشی ترجمه ماشینی در فرآیند پردازش زبان؛ بهره‌گیری از قرینه سیاق در معنایابی...

مشابهی مورد استفاده قرار گرفته‌اند که دو تعبیر "بافت یا فرامتن" با مفهوم و معنای فضای غیرلفظی و همچنین تعبیر "بافت متن" از گفتمان یا فضای لفظی، می‌تواند بخشی از تعبیر سیاق باشد (ایزدی و زندیه، ۱۳۹۱: ۱۹). سیاق در عرصه‌های مختلفی از قرآن کارساز بوده است، از جمله در: تبیین واژه‌ها و لغات قرآنی، ترکیب آیات، توسعه و ضیق معنا، پذیرش یا عدم پذیرش سبب نزول، تعیین معنای جمله، مکی یا مدنی بودن آیات، نقد احادیث تفسیری، وقوع مورد آیه به دلالت سیاق، برداشت خاص از آیه، ترجیح و تعیین یکی از قرائات، تعیین مرجع ضمیر، فهم کلمه یا جمله محذوف، مصداق کلمه، ترتیب نزول آیات (ناصر، ۱۳۸۵). متخصص علوم قرآنی برای بهره‌مندی از ساقی، ضوابط بیان داشته‌اند که در صورت تجاوز از آن‌ها حجیت سیاق از اعتبار ساقط می‌شود، از جمله: استفاده از سیاق در آیات قرآنی نباید منجر به تغییر و تبدیل کلام شود، سیاق با ادله دیگر تعارض نداشته باشد، گاهی ممکن است سیاق اقتضایی داشته باشد که با ادله دیگر (از جمله قواعد مسلم عقلی و روایات صحیح) در تعارض باشد. چون حجیت دلالت سیاق از باب حجیت ظواهر است و معلوم است که ظاهر در تعارض با نص، حجیت خود را از دست می‌دهد (۱۷). استناد به سیاق آیات قبل و بعد در صورتی حجت است که مجموع آیات با هم و یکجا نازل شده باشند (۱۸)، دیگر آنکه مطلبی که از سیاق بدست می‌آید، باید مراد متکلم باشد، یعنی مطمئن شود یا حداقل احتمال قوی بدهد که متکلم آن را اراده کرده است. بنابراین احتمالات و مطالب گوناگونی که اخیراً بعضی به عنوان نکته‌ها یا پیام‌های آیات قرآنی عنوان می‌کنند و آن‌ها را بر قرآن تحمیل می‌نمایند، یک نوع استحسان است و نباید آن را از باب سیاق آیات دانست، زیرا برخی از آن‌ها به گونه‌ای هستند که انسان احتمال هم نمی‌دهد که خداوند آن را اراده کرده است. مورد آخر نشان می‌دهد، واژه "سیاق" تعبیر مناسبی برای پژوهش حاضر (روی پایگاه داده قرآن که مولف آن خداست) نیست، ضمن ادغان به این مطلب بیان می‌شود از واژه "سیاق" به منظور رساندن بهتر و سریع‌تر روند این پژوهش به مخاطبان با توجه به ماهیت میان رشته‌ای آن، بهره برده شد است. هدف طرح پیشنهادی این پژوهش بهر مندی از قرینه سیاق در رفع ابهام چند معنایی قرآن به کمک تکنیک‌های متن کاوی است.

۱-۶. پیاده‌سازی نظری طرح پیشنهادی

در روش پیشنهادی این نوشتار با بهره‌مندی از قاعده سیاق و تکنیک‌های متن کاوی، مترجم ماشینی می‌تواند به فهم دقیق‌تری از واژگان قرآنی برسد و با استناد به آن، واژه معادل مناسب‌تری در زبان مقصد برگزیند. در این طرح، سیاق را در مقیاس کلمات دانسته که می‌توان آن را به شرط احراز شرایط به انواع دیگر توسعه داد. از آنجا که در شرط سیاق آمده مجموع آیات باید یک‌جا نازل شده باشد، سور و مجموعه آیاتی به عنوان ورودی مورد پردازش قرار می‌گیرند که در مورد نزول یک‌باره آن‌ها قطعیت وجود دارد. به طور خلاصه این طرح دو مرحله دارد:

- ۱- اولویت‌بندی (وزن‌دهی) واژگان هم‌جوار یک‌ایک واژگان ورودی (هر واژه در محدوده آیاتی که در مورد نزول یک‌باره آن‌ها اتفاق نظر وجود دارد) و
- ۲- مقایسه با کلماتی که اشتراک لفظی دارند (چندمعنا) و نیز مقایسه همسایگان یک واژه با همسایه‌های سایر واژگان (مترادف‌یابی).

مرحله ۱: داده‌های اولیه برای پردازش را تهیه نموده (۱۹)، که به صورت شکل زیر است:

| ردیف | شماره کلمه | کلمه | جزء (نوشتاری) | جزء (اصلی) |
|------|------------|---------------|---------------|------------|
| 1 | 1 | بسم | ب | ب |
| 2 | 1 | بسم | بسم | بسم |
| 3 | 2 | الله | الله | الله |
| 4 | 3 | الرَّحْمَنُ | ال | ال |
| 5 | 3 | الرَّحْمَنُ | رَحْمَن | رَحْمَن |
| 6 | 4 | الرَّحِيمِ | ال | ال |
| 7 | 4 | الرَّحِيمِ | رَحِيم | رَحِيم |
| 8 | 5 | الْحَمْدُ | ال | ال |
| 9 | 5 | الْحَمْدُ | حَمْدُ | حَمْدُ |
| 10 | 6 | لَهُ | ل | ل |
| 11 | 6 | لَهُ | له | له |
| 12 | 7 | رَبِّ | رَبِّ | رَبِّ |
| 13 | 8 | الْعَالَمِينَ | ال | ال |
| 14 | 8 | الْعَالَمِينَ | عَالَمِينَ | عَالَمِينَ |
| 15 | 9 | الرَّحْمَنُ | ال | ال |
| 16 | 9 | الرَّحْمَنُ | رَحْمَن | رَحْمَن |
| 17 | 10 | الرَّحِيمِ | ال | ال |
| 18 | 10 | الرَّحِيمِ | رَحِيم | رَحِيم |
| 19 | 11 | مَالِكِ | مَالِكِ | مَالِكِ |
| 20 | 12 | يَوْمِ | يَوْمِ | يَوْمِ |

شکل ۱. داده‌های اولیه جهت کاوش متن قرآن

این داده‌ها شامل ۶۳۴۸ آیه، ۸۸۴۵۲ کلمه است، علاوه بر متن قرآن که منبع اصلی داده‌های مورد پردازش است، برای دقیق‌تر شدن نتایج می‌توان مشخصات بیشتری از کلمات را به صورت دستی تهیه نمود، جداولی شامل مواردی چون مکی یا مدنی بود آیات، ترتیب نزول سوره، مفاهیم و تعبیری که در معنای کلمات قرآن در فرهنگ لغاتی چون لسان العرب ابن منظور و فرهنگ لغت راغب اصفهانی آمده است و غیره. برای بدست آوردن داده‌های ورودی از تکنیک‌های نمایه‌سازی استفاده می‌شود. به طور ساده نمایه، سیاهه یا لیستی نظام یافته، برای هدایت منظم یک متن، محتوا، مجموعه‌ای از مدارک یا هر گونه اطلاعات ضبط شده است. در مرحله پیش پردازش باید داده‌هایی که دارای اهمیت کمتری است (Stop Words) (مانند "الذی"، "التي"، "لم"، "کان"، "کانما" و غیره) حذف شود تا خروجی بهتری بدست آید. پایگاه داده قرآن منبع جامعی است که داده ناقص ندارد، تنها برای تغییر شکل داده می‌توان اعراب را حذف کرد تا کدنویسی راحت‌تر انجام شود، برای کاهش نمونه نیز می‌توان از ریشه میانوندی کلمات استفاده نمود. برای اینکه با استناد به قاعده سیاق، برای یکایک کلماتی که به عنوان ورودی مورد پردازش قرار می‌گیرند، رکوردی از مشخصات تهیه نمود لازم ابتدا واحد سازی (Tokenizer) صورت گیرد، در داده‌های اولیه تهیه شده، برای هر کلمه شماره آیات مشخص شده است، یا توجه به شماره سوره و شماره آیه‌ای که کلمه در آن واقع شده است به مرکزیت کلمه مورد نظر، ابتدا یک کلمه قبل و یک کلمه بعد، سپس دو کلمه قبل و دو کلمه بعد در صورت وجود بررسی شود و به همین ترتیب در کل مجموعه آیات ورودی، بر اساس دو معیار قرابت مکانی و فراوانی تکرار، به هر کلمه وزنی اختصاص یابد. هر چه کلمات به کلمه مورد نظر نزدیک‌تر و یا بیشتر تکرار شده باشد، وزن بیشتری به آن اختصاص می‌یابد که معرف ارتباط معنایی قوی‌تر آنان است و برعکس. طبیعتاً کلماتی که در یک آیه قرار دارند (شماره آیه یکسانی دارند) نسب به کلماتی که در آیات دیگر و فاصله دورتر قرار دارند از ظریب تأثیر بیشتری برخوردار هستند. به عنوان مثال یک کلمه خاص (x) را در نظر بگیرید، با بهر مندی از قاعده سیاق و تکنیک‌های پردازش متن، این کلمه با (y): یکایک کلمات موجود در سوره، جفت (x, y) را تشکیل داده و میزان همبستگی بین آن‌ها سنجیده می‌شود. وزن "الف" را به کلماتی با بیشترین وزن، وزن "ب" را به کلماتی که بعد از "الف" در بیشترین وزن قرار دارند اختصاص می‌یابد و همین روند چینش ترتیب القبایی با وزن‌های کمتر را تا میزانی قابل تعیین، در نظر بگیرید. به عنوان نمونه کلمه "رَحِيمٌ" در آیه: نَبِيٌّ عِبَادِي أَنِّي أَنَا الْغَفُورُ الرَّحِيمُ وَأَنَّ عَذَابِي هُوَ الْعَذَابُ الْأَلِيمُ

(حجر/۴۹-۵۰) را در نظر بگیرید، با حذف کلمات بازدارند، کلمات مورد پردازش عبارتند از: "نبی"، "عباد"، "غفور"، "رحیم"، "عذاب" و "الیم". این کار برای یکایک کلمات انجام می‌شود و مشخصات حاصل از سیاق (مانند شکل ۲) برای آن‌ها بدست می‌آید.

| | | | | |
|-------|-----|------------|-------------|---------|
| | پ | ب | الف | رَحِيمٌ |
| | نبی | عباد، الیم | غدادب، غفور | |

شکل ۲. نمونه وزن‌دهی کلمه "رحیم" در (حجر/۴۹)

- در سنجش معیار فراوانی: برای نشان دادن اهمیت کلمه خاص در سوره از فراوانی وزنی (TF/IDF Weight) استفاده می‌شود، مقدار TF/IDF به تناسب تعداد تکرار کلمه در هر سوره یا مجموعه آیات ورودی (سند) افزایش می‌یابد و توسط تعداد آیاتی که در سوره هستند و شامل کلمه نیز می‌باشند متعادل می‌شود.

$$TF = \frac{\text{تعداد تکرار کلمه}}{\text{تعداد کل کلمات آیات مجموعه}} \quad IDF = \log \frac{\text{تعداد کل آیات سند}}{\text{تعداد آیات شامل سوره کلمه نظر}}$$

$$TF/IDF = TF \times IDF$$

- در سنجش معیار قرابت: حال که به تک تک کلمات ورودی وزنی اختصاص یافته، نوبت سنجش میزان همبستگی بین کلمات است، برای این کار الگوریتم‌های قوانین انجمنی کارآمد است، معیار پشتیبان و معیار اطمینان می‌تواند به ارزیابی قوانین انجمنی کمک نماید (اسماعیلی، ۱۳۹۲ش). معیار پشتیبان نشان می‌دهد

اثربخشی ترجمه ماشینی در فرآیند پردازش زبان؛ بهره‌گیری از قرینه سیاق در معنایابی...

که در چند درصد از مجموعه سور یا آیات ورودی می‌توان کلمه مورد نظر (X) را با هر یک از کلمات دیگر (Y) پیدا کرد، و معیار اطمینان در میان آیاتی که کلمه (X) را در خود دارند، به دنبال (Y) می‌گردد.

$$\text{Support}(X, Y) = P(X \cap Y)$$

$$\text{Confidence}(X, Y) = P(X | Y) = \frac{P(X \cap Y)}{P(Y)}$$

یکی از معیارهای ارزیابی همبستگی ساده Lift است:

$$\text{Lift} = \frac{\text{Confidence}(X, Y)}{\text{Support}(Y)}$$

معیارهای دیگری نیز جهت بررسی همبستگی کاربرد دارد مانند ϵ -Coefficient، این معیار مقداری بین ۱- (همبستگی منفی کامل) و ۱+ (همبستگی مثبت کامل) است، در صورتی که از همدیگر مستقل باشند، مقدار آن برابر صفر خواهد بود (اسماعیلی، ۱۳۹۲ ش).

مرحله ۲: حوزه یادگیری ماشین در رابطه با مطالعه و بررسی الگوریتم‌ها و تکنیک‌هایی است که این امکان را برای کامپیوتر فراهم می‌نماید تا بر اساس دانش قبلی آموزش ببیند. هدف یادگیری ماشینی این است که کامپیوتر (در کلی‌ترین مفهوم آن) بتواند به تدریج و با افزایش داده‌ها بازدهی بالاتری در وظیفه مورد نظر پیدا نماید. ماشین بردار پشتیبان (Support Vector Machine: SVM) یکی از روش‌های یادگیری با ناظر (Supervised learning) است که از آن برای طبقه‌بندی (Classification) و رگرسیون (Regression) استفاده می‌شود. روش یادگیری طرح پیشنهادی باناظر است، چرا که از پیش نسبت به نتایج آگاهی وجود دارد. در این مرحله یک کلمه مبهم در یک آیه خاص را در نظر بگیرید، بر اساس یافته‌های مرحله قبل دو جنبه قابل بررسی است:

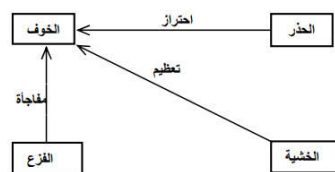
۱- بررسی شود کدام کلمات همسایه‌های مشابهی دارند، به چه میزان مشابهت دارند و وجه تمایزشان در چیست. یعنی از نظر سیاق، در اصطلاح، آن کلمات مترادف هستند. اگر در سطح "الف" وزن مشترکی دارند برود سراغ وزن "ب"، اگر در سطح وزن "ب" اشتراک دارند برود سطح بعدی (در وزن‌دهی شکل ۲)، و به همین ترتیب. هرچه سطوح بالاتر بررسی شود، آن جفت کلمات از نظر قاعده سیاق ارتباط معنوی قوی‌تری با یکدیگر دارند و برعکس. همانند روش مقاله "Computable Difference Matrix for Synonyms in the Holy Quran" می‌توان

مشخص کرد و چه تمایز آن کلمات در چیست و با خوشه‌بندی مشخصات کلمات، مشخص نمود کدام کلمه زیرمجموعه دیگری است.

۲- جستجو شود کلمه X در چه سوره‌هایی تکرار شده است و کلمات همجوارشان با یکدیگر مقایسه شود، یعنی در صورت اشتراک لفظی، پدیده چند معنایی رخ داده است، همسایه-های هرکدام دسته‌بندی شود، همانند روش مقاله "An Automatic Evaluation for Online Quran Case Study Machine Translation: Holy (ردیف‌های جدول شکل ۳) را در نظر گرفته و خصوصیات آن‌ها را از فرهنگ لغات و ترجمه‌های مختلف انسانی گزینش کرده است (ستون‌های جدول شکل ۳)، سپس هریک از آن واژگان را از لحاظ داشتن آن خصوصیت سنجیده است، بعد روابط معنایی بین آن‌ها را از نظر وجوه اشتراک و تمایزشان در داشتن آن خصوصیات و اینکه کدام کلمه از نظر داشتن خصوصیات زیر مجموعه دیگری قرار می‌گیرد، بدست آورده است (شکل ۴).

| الحوف | احتراز | تعظیم | مُفاجأة | عناية | استمرار | استشعار | اضطراب | وقاية |
|-------|--------|-------|---------|-------|---------|---------|--------|-------|
| ✓ | | | | | | | | |
| ✓ | ✓ | | | | | | | |
| ✓ | | ✓ | | | | | | |
| ✓ | | | ✓ | | | | | |
| ✓ | | | | ✓ | | | | |
| ✓ | | | | | ✓ | | | |
| ✓ | | | | | | ✓ | | |
| ✓ | | | ✓ | | | | ✓ | |
| ✓ | | | | | | | | ✓ |
| ✓ | | | ✓ | | | | | |

شکل ۳: سنجش کلمات مترادف از نظر داشتن مشخصات (ATEC)



شکل ۴: ماترس اختلاف بین

اثربخشی ترجمه ماشینی در فرآیند پردازش زبان؛ بهره‌گیری از قرینه سیاق در معنایابی...

جزئیات و فرمول‌های محاسباتی، به دقت در پیاده‌سازی عملی ذکر خواهد شد، در اجرای روش پیشنهادی باید به نکاتی توجه داشت از قبیل اینکه در مورد کلماتی که فقط یک بار در قرآن آمده‌اند (مانند "عَفْرِیْتُ" در نمل: ۳۹، "صَنُكَا" در طه: ۱۲۴، "الرَّقِیْم" در کهف: ۹ و غیره)، فقط بر اساس مشخصات حاصل از سیاق همان واژه می‌توان واژگان مترادف با آن‌ها پیدا نمود.

از فواید این طرح این است که می‌توان آن را برای سایر متون مقدس (نهج البلاغه، صحیفه سجادیه و غیره) بکار بست و حتی می‌توان به صورت تطبیقی و بینامتنی، مفهوم یک واژه را در کتب مختلف مقایسه نمود. باید توجه داشت که شرط حجیت سیاق، عدم تعارض با ادله موثقی مانند روایات صحیح است، لذا یافته‌های حاصل از این پژوهش نیاز به بازبینی توسط متخصصین این حوزه را دارد و به عنوان موید و شاهد قابل بهره‌برداری است.

نتیجه‌گیری

ترجمه ماشینی قرآن علاوه بر مسائلی که هر متن در ترجمه و متن‌کاوی با آن روبروست و این مسائل در متون مقدس به دلیل جنبه‌های وحیانی مضاعف می‌شود، با چالش‌های ویژه‌ای مواجه است، و البته امتیازاتی دارد مانند یکپارچگی منبع داده، عدم تغییر، وجود متن‌هایی کمکی (مانند تفاسیر و ترجمه‌های انسانی)، و غیره که این عوامل راه را برای کاوشگر متن قرآن، هموار می‌سازد. با توجه به استقبال روزافزون جوامع مختلف از ترجمه‌های ماشینی و نیز گسترش اسلام در اقصی نقاط جهان با زبان‌های مختلف، پیش‌بینی می‌شود در آینده نزدیک، ترجمه ماشینی قرآن به عنوان یک ضرورت مطرح شود، لذا عرصه‌ای است که توجه پژوهش‌گران را می‌طلبد چرا که غفلت از آن تبعات ناخوشایندی به همراه دارد. طرح پیشنهاد شده‌ی این نوشتار بهره‌مندی فناورانه از همجواری و تقارب در چینش و اژگان قرآن است، در حقیقت قاعده‌ای عقلایی است که در اصطلاح قرآنی به آن "سیاق" اطلاق می‌شود، در این نوشتار سیاق با تعریف قرینه لفظی متصل مورد نظر است و با توجه به شرایطی که در حجیت آن وجود دارد، تنها در مورد آیات و سوری صادق است که در مورد نزول یک باره آن اتفاق نظر وجود دارد. با فهم بهتر شباهت و تفاوت واژگان و ابعاد مختلف مرتبط به یک واژه، مترجمان زبان‌های مختلف می‌توانند نزدیک‌ترین واژه را در زبان مقصد برگزینند. در چندمعنایی، مسئله هم‌ارزی بین کلمه قرآن و زبان مقصد، پرهیز از کاهش و یا جبران افزایش معنایی و غیره با روش پیشنهاد شده، می‌توان مسیر را هموار نمود. ضمن اذعان به اینکه نتایج

حاصل از اجرای این طرح نیازمند بازبینی متخصصین علوم قرآنی است، بیان می‌شود این نتایج می‌تواند به عنوان شاهد در تأیید مباحث مورد اختلاف صاحب نظران کارساز باشد و یا به عنوان حالت جدید محتمل ارائه شود. ارزیابی دقیق نیاز به پیاده‌سازی عملی و اجرا بر روی نمونه‌ها را دارد. با توجه به کارکردهای سیاق و خروجی کارهای مشابه، پیش بینی می‌شود راهکار موثری در فهم دقیق‌تر واژه‌های قرآن و به تبع آن انتخاب واژه معادل مناسب‌تر در هر زبان مقصدی باشد.

پی‌نوشت‌ها

۱- شامل سه راهبرد: ترجمه تحت اللفظی، گرده برداری (از زبان مبدأ به طور تحت اللفظی، با سازگار ساختن آن با عناصر ساختاری زبان مقصد، به آن زبان انتقال می‌یابد)، وام‌گیری (واژه مبدأ، مستقیم به زبان مقصد انتقال می‌یابد).

۲- شامل چهار راهبرد: جابه جایی (تغییر دادن بخشی از کلام مبدأ که معنا را تحت تأثیر قرار نمی‌دهد)، تغییر مایه (تصرفات موجب بروز برخی تغییرات در ارزش معنایی متن مبدأ می‌شود)، معنایابی (متن مبدأ اساساً به تعبیر دیگری با همان ژرفای معنایی بیان شود)، تطبیق (مرجع فرهنگی تغییر داده شود).

۳- در این روش‌ها قوانین گرامری زبان به عنوان دانش سیستم تعریف شده و فرایند ترجمه بر مبنای آن انجام می‌شود. روش‌های مبتنی بر قانون خود به سه روش کلی: روشهای مستقیم یا مبتنی بر فرهنگ لغت، روش‌های انتقالی، و روش‌های میان‌زبانی، تقسیم می‌شوند.

۴- در این گونه از روشها ماشینهای ترجمه، منابعی شامل مجموعه‌ی گسترده‌ای از متون را به عنوان دانش سیستم ذخیره کرده به منظور ترجمه به کار خواهد برد، روش‌های ترجمه ماشینی مبتنی بر مثال و روش‌های ترجمه ماشینی آماری، نمونه‌ای از این روش‌ها هستند که اساس کار آن-ها مجموعه یا پیکره وسیعی از متون دویا چندزبانه است.

۵- در این روش‌ها مجموعه متونی که به صورت دستی یا خودکار برچسب‌گذاری یا حاشیه نویسی شده‌اند، در یک مدل طبقه‌بندی می‌شوند. به این صورت که مجموعه متون به عنوان یک مجموعه آموزشی در اختیار الگوریتم‌هایی مانند بیز، درخت تصمیم و غیره قرار می‌گیرند تا آموزش

اثربخشی ترجمه ماشینی در فرآیند پردازش زبان؛ بهره‌گیری از قرینه سیاق در معنایی...

بینند و مدل خود را استخراج نمایند که این مدل در یافتن مناسب‌ترین معنا برای کلمه مبهم به مترجم کمک می‌نماید.

۶- در این روش‌ها که بدون نیاز به برچسب گذاری متون است، با استفاده از روش خوشه-بندی، کلمات در خوشه مناسب قرار خواهند گرفت، سپس بررسی می‌شود که کلمه جدید به کدام خوشه نزدیک‌تر است.

7- W. Weaver, Translation. In: Locke W.N. and Booth A.D. (eds.), *Machine translation of languages: fourteen essays* (Cambridge, Mass: Technology Press of the Massachusetts Institute of Technology, 1955), 1949, pp. 15-23

۸- در این موارد مترجم وظیفه دارد معنای متناسب را با توجه به ساختار عبارت و بافت آن گزینش و همان را با استفاده از تعبیرهم ارز در زبان مقصد بازگو نماید.

۹- در این گونه موارد مترجم اول باید بیند تا چه حد فهم آن اغراض برایش دست یافتنی است و تا چه حد موفق خواهد بود در روابط هم ارزی، تعبیری را در زبان مقصد بیابد تا حد ممکن دراصل چند معنایی و حتی میزان چند معنایی با عبارت مبدأ همسانی داشته باشد و عبارت ترجمه نیز بتواند اگر نه همه، بخشی از آن اغراض متکلم را تأمین نماید.

۱۰- برخی از پیروان مکتب ادبی قرآن مانند بنت الشاطی مخالف وجود کلمات مترادف تام در قرآن هستند.

۱۱- مثلاً تردید در تحقق استعاره و مجاز: مانند نظرات متفاوت در مورد لمس زنان در آیه "أَوْ لَا مَسْتَمْتُمُ النِّسَاءَ" (نساء: ۴۳) که موجب اختلاف مذهبی بین مذاهب شده است، مترجمین عملکرد متفاوتی داشته‌اند: برخی از تصریح به معنای لغوی ابایی نداشته، برخی به جنبه کنایی و حفظ عفت تأکید داشته، برخی معنای اولیه را به زبان مقصد انتقال داده و شرایط تحقق استعاره را فراهم آورده‌اند و غیره.

۱۲- تسامح در مقام فهم متن مبدأ: به عنوان نمونه از نظر لغویون "عالمین" در قرآن از جمله در "الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ" (حمد: ۲) به معنای جمع عالم نیست، اما برخی مترجمین در زبان-های مختلف، به رغم تصریح لغویون، برخلاف این مطلب برگردانده‌اند.

۱۳- برای کسب اطلاعات بیشتر به پاکتچی (۱۳۹۲: ۱۷۰-۳۱۵) مراجعه شود.

۱۴- زیرا مآتن آن خداوند واحد و حکیمی است، "أَفَلَا يَتَذَكَّرُونَ الْقُرْآنَ وَلَوْ كَانَ مِنْ عِنْدِ غَيْرِ اللَّهِ لَوَجَدُوا فِيهِ اخْتِلَافًا كَثِيرًا" (نساء: ۸۲).

۱۵- درباره گردآوری و تنظیم سور بین عالمان مسلمان اختلاف نظرهایی وجود دارد، سه دیدگاه ارائه شده در مورد ترتیب سور عبارتند از: توقیفی و مستند به پیامبر (ص)، اجتهادی و مستند به صحابه، در بخشی و حیانی و در بخشی دیگر اجتهادی. منظور از توقیفی بودن سوره‌های قرآن این است که ترتیب سوره‌ها از سوی خدا به پیامبر (ص) وحی شده است، بسیاری از دانشمندان شیعه و سنی چه در گذشته و چه در حال بر جمع قرآن در زمان پیامبر (ص) تأکید داشته‌اند، علامه حسن زاده در کتاب "فصل الخطاب فی عدم تحریف کتاب رب الارباب" دلایل متعددی برای توقیفی بودن سور برشمرده است، برای کسب اطلاعات بیشتر به حسن خرقانی، و حیانی بودن ساختار قرآن، فصل‌نامه پژوهش‌های قرآنی، شماره ۱۹ و ۲۰، ۱۳۷۸ ش، مراجعه شود.

۱۶- مقصود از قراین، اموری است که به نحوی با کلام ارتباط لفظی یا معنوی داشته و در فهم مفاد کلام و درک مراد گوینده مؤثر باشد. قراین به دو دسته تقسیم می‌شوند: ۱. پیوسته به کلام (قراین متصل) و ۲. گسسته از کلام (قراین منفصل). قراین پیوسته بیشتر به قراین لفظی و غیر لفظی تقسیم می‌شود و سیاق، نوعی از قرینه‌ی لفظی متصل در کلام است.

۱۷- مثلاً آیه تطهیر (احزاب: ۳۳) که در ضمن آیات مربوط به همسران پیامبر قرار گرفته است، از نظر سیاق می‌تواند مربوط به آنها باشد، ولی با توجه به روایات مستفیضه اطمینان بخش که شیعه و سنی نقل کرده‌اند که این بخش از آیه درباره اهل بیت پیامبر (ص)، یعنی خود آن حضرت و علی و فاطمه و حسن و حسین (ع) است.

۱۸- اکثر سوره‌های قرآن به طور معمول یکجا و به صورت یک سوره کامل نازل نشده‌اند؛ بلکه آیات آن کم کم به مناسبت‌ها و در زمان‌ها و مکان‌های مختلف و با فاصله زمانی نازل شده تا به صورت یک سوره کامل در آمده است. از این میان فقط چهارده سوره قرآن است که به طور مجموع و یکجا نازل شده و به سور جمعی النزول شهرت یافته‌اند؛ فاتحه، انعام، صف، مرسلات، ضحی، بینه، عادیات، کوثر، کافرون، نصر، تبت، اخلاص، فلق، ناس، برخی دو سوره مائده و توبه را نیز گفته‌اند که به طور مجموع نازل شده‌اند (سیوطی، ۱۴۲۱: ۱۳۶).

منابع و مأخذ

- اسماعیلی، مهدی. (۱۳۹۲). مفاهیم و تکنیک‌های داده کاوی، کاشان: دانشگاه آزاد واحد کاشان.
- ایزدی، مهدی. زندیه، نورالدین. (۱۳۹۱). "سیاق و سباق در مکتب تفسیری علامه. مطالعات قرآن و حدیث"، ۵ (۲)، صص ۳۳-۵.
- پاکتچی، احمد. (۱۳۹۲). ترجمه‌شناسی قرآن کریم؛ رویکردی نظری و کاربردی (مطالعه از سطح واژگان تا ساخت جمله)، چاپ ۳. تهران: دانشگاه امام صادق (ع).
- تدین، محمد علی. (۱۳۹۱). مطالعه مسائل جاری رفع ابهام در ترجمه ماشینی، مبتنی بر مجموعه متون و ارائه راهکار جدید. پایان نامه کارشناسی ارشد، دانشگاه شیراز.
- جوادی پور، مسلم. (۱۳۹۴). "مروری بر ترجمه ماشینی"، دومین همایش الکترونیکی پژوهش‌های نوین در علوم و فناوری.
- رجبی، محمود. (۱۳۸۷). روش تفسیر قرآن، قم: پژوهشگاه حوزه و دانشگاه.
- رضایی اصفهانی، محمد علی. (۱۳۹۱). منطق ترجمه قرآن. چاپ ۲، قم: مرکز بین‌المللی ترجمه و نشر المصطفی.
- سیوطی، عبدالرحمان بن ابی بکر. (۱۴۲۱). الاتقان فی علوم القرآن، چاپ ۲ بیروت: دارالکتب العربی.
- طبری، محمد بن جریر. (۱۴۰۵). التفسیر الجامع (جامع البیان). بیروت: دارالفکر.
- فلاحی، محمد رضا. (۱۳۸۵). "ابهام در ماشین ترجمه"، نشریه کتابداری و اطلاع‌رسانی، دوره ۹، شماره ۳ (مسلسل ۳۵)، صص ۴۰-۴۲.
- ناصر، علی احمد. (۱۳۸۵). "کارکردهای سیاق در تفسیر قرآن کریم"، پژوهش نامه قرآن و حدیث، دوره ۱، شماره ۱، صص ۱۰۷-۱۳۰.

Acknowledgements

We would like to express our thanks to reviewers for their valuable suggestions on an earlier version of this paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

Alshaari, M., Elfitori, K., (2014). "Computable Difference Matrix for Synonyms in the Holy Quran". World Academy of Science, Engineering and Technology International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, Vol. 8, No. 5, Pp. 1401-1404. Doi: 10.1999/1307-6892/9998243

AlSukhni, E. A., Mohammed N. Alsmadi, Izzat M. (2016). "An Automatic Evaluation for Online Machine Translation: Holy Quran Case Study", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 6, Doi: 10.14569/IJACSA.2016.070614

Esmaili, M., (2012). *Concepts and techniques of data mining*. Kashan: Kashan Azad University.

Falahati, M., (2006). "Ambiguity in machine translation", Library and Information Sciences, Vol. 9, No. 3, pp. 21-38.

Izadi, M., & Zandieh, N. (2012). "Context and Spontaneous Occurrence of Meaning in the Exegetic School of Allameh Tabatabaee". Quran and Hadith Studies, Vol. 5, No. 2, pp. 5-23. Doi: 10.30497/quran.2012.1005

Javadipour, M., (2014). *Overview of machine translation*. The second electronic conference of new researches in science and technology.

Lyons, J. (1970). *Semantics*, 2th edition. Cambridge: Cambridge University Press

Naseh, A. A., (2006). "The functions of context in the interpretation of the Holy Quran". *Qur'an and Hadith research*, Vol. 1, No. 1, pp. 107-130.

Paktchi, A., (2012). *Translation of the Holy Quran; A theoretical and practical approach (studying from vocabulary level to sentence construction)*, 3th edition Tehran: Imam Sadegh University (AS).

Rajabi, M., (1967). *The method of interpretation of the Qur'an*. Qom: University and District Research Institute.

Rezaei Esfahani, M. A., (2011). *The logic of Quran translation*, 2th edition, Qom: Al-Mustafa International Translation and Publishing Center.

Saeed, J. I. (2003). *Semantics*, Malden (MA)/Oxford: Blackwell.

Soyuti, A., (2000). *Al-Itqan fi Ulum al-Qur'an*, 2th edition, Beirut: Dar al-Kitab al-Arabi,

Tabari, M., (1984). *Al-Tafsir al-Jame'i (Jame'i al-Bayan)*, Beirut: Dar al-Fekr.

Tadin, M. A., (2011). *Studying the current problems of solving ambiguity in machine translation*, Master thesis, Shiraz University.