



## Tracing the historical origins of machine translation: A bibliometric analysis via RPYS

Ebrahim Ezzati Larsari<sup>1\*</sup>, Ali Akbar Khasseh<sup>2</sup>

1. Assistant Professor, Department of linguistics and foreign languages, Payame Noor University, Tehran, Iran
2. Associate Professor, Department of Knowledge and Information Science, Payame Noor University, Tehran, Iran

### Article Info.

Received: 2023/06/27

Accepted: 2024/04/18

### Abstract

**Background and Objectives:** Among the features that have doubled the importance of the field of machine translation is its interdisciplinary nature; so that the influence of this field from sciences such as mathematics, statistics, probability, natural language processing, formal syntax, corpus linguistics and information technology is inevitable. Each of the aforementioned sciences has played a significant role in the emergence, formation and development of machine translation in some way. For this reason, researchers from different disciplines, each with their own expertise and interest, have turned to research in the field of machine translation and have caused the rapid development of this field. This study, employing a scientometric technique called Referenced Publication Years Spectroscopy (RPYS), concerns analyzing the most important historic works published in the area of Machine Translation.

**Methodology:** This study is conducted by a scientometric approach. Preliminary data of this study have been extracted from Web of Science. For this purpose, the references used in all the papers in Machine Translation since 1945 to the end of 2021 were studied. This search strategy led to 7899 records, covering 193912 references. Using RPYS software named CRExplorer, the revised data were analyzed.

**Findings:** Results exposed that the supreme enterprise of MT was initiated by Shannon in 1948; the one who illuminated the notion of entropy in information technology. After that, Weaver contributed to the appearance of statistical word by word MT in 1949. Moreover, it was Levenshtein (1966) who introduced the revolutionary notion of binary codes capable of correcting deletions, insertions, and reversals in MT. In addition, Dempster et al. (1977) presented a novel statistical EM algorithm concerning maximum likelihood from incomplete data. Later, the works of scholars Brown et al. (1993), Hochreiter (1997), Papineni et al. (2002), and Koehn et al. (2007) have been instrumental in shaping and developing MT.

**Discussion:** Based on the study of influential works in the field of machine translation, the role of mathematics and statistics in the formation of this field is evident. However, after advances in information technology and body linguistics, machine translation translates more accurately and conformingly to the source language.

**Keywords:** Citation analysis, machine translation, scientometrics, Referenced Publication Years Spectroscopy (RPYS)

---

\*Corresponding author    Email: [ebrahimezzati2015@gmail.com](mailto:ebrahimezzati2015@gmail.com)

---

**How to Cite:** Ezzati Larsari, E., & Khasseh, A. A. (2024). Tracing the historical origins of machine translation: A bibliometric analysis via RPYS. *Journal of Studies in Library and Information Science*, 16(2), 37-56.

---

## فصلنامه

مطالعات کتابداری و علم اطلاعات. سال ۱۶، شماره ۲، تابستان ۱۴۰۳



مقاله پژوهشی

## شناسایی ریشه‌های تاریخی حوزه‌ی ترجمه ماشینی: یک تحلیل علم سنجی از طریق آر.پی.وای.اس

ابراهیم عزتی لارسی<sup>۱\*</sup>، علی اکبر خاصه<sup>۲</sup>

۱. استادیار، گروه زبان‌شناسی و زبان‌های خارجی، دانشگاه پیام نور، تهران، ایران

۲. دانشیار، گروه علم اطلاعات و دانش‌شناسی، دانشگاه پیام نور، تهران، ایران

## اطلاعات مقاله

تاریخ پذیرش: ۱۴۰۳/۰۱/۳۰

تاریخ دریافت: ۱۴۰۲/۰۴/۰۶

## چکیده

**هدف:** در این پژوهش، با استفاده از تکنیک علم سنجی به نام طیف سنجی سال انتشار ماخذ، مهم‌ترین آثار تاریخی در حوزه‌ی ترجمه‌ی ماشینی مورد شناسایی و تحلیل قرار گرفته‌اند.

**روش شناسی:** این پژوهش از نوع مطالعات علم سنجی می‌باشد که با استفاده از تکنیک RPYS انجام شده است. جامعه این پژوهش را تعداد ۷۸۹۹ مقاله در حوزه‌ی ترجمه‌ی ماشینی تشکیل می‌دهد که بین سال‌های ۱۹۴۵ تا ۲۰۲۱ در وب آو ساینس نمایه شده‌اند. در راستای هدف پژوهش، ارجاعات استفاده شده در کلیه مقالاتی که در حوزه ترجمه‌ی ماشینی به چاپ رسیده‌اند مورد بررسی قرار گرفت (۱۹۳۹۱۲ ارجاع). پس از استخراج کلیه آثاری که در قسمت فهرست منابع این مقالات درج شده بودند، با استفاده از برنامه نرم‌افزاری CRExplorer نتایج مورد تجزیه و تحلیل قرار گرفت.

**یافته‌ها:** نتایج نشان داد که (1948) Shannon نخستین کسی بود که با معرفی مفهوم آنتروپی در فناوری اطلاعات مقدمات ترجمه‌ی ماشینی را فراهم ساخت. پس از ایشان، ویور در سال ۱۹۴۹ با طرح ترجمه‌ی کلمه به کلمه در ترجمه‌ی ماشینی آماری، این حوزه را متحول ساخت. اما (1966) Levenshtein بود که با معرفی انقلابی "کدهای باینری" باعث گردید که ترجمه‌ی ماشینی قادر به تصحیح، حذف و بازنگری در ترجمه شود. همچنین در سال ۱۹۷۷، دمپستر و همکارانش الگوریتم آماری جدیدی را جهت دستیابی به احتمال حداکثر در داده‌های ناقص ارائه کردند. در خاتمه، آثار دانشمندانی همچون (Hochreiter & Schmidhuber, 1997; Papineni et al., 2002; Koehn et al., 2007) در شکل دهی و توسعه‌ی ترجمه‌ی ماشینی از تاثیرگذاری بسزایی برخوردار بودند.

**نتیجه‌گیری:** براساس بررسی و مطالعه آثار تاثیرگذار در حوزه ترجمه‌ی ماشینی، نقش علم ریاضیات و آمار در شکل‌گیری این حوزه مشهود است. هر چند پس از پیشرفت‌ها در علم فناوری اطلاعات و زبان‌شناسی پیکره‌ای، ترجمه‌ی ماشینی، دقیق‌تر و با صحت بیشتر و متناسب با زبان مبدا به ترجمه می‌پردازد.

**کلیدواژه‌ها:** تحلیل استنادی، ترجمه‌ی ماشینی، علم سنجی، طیف سنجی سال انتشار ماخذ.

ایمیل: ebrahimezzati2015@gmail.com

\*نویسنده مسئول

**استناد به این مقاله:** عزتی لارسی، ابراهیم و خاصه، علی اکبر. (۱۴۰۳). شناسایی ریشه‌های تاریخی حوزه ترجمه ماشینی: یک تحلیل علم سنجی از طریق

آر.پی.وای.اس. مطالعات کتابداری و علم اطلاعات، ۱۶(۲)، ۳۷-۵۶.

## مقدمه و بیان مسئله

ترجمه، نظامی است که با روشی بسیار پیچیده بین جهان زبان مبدأ و جهان زبان مقصد پیوند برقرار می‌کند (Petrilli, 2021). امروزه ترجمه‌ی آنلاین یا خودکار به نام ترجمه‌ی ماشینی، که قدمت آن به دهه‌ی ۱۹۵۰ بازمی‌گردد، به تدریج به منظور تسهیل ارتباطات زبانی بین فرهنگ‌ها و جوامع مختلف طراحی شده است. در واقع، دلیل کاربردی استفاده از ترجمه‌ی ماشینی، ترجمه‌ی سریع داده‌ها و یا اطلاعات از زبانی به زبانی دیگر است، بدون اینکه هزینه‌ای به طور مستقیم برای کاربرانش در اینترنت منظور گردد (Vieira et al., 2021).

از جمله ویژگی‌هایی که اهمیت حوزه ترجمه‌ی ماشینی را دوچندان کرده است، ماهیت میان رشته‌ای آن است؛ به طوری که تأثیرپذیری این حوزه از علوم هم‌چون ریاضیات، آمار، احتمالات، پردازش طبیعی زبان، نحو صورت‌گرا، زبانشناسی پیکره‌ای و فناوری اطلاعات اجتناب‌ناپذیر است. هر کدام از علوم مذکور به نحوی در پیدایش، شکل‌گیری و توسعه ترجمه‌ی ماشینی نقش بسزایی داشته‌اند. به همین دلیل، پژوهشگران از رشته‌های مختلف، هر یک با توجه به تخصص و علاقه‌مندیشان به پژوهش در حوزه‌ی ترجمه‌ی ماشینی روی آورده‌اند و باعث توسعه سریع این حوزه شده‌اند.

## مبانی نظری پژوهش

اکنون که مطالعات و پیشرفت‌های فناورانه در حوزه ترجمه‌ی ماشینی در حال گسترش روزافزون است و قدمت نسبتاً طولانی به عنوان یک حوزه‌ی پژوهشی دارد، انتظار از پژوهشگران این حوزه‌ی میان رشته‌ای این است که کمی به عقب برگشته و با بهره‌گیری از ظرفیت‌های مطالعاتی موجود، ضمن تحلیل نظام‌مند دستاوردهای کسب شده در طول این سال‌ها، نقش و میزان اثرگذاری چهره‌های اثرگذار بر ساختار فکری و روند تکامل این علم پیچیده، با نگاهی عمیق‌تر مورد بررسی قرار گیرد و ماهیت و چگونگی تأثیرگذاری هر یک از علوم شناسایی و نمایان شوند.

نیل به چنین اهدافی را می‌توان با استفاده از فنون علم‌سنجی محقق ساخت. از علم‌سنجی به عنوان یکی از روش‌های کمی یاد می‌شود که به سرعت در حال رشد است. این روش نه تنها برای ارزیابی علم کاربرد دارد، بلکه در راستای اهداف دیگری از قبیل تجزیه و تحلیل تاریخ علم نیز مفید واقع می‌گردد (Marx et al. 2014). از این رو، رویکرد پژوهش حاضر به مقوله‌ی ترجمه‌ی ماشینی، رویکردی است مبتنی بر تاریخ علم که بر تاریخ ترجمه‌ی ماشینی تمرکز دارد و در صدد کشف و شناسایی آن دسته از آثار تاریخی این حوزه است که کماکان در مقالات جدید این حوزه از آن‌ها به کرات یاد می‌شود و سهم بسزایی در تکامل حوزه‌ی ترجمه‌ی ماشینی دارند. به عبارت دیگر، این پژوهش بنا دارد با استفاده از تکنیکی به نام «طیف‌سنجی سال انتشار مآخذ» (Referenced Publication Years Spectroscopy)، از طریق تعمق در سیر مطالعاتی پژوهش‌های صورت گرفته در حوزه‌ی ترجمه‌ی ماشینی، آثار و چهره‌های تأثیرگذار در روند تکامل تاریخی این رشته را از منظر داده‌های کتابشناختی مورد شناسایی قرار دهد.

روشی کمی RPYS را نخستین بار Marx et al. (2014) برای بررسی خاستگاه‌های تاریخی یک حوزه‌ی علمی معرفی کرده‌اند. با استفاده از این روش می‌توان ریشه‌های تاریخی حوزه‌های پژوهشی را نگاشت و تأثیر این خاستگاه‌ها بر پژوهش‌های کنونی را با رویکردی علمی بر شمرد. به بیان ساده‌تر، این تکنیک امکان تحلیل تاریخی براساس داده‌های کتاب‌سنجی را فراهم می‌کند. با این روش، سیر تاریخی را می‌توان به شکل منحنی‌های نوسانی ترسیم کرد که جهش آن‌ها نشان دهنده‌ی سالی است که آثار مهم و پرآستانه در آن سال چاپ شده‌اند. پس از شناسایی این سال‌ها که نقش مهمی در شکل‌گیری حوزه‌ی مربوطه داشته‌اند، باید آثار را به تفکیک هر سال شناسایی و مورد بررسی قرار داد و سهم آن‌ها در شکل‌گیری جهش مورد بررسی را نیز در نظر گرفت (Khasseh & Mokhtarpour, 2016). RPYS به عنوان یک ابزار قدرتمند برای شناسایی خاستگاه‌های تاریخی یک حوزه‌ی پژوهشی معرفی

شده است (Scheidsteger & Haunschild, 2020). به طور خلاصه، با استفاده از RPYS می‌توان الگوهای استنادی را بررسی کرد، آثاری که بیشترین استناد را دارند شناسایی نمود و دلایل استناد آن‌ها را بررسی کرد (Fiala & Bornmann, 2020). RPYS بر این فرض استوار است که آثار پژوهشی مهم در یک حوزه مطالعاتی (در اینجا: ترجمه‌ی ماشینی) اغلب در انتشارات آن حوزه مورد استناد قرار می‌گیرند (Scheidsteger & Haunschild, 2020; Bornmann & Marx, 2014).

با توجه به مطالب فوق، در پژوهش حاضر از تکنیک علم‌سنجی RPYS که رویکردی مبتنی بر تاریخ علم دارد، استفاده شده است تا سیر تکاملی و تاریخی پژوهش‌های صورت گرفته در حوزه‌ی ترجمه‌ی ماشینی شناسایی شود. با توجه به هدف فوق، این پژوهش جستاری است در راستای یافتن پاسخ به پرسش زیر:

- بر اساس تکنیک RPYS، مشخصات مهم‌ترین آثار و جهش‌ها در روند تکاملی حوزه ترجمه‌ی ماشینی چیست و به چه مباحث موضوعی پرداخته‌اند؟

در سالیان اخیر، روش RPYS برای شناسایی انتشارات ریشه‌ای چندین حوزه تحقیقاتی به طور موفقیت‌آمیزی استفاده شده است که برخی از این حوزه‌ها عبارتند از علم‌سنجی (Leydesdorff et al. 2014)، فلسفه علم (Wray & Bornmann, 2014)، رفتار اطلاعاتی (Soheili & Khasseh, 2015)، مدیریت دانش (Khasseh & Mokhtarpour, 2016)، روش پژوهش (Mokhtarpour & Mousavi, 2018)، اقتصاد اکولوژیک (Ballandonne, 2019)، کارآفرینی (Habibi et al., 2018)، خدمات مرجع (Mousavi, 2018)، برابری بهداشتی (Yao et al., 2019)، کاردرمانی (Khasseh et al. 2019)، و ارگونومی (Heidarimoghadam et al. 2021). از طرفی، هیچ تحقیقی به طور صریح به مراجع ذکر شده در ترجمه‌ی ماشینی به منظور کشف ریشه‌های تاریخی آن پرداخته است، که دغدغه‌ی تحلیل حاضر است. به بیان دقیق‌تر، مطالعه‌ی حاضر به شناسایی و تحلیل آثاری می‌پردازد که تا حد زیادی به تکامل و توسعه‌ی ترجمه‌ی ماشینی کمک کرده است. با توجه به کنکاش نویسندگان، این مطالعه در طیف ترجمه‌ی ماشینی منحصر به فرد و جدید است که با استفاده پیشگامانه از روش RPYS به بررسی آثار اساسی در این زمینه پرداخته است.

البته، طبق بررسی‌ها، تاکنون محدود مطالعاتی با استفاده از فنون علم‌سنجی اقدام به بررسی پژوهش‌های حوزه ترجمه‌ی ماشینی نموده‌اند. به بیان دقیق‌تر، به نظر می‌رسد استفاده از شاخص‌های علم‌سنجی در مطالعه فعالیت‌های پژوهشی بویژه در حوزه‌ی ترجمه‌ی ماشینی به دهه‌های اخیر بازمی‌گردد، زمانیکه جمع‌آوری اطلاعات از طریق پردازش داده‌ها امکان‌پذیر گردید. بر این اساس، با توجه به اهمیت ترجمه‌ی ماشینی در ارتباطات دنیای معاصر، محققان قصد دارند تحلیلی علم‌سنجی از این پدیده‌ی جدید فناورانه به منظور تبلور ابعاد علمی متنوع و ترسیم دامنه‌ی تحقیقات پیش‌بینی‌کننده آن در آینده، انجام دهند؛ نکته‌ای که توسط سایر محققین، هرچند نه به تعداد زیاد، به شرح زیر تأکید شده است:

اولین و مهمترین تحقیق در مورد تجزیه و تحلیل علم‌سنجی مطالعات ترجمه‌ی ماشینی توسط Gupta & Dhawan (2019) انجام شده است. آن‌ها در مطالعه خود گزارش کمی و کیفی از مطالعات ترجمه‌ی ماشینی نمایه شده در پایگاه اسکوپوس از سال ۲۰۰۷ تا ۲۰۱۶ ارائه کرده‌اند. در واقع، آن‌ها تحلیل علم‌سنجی را براساس زنجیره‌ای از رویه‌ها، مانند نرخ رشد انتشارات، سهم جهانی، تأثیر استناد، سهم مقالات بین‌المللی مشارکتی و توزیع نشریات براساس حوزه‌های فرعی ترسیم کردند. همچنین کشورها، سازمان‌ها و نویسندگان برتر را در تحقیقات ترجمه‌ی ماشینی براساس برخی از شاخص‌های کتاب‌سنجی شناسایی نمودند. در نهایت، پژوهش آنان ویژگی‌های مقالات پراستناد را در ترجمه‌ی ماشینی ارائه کرد (Gupta & Dhawan, 2019). در پژوهش دیگری، Gile (2015) یک بررسی درباره‌ی پیشینه مطالعات ترجمه با استفاده از روش‌های علم‌سنجی برای آشکار کردن و توضیح تاریخ و تحولات اخیر آن انجام داد و تلاش کرد تا مطالعات ترجمه را از دیدگاه CIRIN (شبکه اطلاعات پژوهشی تفسیر کنفرانس) با تحلیل استنادی توصیف کند. او برخی از مسائل فنی را در مورد استفاده از تحلیل علم‌سنجی برای تعداد محدودی از انتشارات که در آن‌ها تجزیه و تحلیل

کتابشناختی کافی وجود نداشت، مورد بحث و بررسی قرار داد. آخرین نکته تحقیق او بر نیاز ضروری آموزش علم‌سنجی برای پژوهشگران جوان متمرکز بود (Gile, 2015). پس از آن، Dew et al. (2018) مروری نظام‌مند درباره‌ی توسعه‌ی فناوری ترجمه‌ی ماشینی مربوط به ارتباطات سلامت در بازه‌ی زمانی ۲۰۰۶ و ۲۰۱۶ انجام دادند. به بیان دیگر، براساس متغیرهای مختلف مانند تنظیمات برنامه، کاربران هدف، فناوری اساسی، اینکه آیا ترجمه‌ی ماشینی به صورت مجزا یا در ترکیب با ویرایش انسانی استفاده شده است، زبان‌های آزمایش شده، روش‌های ارزیابی، یافته‌ها، و شناسایی شکافهای پژوهش، مشخص شد که اکثر مطالعات از طیف وسیعی از روشهای ترجمه‌ی ماشینی انسانی و خودکار استفاده می‌کنند. متعاقباً، سیستم‌های ترجمه‌ی ماشینی آماری دقیق‌تر از سیستم‌های مبتنی بر قانون در مجموعه‌های بزرگ بودند. علاوه بر این، در بیشتر سیستم‌ها، بهترین کارایی برای ترجمه‌ی جملات ساده از انگلیسی به زبان‌های اروپای غربی بود. در نهایت به این نتیجه رسیدند که جدای از نقش ترجمه‌ی ماشینی در ارتقای ارتباطات چندزبانه در زمینه‌های سلامت، هنوز یک نکته نگران‌کننده در مورد دقت و ارزیابی استاندارد ترجمه‌های ماشینی در ترجمه‌ی متون رشته‌های بهداشت و پزشکی وجود دارد (Dew et al., 2018).

### روش‌شناسی پژوهش

این پژوهش از نوع کاربردی بوده و با استفاده از یکی از فنون علم‌سنجی به نام RPYS انجام شده است. جامعه این پژوهش را تعداد ۷۸۹۹ مقاله در حوزه ترجمه‌ی ماشینی تشکیل می‌دهد که بین سال‌های ۱۹۴۵ تا ۲۰۲۱ در وب آو ساینس نمایه شده‌اند. به منظور بازیابی مقالات حوزه ترجمه‌ی ماشینی از راهبرد جستجوی زیر استفاده شده است:

Topic: "machine translation" OR "automatic translation" OR "automated translation" OR "online translation"

Time Span: 1945-2021

Doc Types: All

Web of Science Core Collection: Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Emerging Sources Citation Index (ESCI), Book Citation Index – Social Sciences & Humanities (BKCI-SSH), Book Citation Index – Science (BKCI-S), Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH), Conference Proceedings Citation Index – Science (CPCI-S)

جستجوی موضوعی فوق در WoS که در ۱۲ نوامبر ۲۰۲۱ انجام شد، نشان داد که نخستین مقاله در حوزه ترجمه‌ی ماشینی در سال ۱۹۵۶ در این نمایه نامه ثبت شده است. پس از استفاده از راهبرد فوق، تمامی مقالات (۷۸۹۹ مقاله) و مراجع ذکر شده آن‌ها (۱۹۳۹۱۲ مورد) از پایگاه داده WoS استخراج و با استفاده از نرم‌افزار مربوطه تجزیه و تحلیل شد.

تاکنون چندین نرم‌افزار برای انجام تحلیل‌های RPYS طراحی شده‌اند. از نخستین برنامه‌های نرم‌افزاری مختص این روش می‌توان به RPYS اشاره کرد که در سال ۲۰۱۴ توسط «لیدسدورف» طراحی شد. سپس در سال ۲۰۱۶ نرم‌افزار دیگری به نام RPYS i/o (http://www.leydesdorff.net/comins/rpys/index.html) توسط Comins & Leydesdorff (2016) معرفی شد. (http://andreas-thor.github.io/cre/#) CRExplorer نیز نرم‌افزار دیگری است که بدین منظور طراحی شده است و قابلیت‌های بیشتری نسبت به دو نرم‌افزار قبلی ارائه می‌دهد. به همین دلیل، در این پژوهش از برنامه نرم‌افزاری CRExplorer (http://andreas-thor.github.io/cre) برای اعمال RPYS استفاده شد. بدین منظور، فایل‌های دانلود شده برای پردازش بیشتر در CRExplorer وارد شدند. سپس از ویژگی‌های خوشه‌بندی و ادغام CRExplorer برای پاکسازی مجموعه داده‌های مراجع استناد شده با حذف موارد تکراری یا انواع مرجع همان نشریه استناد شده استفاده شد. واردات داده‌ها به RPYS در دوره‌ی بین ۱۹۰۰ و ۲۰۲۰ محدود شد تا بهتر بتوان بر خاستگاه‌های تاریخی این دوره متمرکز شد. براساس تکنیک RPYS، قله‌ها و

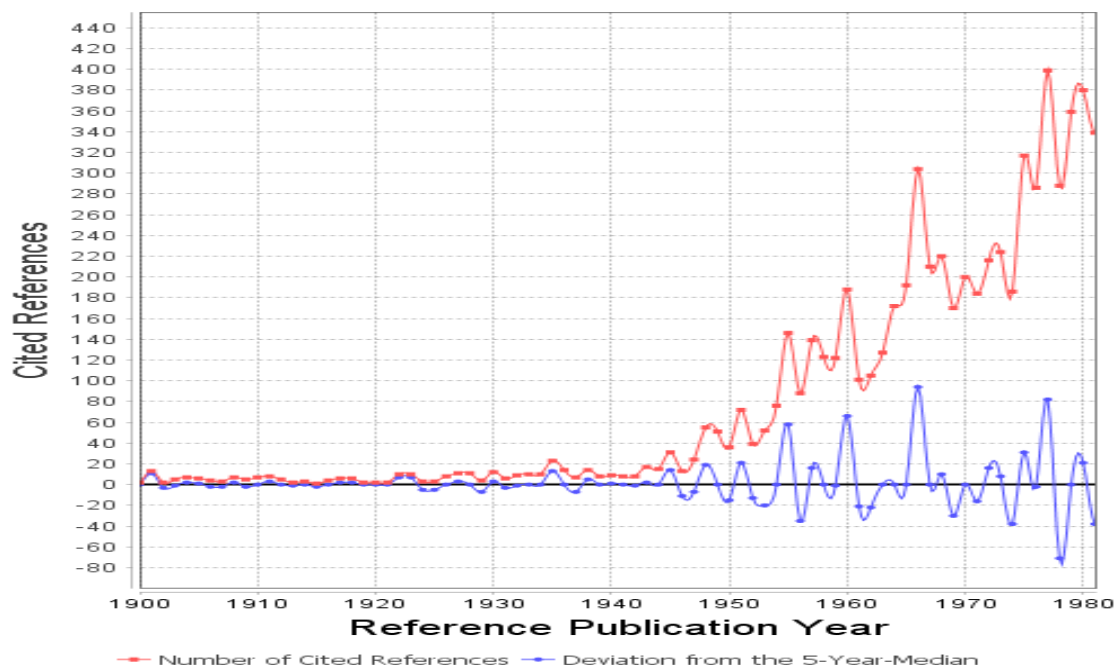
جهش‌های مندرج بر نمودار مبین این نکته هستند که در سال مربوطه اثر/آثار مهمی در آن زمینه‌ی تحت مطالعه چاپ شده‌اند. به بیان دیگر، اوج‌ها در RPY های اولیه به ریشه‌های تاریخی حوزه‌ی تحت مطالعه اشاره دارد: آثاری که پس از انتشارشان، در سال‌های بعد به کرات مورد استناد قرار گرفته‌اند و جزو آثار شاخص و تاثیرگذار آن حوزه‌ی پژوهشی تلقی می‌شوند.

### یافته‌های پژوهش

به منظور بررسی منابع استناد شده (CRs) به شیوه‌ای دقیق‌تر، RPYS به دو دوره به شرح زیر تقسیم شد: ۱۹۸۰-۱۹۰۰ و ۱۹۸۱-۲۰۲۰. همچنین شایان ذکر است که در تحقیق حاضر به دلیل کم بودن تعداد استنادات قبل از دوره ۱۹۰۰، مطالعات بعد از این سال مورد توجه بوده است. طبق تجزیه و تحلیل داده‌ها، بین سال‌های ۱۹۰۰ تا ۱۹۸۰، ۱۰ نقطه‌ی اوج شامل ۱۲ اثر تاثیرگذار در ترجمه‌ی ماشینی شناسایی شد. همچنین، در بازه زمانی ۱۹۸۱ تا ۲۰۲۰، ۱۰ نقطه‌ی اوج شامل ۱۳ اثر تاثیرگذار یافت شد. شرح مفصل این دو بازه‌ی زمانی در ادامه ارائه شده است.

### - بازه زمانی اول: از سال ۱۹۰۰ تا ۱۹۸۰

نمودار ۱ طیف‌سنجی سال انتشار مآخذ در حوزه ترجمه‌ی ماشینی بین سال‌های ۱۹۰۰ تا ۱۹۸۰ را نشان می‌دهد. RPYS بر این اصل استوار است که نقطه‌های اوج اشاره به انتشارات مهم در یک زمینه تحقیقاتی دارند. اوج‌ها در سال‌های موجود در نمودار به ریشه‌های تاریخی این رشته اشاره می‌کنند (Fiala & Bornmann, 2020). براساس این نمودار، بین سال‌های مذکور ده جهش رخ داده است که به ترتیب در سال‌های ۱۹۴۸، ۱۹۴۹، ۱۹۵۱، ۱۹۵۵، ۱۹۵۷، ۱۹۶۰، ۱۹۶۶، ۱۹۷۵، ۱۹۷۷، و ۱۹۸۰ رخ داده است. بنابراین، به احتمال فراوان در این سال‌ها رخداد مهمی در شکل‌گیری مباحث MT روی داده است. به عبارتی، این آثار تاثیرگذاری بیشتری در این حوزه داشته‌اند که بعدها به کرات بدان استناد شده است. به همین دلیل آثاری که در این سال‌ها چاپ شده‌اند مورد بحث قرار می‌گیرند تا تأثیر آن بر حوزه MT مشخص شود.



نمودار ۱. نتایج RPYS در حوزه ترجمه‌ی ماشینی (بازه زمانی ۱۹۸۰-۱۹۰۰)  
 Chart 1: RPYS results in the field of machine translation (period 1980-1900)

با نگاهی به نمودار طیفی فوق، علیرغم اینکه نوساناتی در مطالعات ترجمه‌ی ماشینی بین سال‌های ۱۹۲۰ و ۱۹۴۵ وجود داشت، بدیهی است که آثار پژوهشی از سال ۱۹۰۰ تا حدود ۱۹۴۸ الگوی مشابهی را دنبال کرده‌اند. از سال ۱۹۴۸ تا ۱۹۶۰، مطالعات ترجمه‌ی ماشینی نوسانی ۱۲ ساله را تجربه کرد. با این حال، هر چه بیشتر به ربع سوم قرن می‌رسد، این نوسان به‌طور نامنظم و با نرخ فزاینده‌ای شدیدتر می‌شود که در آن تعداد استنادها از ۱۴۶ به حدود ۳۸۰ مورد در طول دوره ۱۹۶۰ تا ۱۹۸۰ افزایش می‌یابد. برای توضیح دقیق‌تر این موضوع، جدول ۱ آثار پر استناد را در بازه زمانی ۱۹۰۰ تا ۱۹۸۰ نشان می‌دهد.

**جدول ۱. جهش‌های تاریخی حوزه‌ی ترجمه‌ی ماشینی به همراه پراچ‌ترین آثار در بازه زمانی ۱۹۰۰ تا ۱۹۸۰**

*Table 1. Historical leaps in the field of machine translation along with the most referenced works in the period from 1900 to 1980*

RPY	Cited Reference	Cited References Count	No. of all citations of the year	Document Type
1948	Shannon, C. E. (1948). A mathematical theory of communication. <i>The Bell system technical journal</i> , 27(3), 379-423.	23	55	Journal Paper
1949	Weaver, W. (1949). <i>Machine Translation of Languages</i> , MIT Press, Cambridge, MA, Reprinted from a memorandum.	17	51	Book
1951	Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. <i>The annals of mathematical statistics</i> , 22(1), 79-86.	17	72	Journal Paper
1956	Locke, W. N., & Booth, A. D. (1956). Machine translation of languages. <i>American Documentation (pre-1986)</i> , 7(2), 135.	26	146	Journal Paper
1957	Chomsky, Noam (1957), <i>Syntactic Structures</i> , The Hague/Paris: Mouton,	28	139	Book
1960	Cohen, J. (1960). A coefficient of agreement for nominal scales. <i>Educational and psychological measurement</i> , 20(1), 37-46.	38	188	Journal Paper
	Bar-Hillel, Y. (1960). The present status of automatic translation of languages. <i>Advances in computers</i> , 1, 91-163.	29		Journal Paper
1966	Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions, and reversals. In <i>Soviet Physics Doklady</i> (Vol. 10, No. 8, pp. 707-710).	112	304	Journal Paper
1975	Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. <i>Communications of the ACM</i> , 18(11), 613-620.	26	317	Journal Paper
1977	Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. <i>Journal of the Royal Statistical Society: Series B (Methodological)</i> , 39(1), 1-22.	95	399	Journal Paper
	Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. <i>biometrics</i> , 159-174.	38		Journal Paper
1980	Porter, M.F. (1980). An algorithm for suffix stripping. <i>Program: electronic library and information systems</i> , Vol. 14 No. 3, pp. 130-137.	56	380	Journal Paper

جدول ۱، ۱۰ جهش شامل ۱۲ اثر اساسی منتشر شده در زمینه مطالعات ترجمه‌ی ماشینی در طول ۸۰ سال بین سال‌های ۱۹۰۰ و ۱۹۸۰ را مقایسه می‌کند. این سال‌ها عبارتند از: ۱۹۴۸، ۱۹۴۹، ۱۹۵۱، ۱۹۵۶، ۱۹۵۷، ۱۹۶۰، به ترتیب ۱۹۶۶، ۱۹۷۵، ۱۹۷۷ و ۱۹۸۰ در سال ۱۹۴۸، حدود ۵۵ استناد وجود داشت که از این تعداد، ۲۳ مورد، تقریباً نیمی از استنادها، به مقاله‌ای با عنوان: "نظریه ریاضی ارتباطات" که نویسنده‌ی آن شانون بود، اختصاص یافت، با توجه به اینکه مقاله‌ی او، اولین و قابل توجه‌ترین مطالعه بود، به طور قابل تحسینی بیشترین تعداد استناد را در قرن بیستم به دست آورد. بیان این نکته کاملاً ضروری



است که ابتدا و مهمترین تلاش در زمینه‌ی نظریه‌ی اطلاعات و متعاقباً ترجمه‌ی ماشینی توسط کلودای. شانون پیشنهاد شد. مقاله او به طور حتم مورد موافقت بسیاری از دانشمندان قرار گرفت که پایه‌گذار و سازنده‌ی نظریه‌ی اطلاعات گردید (Shannon, 1948). یک سال بعد این مقاله به صورت کتابی با عنوان «نظریه ریاضی ارتباطات» منتشر شد. در این کتاب، ویور با ارائه توصیفی جامع و شفاف از نظریه شانون به گسترش و توسعه نظریه‌ی ریاضی ارتباطات پرداخت. در واقع، ویور اصطلاحات فنی شانون را برای عموم مخاطبان قابل درک تر توضیح داد. به طور مشخص، برجسته‌ترین مفهوم در کار شانون، معرفی یک واحد اندازه‌گیری (آنترپی) (Entropy) برای نرخ انتقال اطلاعات بود که بعدها توسط جان توکی (John Tukey) (Saha & Jose, 2020) به عنوان «بیت» (Bit) نام‌گذاری شد. شانون همچنین اکتشافات خود را با همکاری رابرت فانو (Robert Fano) توسعه داد تا سیستمی را برای کدگذاری معرفی کند که بعدها اساس کدگذاری در فناوری اطلاعات گردید و منجر به تبدیل اطلاعات متناسب با زبان ماشین‌ها شد (Manz, 2021).

تنها یک سال بعد، Weaver (1949) کتاب «ترجمه‌ی ماشینی زبان‌ها» را نوشت که حدود ۱۷ بار از مجموع ۵۱ مورد بدان استناد شد. این کتاب اساساً تقویت ایده‌های ویور بود که در نامه‌ای به نوربرت وینر (Norbert Wiener)، متخصص سایبرنتیک نوشته شده بود. او در آن نامه در سال ۱۹۴۷ مفهوم ترجمه‌ی متون از زبانی به زبان دیگر را با کمک رایانه مطرح کرد. پس از آن، ایده‌های ارائه شده را در یادداشتی تنظیم کرد که بعداً در سال ۱۹۴۹ در قالب کتاب تجدید چاپ شد. همان‌طور که قبلاً ذکر شد، به طور اساسی، این کتاب ضروری‌ترین و تأثیرگذارترین اثر در رشد اولیه ترجمه‌ی ماشینی، به ویژه رویکرد ترجمه‌ی ماشینی آماری (Statistical Machine Translation (SMT)) بود. برجسته‌ترین دیدگاه‌های این کتاب در ارتباط با ارائه طرح‌های دقیق و کارآمد برای ترجمه‌ی ماشینی به صورت چهار پیشنهاد است: اولین مورد این بود که چگونه معادل مناسب یک کلمه را براساس کلمات موجود در مجاورت کلمه مورد ترجمه تشخیص دهیم و به اصطلاح معادل سازگار با آن را پیدا کنیم. البته پیشتر ویور بیان کرد که معنای کلمات را می‌توان به صورت مجموعه‌ای از شبکه‌های عصبی (Neural Network) در نظر گرفت، ایده‌ای که از مک کالوچ (MacCulloch) و پیتس (Pitts) الهام گرفته شده بود (Chakraverty et al., 2019). در پیشنهاد دوم، ترجمه را می‌توان به عنوان یک "استنتاج" (Deduction) در زبان مقصد از "مقدمات" (Premises) در زبان مبدأ تعریف کرد، استنتاجی که براساس احتمال و پیش‌بینی‌پذیری در ترجمه‌ی ماشینی معرفی شده است. پیشنهاد سوم، کاربرد تاکتیک‌های رمزنگاری در ترجمه‌ی ماشینی است. به عنوان مثال، برای ترجمه از زبان A به زبان B، زبان A می‌تواند به یک متن ساده رمزگذاری شده از زبان B تبدیل شود، که اساس معرفی مفهوم رمزگذار و رمزگشا در ماشین‌های ترجمه‌ی عصبی است. آخرین پیشنهاد مربوط به بهره‌برداری از ایده جهانی‌های زبانی در ترجمه‌ی ماشینی بود که زیربنای همه زبان‌های انسانی است (Umeozor, 2020). به عنوان مثال، Reifler (1950) پیشنهاد ویور را پذیرفت که محصولی براساس ترجمه‌ی کلمه به کلمه، ممکن است نادرستی‌های زیادی ایجاد کند. بر این اساس، ایده‌ی پیاده‌سازی یک پیش-ویرایشگر انسانی متخصص در زبان مقصد را برای بررسی و تصحیح خطاهای دستوری، واژگانی و منطقی ارائه کرد (Reifler, 1950). علاوه بر این، یک پسا-ویراستار را نیز پیشنهاد کرد که در زبان مبدأ مهارت داشته باشد تا منطقی بودن و همچنین طبیعی بودن متن ترجمه شده را در مقایسه با متن اصلی آن ارزیابی کند (Hilbert, 2021).

دو سال بعد، Kullback & Leibler (1951) مقاله‌ای منتشر کردند تحت عنوان «درباره اطلاعات و کفایت» (On Information and Sufficiency) که توانست حدود ۱۷ استناد را از ۷۲ مورد در سال ۱۹۵۱ به دست آورد. در واقع، آنترپی شانون را در سیستم‌های اطلاعاتی با معیار  $(D_{KL}(Q||P))$  برای میانگین اختلاف تعداد بیت‌های مورد نیاز برای یک فرآیند رمزگذاری مطرح ساختند (Kullback & Leibler, 1951). در سال ۱۹۵۶، لاک و بوث "ترجمه‌ی ماشینی زبان‌ها" را منتشر کردند. این مقاله ۲۶ استناد از ۱۴۶ مورد را دریافت کرد. در مرحله بعد، مشکلات زبانی مرتبط با ترجمه خودکار مانند شناسایی معادل مناسب برای همخط‌ها، افعال عبارتی، کلمات مرکب و غیره مثال زده و توضیح داده شد (Locke & Booth, 1956).

در سال ۱۹۵۷، نوام چامسکی کتابی به نام "ساختارهای نحوی" نوشت که برای معرفی زبان‌شناسی صورت‌تگر در حوزه علم اطلاعات بسیار کارآمد بود. این کتاب ۲۸ استاد از مجموع ۱۳۹ استاد به دست آورد. در طیف پیوند زبان‌شناسی و فناوری که در ابتدا توسط لاک و بوث ارائه شد، ساختارهای نحوی چامسکی به طرز چشمگیری درخشان تا علوم نوین رایانه و زبان‌شناسی را به صورت تعاملی تر به هم پیوند دهد. به طور خاص، توضیح رسمی و منطقی چامسکی از نحو جملات براساس قواعد ساختاری عبارت-محور، راه را برای برنامه‌نویسان رایانه در جهت تفسیر و تبدیل دستور زبان به یک الگوریتم سازگار و ماشین پسند، هموار کرد که مطابق آن، تجزیه و تحلیل صوری زبان از طریق رایانه توسعه یافت (Wu & Xu, 2011).

سه سال بعد، دو مقاله با عناوین: «ضریب توافق برای مقیاس‌های اسمی» و «وضعیت فعلی ترجمه‌ی ماشینی زبان‌ها» به ترتیب توسط کوهن و بار-هیل پیشنهاد شد. در مقاله‌ی اول، کوهن روشی محاسباتی برای توافق تناسبی میان مقیاس‌های اسمی پیشنهاد کرد که در آن به هر کلمه، عددی اختصاص داده می‌شود به گونه‌ای که تفاوت‌های متناسب بین اعداد نشان‌دهنده ضریب میزان توافق بین آن‌هاست. در نتیجه، در این محاسبه، کلمات در همان شبکه‌ای انتخاب می‌شوند که ضرایب ماتریس توافق آن‌ها در مقایسه با کلمات دیگر از همبستگی بیشتری برخوردار است. کاملاً بدیهی است که این روش به نقطه‌ی شروع بسیاری از تکنیک‌ها در تخمین خطای استاندارد و آزمون فرضیه محاسبات خودکار تبدیل شد (Cohen, 1960). در دیگری، بار-هیل بیان کرد که هر چه کیفیت ترجمه بالاتر باشد، پردازش در فرآیند ترجمه‌ی ماشینی کمتر می‌شود. این امر به وضوح ثابت می‌کند که به بازنگری خودکار متن خروجی برای اصلاح و پساویراستاری نیاز مبرم وجود دارد تا کیفیت و کارایی ترجمه‌ی ماشینی بهبود یابد (Bar-Hillel, 1960).

با اختلافی بارز در میزان استنادات، لوشنتاین (Levenshtein (1966) در مقاله خود ایده حیرت‌انگیزی را با عنوان: "کدهای باینری جهت تصحیح، حذف، درج و بازنگری" معرفی کرد که به طرز شگفت‌انگیزی توانست ۱۱۲ استاد را در بین ۳۰۴ مورد به دست آورد که به طور چشمگیری بیش از یک سوم کل استنادات است. از کل نقل‌قول‌ها لازم به ذکر است که مقاله لوشنتاین به عنوان دومین اثر پس از مقاله‌ی شانون با بیشترین تعداد استناد در دوره ۱۹۰۰ تا ۱۹۸۰ رتبه‌بندی می‌شود. به بیان دقیق‌تر، لوشنتاین آنتروپی رشته (String) را برای اندازه‌گیری فاصله بین دو دنباله، معرفی کرده است. به این معنا که فاصله‌ی بین دو کلمه، به عنوان مثال  $X$  و  $Y$ ، حداقل تعداد ویرایش‌هایی است که برای تبدیل  $X$  به  $Y$  یا یک کلمه به کلمه دیگر مورد نیاز است (Levenshtein, 1966). در واقع، یافته‌های لوشنتاین در مورد کدهای باینری بدون شک بر بسیاری دیگر از فناوران اطلاعات و متخصصان پردازش زبان تأثیر گذاشته است تا دیدگاه تازه‌ای نسبت به ترجمه‌ی ماشینی به دست آورند (Hardiyanti, 2021).

Salton et al. (1975) مقاله‌ای برای توضیح "مدل فضای برداری برای نمایه‌سازی خودکار" با ۲۶ استناد از ۳۱۷ مورد ارائه کردند. در مدل فضای برداری برای نمایه‌سازی خودکار، چگالی فضای بین اسناد به روشی نوآورانه متمایز می‌شود تا شاخص تطبیق بهینه، بازیابی گردد. بنابراین وقتی قرار است اسناد با یکدیگر مقایسه شوند یا زمانی که درخواست جستجو آغاز می‌شود، بهترین نمایه، نمایه‌ای است که بتوان آن را به طور مستقل و به گونه‌ای منحصر و مجزای از سایر نمایه‌ها بازیابی کرد. با این مدل پیشنهادی، می‌توان نزدیکترین معادل برای کلمه را در ماشین ترجمه انتخاب کرد (Bolte & Pauwels, 2020).

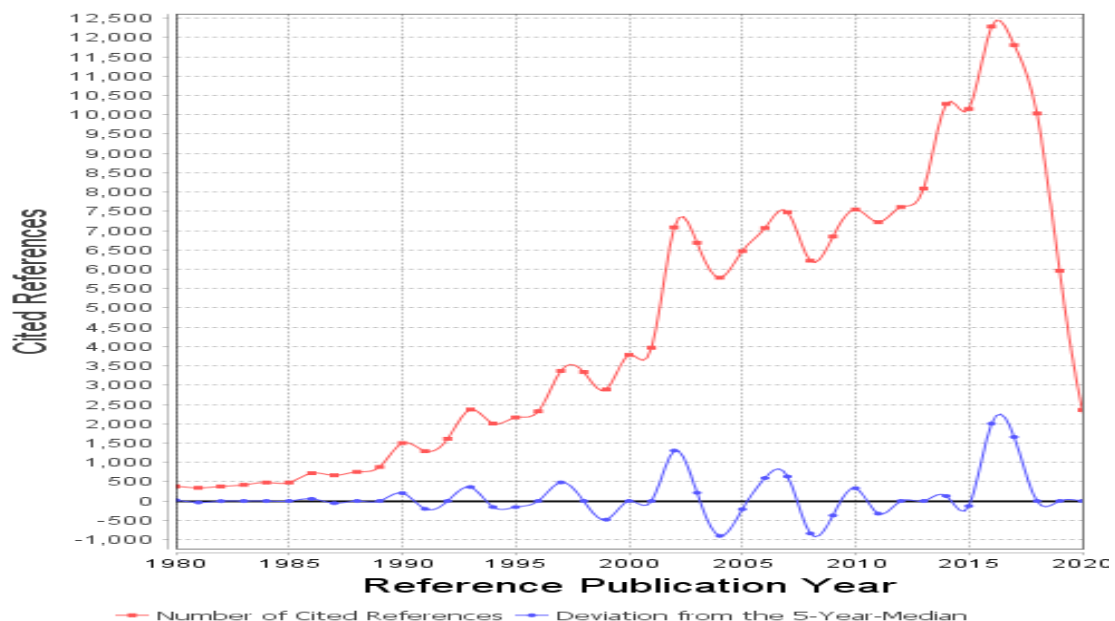
در سال ۱۹۷۷، دمپستر و همکارانش "حداکثر احتمال از داده‌های ناقص را از طریق الگوریتم حداکثرسازی انتظارات (EM)" پیشنهاد کردند که به اندازه‌ی کافی حدود یک سوم کل استنادها یعنی ۹۵ مورد از ۳۹۹ مورد را بدست آورد. آن‌ها به روشی آماری به نام الگوریتم حداکثرسازی انتظارات (EM) در جهت افزایش احتمال داده‌های ناقص اشاره کردند. به طور واضح، حداکثرسازی انتظارات قابلیت انتظار یک سیستم را در بازیابی میزان داده‌های از دست رفته براساس داده‌های مشاهده-شده ارتقا می‌دهد. سپس، داده‌های ناقص را با استفاده از فرآیند حداکثرسازی به داده‌های کامل تبدیل می‌کند. در نهایت، همگرایی میزان مطابقت بین مقادیر مورد نظر را نشان می‌دهد. ولی همگرایی کافی نبود، تکرار مراحل حداکثرسازی انتظارات

تا زمانی ادامه می‌یابد که کمترین اختلاف احتمال بین مقادیر حاصل شود (Dempster et al., 1977). شایان ذکر است که از این الگوریتم در یادگیری ماشین و فرآیند معادل‌یابی ماشین ترجمه استفاده می‌شود (Kwon et al., 2021). در همان سال، لنڈیس و کخ "اندازه‌گیری توافق ناظر برای داده‌های مقوله‌ای" را ارائه کردند که در مجله "بیومتریک" چاپ شده بود و توانست ۳۸ استناد از ۳۹۹ مورد را دریافت کند. در واقع روشی آماری را معرفی کردند که به انتخاب بهینه‌ی معادل کلمه می‌انجامد و باعث تسهیل و توسعه در فرآیند ترجمه می‌شود (Landis & Koch, 1977). چنین روش ابتکاری عمدتاً در انتخاب بهترین معادل برای یک کلمه در ماشین ترجمه کاربرد داشته است (Carl & Báez, 2019).

در آخرین مورد و البته نه کم اهمیت‌ترین مورد در بازه‌ی زمانی اول، پورتر "الگوریتمی برای حذف پسوند" با ۵۶ استناد در مجموع ۳۸۰ مورد ارائه کرد. در واقع، الگوریتمی برای حذف پسوند ابداع کرد که در زمان خود واقعاً کار قابل توجهی بود. به بیان دیگر، پورتر روش جدیدی را در بازیابی اطلاعات معرفی کرد و حذف خودکار پسوندها را بهبود بخشید که منجر شد داده‌های کمتری مورد نیاز باشد و همچنین فرایندها محدود، ساده و سریع‌تر عمل کنند (Porter, 1980).

### - بازه زمانی دوم: از سال ۱۹۸۱ تا ۲۰۲۰

نمودار ۲ توزیع تعداد مآخذ موجود در رکوردهای حوزه‌ی ترجمه‌ی ماشینی را در بازه‌ی ۱۹۸۱-۲۰۲۰ نشان می‌دهد. در نگاه اول مشاهده می‌شود که ده‌جمله بین سال‌های یاد شده به ترتیب در سال‌های ۱۳۶۵، ۹۰، ۹۳، ۱۳۷۶، ۱۳۷۹، ۱۳۸۱، ۱۳۸۶، ۱۳۸۹، ۱۳۹۳ و ۱۳۹۵ اتفاق افتاده است. در همان سال‌های اولیه نسبت به دوره‌ی ۱۹۸۰ تا ۱۹۸۹، افزایشی نسبی در تعداد استنادها از ۵۰۰ مورد به ۹۰۰ مورد دیده می‌شود و این رشد ادامه می‌یابد تا اینکه در سال ۱۹۹۳ به ۲۵۰۰ مورد می‌رسد. سپس، با نوساناتی جزئی تا سال ۱۹۹۷، تعداد استنادها به ۳۵۰۰ مورد افزایش یافت. در سال ۲۰۰۰، استنادها افزایش ملایمی تا رسیدن به مرز ۴۰۰۰ مورد داشتند. با این حال، افزایش نامنظم ولی چشمگیر در سال ۲۰۰۲ رخ داد که بیش از ۷۰۰۰ استناد وجود داشت. متعاقباً، استنادها تا سال ۲۰۱۳ حداقل افزایش و کاهش را تجربه کردند. در نهایت، مسیر استناد در سال‌های بعد به حدی توسعه یافت که در سال ۲۰۱۶ به سرعت به بیش از ۱۲۰۰۰ مورد افزایش یافت. اما پس از آن، استنادها و همچنین آثار تحقیقاتی تأثیرگذار کاهش یافت و به طور شگفت‌انگیزی به حدود کمتر از ۲۵۰۰ مورد رسید.



نمودار ۲. نمودار طیفی RPYS ترجمه‌ی ماشینی از سال ۱۹۸۱ تا ۲۰۲۰

Chart 2. RPYS spectral chart of machine translation from 1981 to 2020

همچنین، همان‌طور که در جدول ۲ مشاهده می‌شود، نتایج مربوط به کارهای تحقیقاتی تأثیرگذار در سال‌های مطرح بین سال‌های ۱۹۸۱ تا ۲۰۲۰ نشان داده می‌شود که بیشترین استناد را به دست آورده است.

**جدول ۲. جهش‌های تاریخی حوزه ترجمه‌ی ماشینی به همراه پرارجاع‌ترین آثار در بازه زمانی ۱۹۸۱ تا ۲۰۲۰**

**Table 2. Historical leaps in the field of machine translation along with the most referenced works in the period from 1981 to 2020**

RPY	Cited Reference	Cited References Count	No. of all citations of the year	Document Type
1986	Hutchins, W. J. (1986). Machine translation: past, present, future (p. 66). Chichester: Ellis Horwood.	68		Book
	Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. <i>Nature</i> , 323(6088), 533-536.	43	719	Journal Paper
	Lesk, M. (1986, June). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In Proceedings of the 5th annual international conference on Systems documentation (pp. 24-26).	42		Conference Paper
1990	Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J., ... & Roossin, P. S. (1990). A statistical approach to machine translation. <i>Computational linguistics</i> , 16(2), 79-85.	216	1498	Journal Paper
1993	Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. <i>Computational linguistics</i> , 19(2), 263-311.	709	2367	Journal Paper
1997	Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. <i>Neural computation</i> , 9(8), 1735-1780.	794	3370	Journal Paper
2000	F J Och, H Ney: Improved Statistical Alignment Models. Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, p.440-447, October 03-06, 2000, Hong Kong	192	3781	Conference Paper
2002	Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).	1698	7083	Conference Paper
	Stolcke, A. (2002). SRILM-an extensible language modeling toolkit. In Seventh international conference on spoken language processing.	436		Conference paper
2007	Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., & Herbst, E. (2007, June). Moses: Open source toolkit for statistical machine translation. In Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions (pp. 177-180).	1098	7478	Conference Paper
2010	Koehn Philipp. Statistical Machine Translation. Cambridge: Cambridge University Press, 2010	239	7547	Book
2014	Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).	745	10280	Book
2016	Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. <i>arXiv preprint arXiv:1609.08144</i> .	414	12284	Journal Paper

اولین اثر تأثیرگذار، کتاب جان هاجینز با عنوان «ترجمه‌ی ماشینی: گذشته، حال، آینده» است که ۶۸ استناد از مجموع ۷۱۹ مورد را به دست آورد. هاجینز، در این کتاب دیدگاه کلی مطالعات ترجمه‌ی ماشینی را براساس رویکردها و طرح‌های آن‌ها در طول ۴۰ سال از ۱۹۴۶ تا ۱۹۸۶ ارائه کرده است (Hutchins, 1986). در همین حین، روملهارت، هیتون و ویلیامز مقاله‌ای در مجله‌ای در مورد

"یادگیری بازنمایی‌ها از طریق بازسازی خطاها" ارائه کردند که توانست ۴۳ استاد به دست آورد. در حقیقت، سیستم یادگیری نوآورانه‌ای برای پردازش ماشین ابداع کردند که ماشین خود را براساس خطاهایی که قبلاً در پردازش مرتکب شده بود، اصلاح کند. این سیستم ابداعی به نام "بازساز" (Backpropagation)، سیستم پردازش شبکه‌های عصبی را بهبود می‌بخشد (Rumelhart et al., 1986). همچنین، در همان سال، لسک مقاله‌ای در مورد شناسایی و تشخیص درست معنای کلمه ارائه داد که ۴۲ استاد را از مجموع ۷۱۹ مورد استاد دریافت کرد. لسک در مقاله‌ی خود سیستم پردازش خودکار متن را پیشنهاد کرد که در آن می‌توان معنای صحیح یک کلمه را با کمک فرهنگ لغت آنلاین و همچنین استفاده از بافت کلمه تشخیص داد که مبنای استفاده از بافت در تشخیص معنای کلمه، از همین مقاله در ترجمه‌ی ماشینی مطرح گردید (Lesk, 1986).

در سال ۱۹۹۰، براون و همکارانش مقاله‌ای در مورد "رویکرد آماری در ترجمه‌ی ماشینی" نوشتند که در مجله "زبان‌شناسی محاسباتی" (Computational Linguistics) منتشر شد. این مقاله ۲۱۶ مورد استاد را از میان ۱۴۹۸ مورد استاد به خود اختصاص داد. نویسندگان از الگوریتم حداکثر انتظارات پیشنهاد شده توسط (Dempster et al., 1977) در توسعه‌ی مدل تخمین آماری جهت افزایش احتمال انتخاب مناسب‌ترین و دقیق‌ترین معادل برای یک کلمه استفاده کردند. در نهایت چنین مدلی به بهترین ترجمه برای یک جمله دست می‌یابد (Brown et al., 1990).

شاید مهم‌ترین مطالعه در خصوص ترجمه‌ی ماشینی در طول چهل سال اخیر مربوط به مقاله‌ای است با عنوان: "ریاضیات ترجمه‌ی ماشینی آماری: تخمین پارامتر" که ۷۰۹ مورد استاد را از بین ۲۳۶۷ مورد را به دست آورد؛ تقریباً یک سوم کل استادها. این مقاله توسط براون و همکاران در سال ۱۹۹۹ نوشته شده است. در این مقاله برخی از مدل‌های ریاضی خاص را ابداع کردند که با آن‌ها می‌توان کلمات را یکی یکی از زبان‌های مبدأ به زبان‌های مقصد با دقت بیشتری از طریق مجموعه‌ای از متون که قبلاً در اینترنت تغذیه شده‌اند، مطابقت داد. سپس، کلمات داده شده براساس الگوریتمی دقیق مرتب می‌شوند؛ به طوری که هم متناسب با نحو زبان مبدأ و همچنین زبان مقصد باشد؛ چنین مدلی ساخت جمله‌ی ترجمه شده را براساس زبان مقصد بهبود بخشید (Brown et al., 1993). در سال ۱۹۹۷، هوکرایتر و همکارانش مقاله‌ای با عنوان «حافظه‌ی کوتاه مدت و بلندمدت» معرفی کردند که توانست حدود یک چهارم کل استادها را به خود اختصاص دهد، یعنی ۷۹۴ مورد از ۳۳۷۰ مورد. به نظر می‌رسد این مقاله رتبه‌ی دوم از نظر تعداد استادها در یافتی در بیست سال گذشته را داشته باشد. در واقع، سیستم "بازساز" (Rumelhart et al., 1986) به منظور حفظ ذخیره‌سازی اطلاعات در طول مکث‌های طولانی را اصلاح و بهبود بخشیدند (Hochreiter & Schmidhuber, 1997).

در سال ۲۰۰۰، اوک و نی مقاله‌ای با عنوان "مدل بهبود یافته‌ی مرتب‌سازی" را در مجموعه مقالات سی و هشتمین نشست سالانه انجمن زبان‌شناسی محاسباتی ارائه کردند که به سادگی ۱۹۳ استاد از مجموع ۳۷۸۱ استاد دریافت کرد. در واقع، آن‌ها روشی برای ارزیابی و اندازه‌گیری پیشنهاد کردند که معیاری برای ارزیابی کیفیت پنج مدل همسویی IBM در ترجمه‌ی ماشینی آماری شد (Och & Ney, 2000).

Papineni et al. (2002) مقاله‌ای با عنوان "بلو" (Bleu): روشی برای ارزیابی خودکار ترجمه‌ی ماشینی" را در جلسات چهلیمین نشست سالانه انجمن زبان‌شناسی محاسباتی ارائه کردند که با تمایزی بسیار زیاد و به طور قابل توجهی، ۱۶۹۸ استاد را از کل تعداد ۷۰۸۳ مورد به دست آورد. با توجه به موضوعات مطرح شده در مقاله، لازم به تاکید است که محققین سعی در حذف دستکاری و ارزیابی انسانی برای خروجی‌های ترجمه‌ی ماشینی داشتند که پرهزینه و وقت‌گیر بود (Papineni et al., 2002). به طور واضح، مدل پیشنهادی آن‌ها کیفیت ترجمه‌ی ماشینی را ارتقا می‌دهد (Ulitkin et al., 2021). در همین سال و به طور همزمان، استولک مدل زبانی آماری برای تشخیص گفتار به نام "SRILM" را با استفاده از کتابخانه‌های ++C، برنامه‌های اجرایی و اسکریپت‌های کمک-رسان (Helper Scripts) توسعه و گسترش داد (Stolcke, 2002).

در سال ۲۰۰۷، کوهن و همکارانش، مقاله‌ای برای چهل و پنجمین نشست سالانه انجمن زبان‌شناسی محاسباتی به نام «موزز» (Moses): ابزار منبع باز برای ترجمه‌ی ماشینی آماری» ارائه کرد که نزدیک به ۱۰۹۸ استاد در مجموع ۷۴۷۸ مورد را به خود اختصاص داد. در

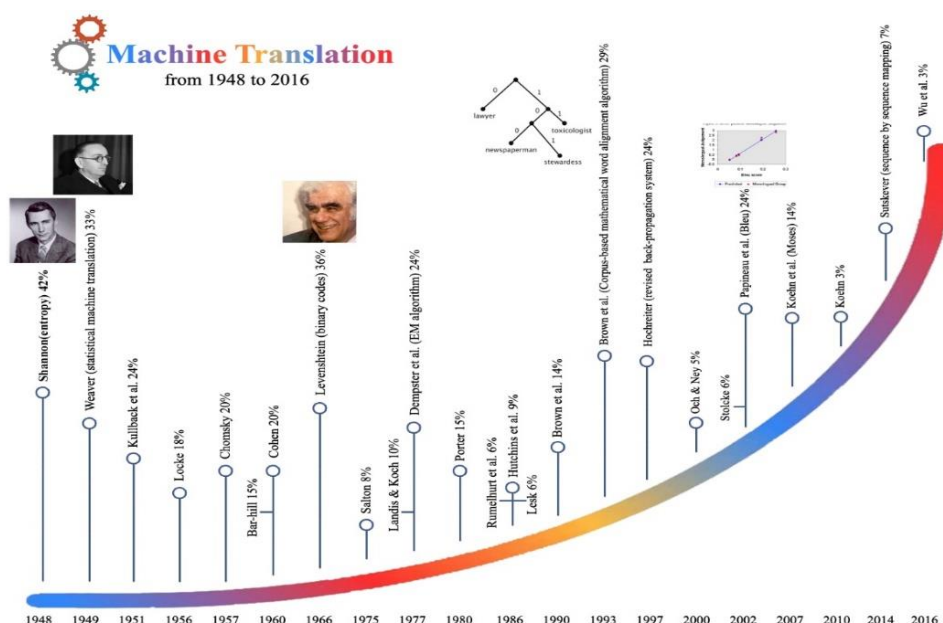
واقع، آن‌ها ترجمه‌ی ماشینی آماری را معرفی کردند که یک جعبه‌ی ابزار متن باز خلاق و انعطاف‌پذیر دارد و می‌تواند به طور سازگار با عوامل زبانی متناسب شود، خطاهای شبکه را مستقیماً رمزگشایی کند و قالب‌های مناسب داده‌ها را برای ترجمه‌ی ماشینی ارائه دهد (Koehn et al., 2007).

کوهن در سال ۲۰۱۰ کتابی با عنوان «ترجمه‌ی ماشینی آماری» که توسط انتشارات دانشگاه کمبریج منتشر شد را نوشت و توانست ۲۳۹ استاد از ۷۵۴۷ مورد را به دست آورد. او در کتاب خود، ترجمه‌ی ماشینی آماری را در سه بخش اصلی: مبانی ترجمه‌ی ماشینی آماری، روش‌های اصلی در ترجمه‌ی ماشینی آماری و موضوعات پیشرفته مرتبط با ترجمه‌ی ماشینی آماری مورد مطالعه و بررسی قرار داد (Koehn, 2010).

در روند انتقال از ترجمه‌ی ماشینی آماری به رویکردی جدید به نام ترجمه‌ی ماشینی عصبی، برخی از روش‌ها مانند شبکه‌های عصبی عمیق (Deep) در ترجمه‌ی ماشینی معرفی شدند. با در نظر گرفتن این موضوع، (Sutskever et al. (2014) مقاله‌ای با عنوان «یادگیری توالی توالی مرحله به مرحله در شبکه‌های عصبی» نوشت که توانست ۷۴۵ استاد در ۱۰۲۸۰ تعداد کل استاد به دست آورد. در روش سوتسکور، مجموعه‌های برجسب‌گذاری شده‌ی بزرگ را می‌توان در قالب توالی‌های مرحله به مرحله به ترتیب پردازش کرد؛ ویژگی که به عنوان کمبود در روش‌های قبلی یعنی شبکه‌های عصبی عمیق در نظر گرفته شد (Sutskever et al., 2014).

علیرغم این واقعیت که ترجمه‌ی ماشینی عصبی بر بسیاری از معایب سیستم‌های ترجمه عبارت-محور مرسوم غلبه کرده است، هنوز هم معایب و کمبودهای خاص خود را دارد. بر این اساس، (Wu et al. (2016) مقاله‌ای در مورد "ترجمه‌ی ماشینی عصبی گوگل (GNMT)" با تعداد استاد ۴۱۴ از ۱۲۲۸۴ مورد، معرفی کردند تا سیستم ترجمه‌ی ماشینی عصبی هم از لحاظ آموزش و همچنین استخراج ترجمه از نظر محاسباتی کم هزینه باشد. علاوه بر این، سیستم جدید قادر به پردازش کلمات کمیاب نیز می‌باشد که باعث بهبود دقت و سرعت ترجمه‌ی ماشینی عصبی است (Wu et al., 2016).

با توجه به تحلیلی که از هر دو دوره‌ی زمانی فوق به دست آمد، در اینفوگرافی زیر که توسط نویسندگان طراحی شده است، سعی بر آن است که خاستگاه‌های اصلی و تاثیرگذار در روند ترجمه‌ی ماشینی را که به روش RPYS شناسایی شده‌اند، به تصویر کشانده شود.



نمودار ۳. اینفوگرافی تاریخی ترجمه ماشینی براساس RPYS  
Figure 3. Machine translation history infographic based on RPYS

به طور کلی، نمودار ۳ نشان می‌دهد که بیشترین جرقه‌ی ابتکاری ترجمه‌ی ماشینی توسط سی. ای. شانون در این مقاله ایجاد شد: "نظریه ریاضی ارتباطات" که ۴۳ درصد از کل اسنادها را در سال ۱۹۴۸ به دست آورد و مفهوم آنتروپی در فناوری اطلاعات را معرفی کرد. علاوه بر این، ویور در گسترش ایده‌های شانون و همچنین معرفی ترجمه‌ی ماشینی آماری کلمه به کلمه نقش موثری داشت. ایده‌های او در سال ۱۹۴۹، ۳۳ درصد اسنادها را به دست آوردند. (Kullback & Leibler (1951)؛ (Locke & Booth (1956)؛ (Chomsky (1957)؛ (Cohen (1960) و (Bar-Hillel (1960) در عملیاتی کردن ایده‌های خلاقانه در ترجمه‌ی ماشینی کمک شایانی داشته‌اند، اما لوونشتاین بود که مفهوم انقلابی کدهای باینری را معرفی کرد تا عملکردهای تصحیح، حذف، درج و بازنگری وارد ترجمه‌ی ماشینی شوند؛ ایده‌ای که توانست بیشترین تعداد اسناد (۳۶ درصد) را در سال ۱۹۶۶ به دست آورد. علاوه بر این، (Salton et al. (1975) مدل فضای برداری برای نمایه‌سازی خودکار را معرفی کرد. با این حال، (Dempster et al. (1977) بودند که توانستند الگوریتم آماری جدیدی را به نام "حداکثرسازی انتظارات" به منظور بهینه کردن احتمال بر اساس داده‌های ناقص، ارائه دهند که به ۹۵ (۲۴ درصد) اسناد از ۳۹۹ مورد اسناد دست یافت. سپس در چهل سال اخیر، محققان بسیاری مانند (Landis & Koch (1977)؛ (Porter (1980)؛ (Hutchins (1986)؛ (Rumelhart et al. (1986)؛ (Lesk (1986) و (Brown et al. (1993) کمک شایانی به توسعه‌ی ترجمه‌ی ماشینی کردند. با این حال، مقاله براون و همکارانش در مورد "ریاضیات ترجمه‌ی ماشینی آماری" بود که توانست بیشترین تعداد اسناد (حدود ۲۹ درصد) را در مقایسه با سایر مطالعات در سال ۱۹۹۳ به دست آورد، مطالعه‌ای که کیفیت ترجمه‌ی ماشینی را بهبود و ارتقاء داد. این الگوریتم که مبتنی بر پیکره‌ی ریاضی است، ترتیب دقیق قرارگیری کلمات را در کنار یکدیگر متناسب با نحو زبان مبدأ و مقصد تنظیم می‌کند. علاوه بر این، مقاله (Hochreiter & Schmidhuber (1997) در مورد "حافظه‌ی کوتاه مدت بلندمدت" نیز تعداد قابل توجهی از اسنادها (حدود ۲۴ درصد) را در پایان قرن بیستم دریافت کرد که در واقع سیستم بازسازی روملهارت را تقویت کرد. اگرچه (Och & Ney (2000) مدل همسویی آماری بهبود یافته را پیشنهاد کردند، اما شایان ذکر است که مقاله‌ی (Papineni et al. (2002) در مورد "Bleu" بود که تعداد قابل توجهی از اسنادها (۲۴ درصد) را در ابتدای قرن بیست و یکم به خود اختصاص داد. مقاله‌ای که در ایجاد مدل ارزیابی فوری و کم هزینه برای خروجی‌های ترجمه‌ی ماشینی بسیار موثر واقع شد. سپس، روند اسنادها توسط (Stolcke (2002) ادامه یافت، تا اینکه مقاله ارائه شده توسط کوهن و همکاران، به اوج قابل توجهی از اسنادها (۱۴٪) در سال ۲۰۰۷ رسید. کوهن با معرفی "موزز"، جعبه ابزار منبع باز متشکل از ابزارهای مختلفی را ارائه کرد که توانست سیستم‌های ترجمه‌ی ماشینی را منظم‌تر کند. به طور عادی، تعداد اسنادها در ربع اول قرن حاضر با آثار مربوط به (Koehn (2010) و (Wu et al. (2016) اندکی پیشرفت کرد. با این حال، مقاله (Sutskever et al. (2014) حدود ۷۴۵ اسناد از ۱۰۲۸۰ را برای ارائه فرمت توالی مرحله به مرحله، به عنوان آخرین مطالعه قابل توجه در ترجمه‌ی ماشینی در طول چهل سال اخیر، دریافت کرد.

### بحث و نتیجه‌گیری

بدیهی است که کارهای تحقیقاتی بسیار برجسته و کاربردی در زمینه‌ی علوم کامپیوتر و فناوری اطلاعات پس از جنگ جهانی دوم به وجود آمد که به وضوح در راستای پیشرفت ترجمه‌ی ماشینی شکوفا شد؛ مطلبی که (Gupta & Dhawan (2019) و (Gile (2015) هم موکد تاکید داشتند. به طور دقیق‌تر، منصفانه به نظر می‌رسد که بگوییم این زبان ریاضیات (۰ و ۱)، و همچنین آمار (تخمین بهینه) بود که توانست هرگونه اطلاعات زبانی را وارد ماشین‌ها و رایانه‌ها کند تا به طور طبیعی پردازش شوند (Manz, 2021). علاوه بر این، لازم به ذکر است که اولین تلاش برای ترجمه‌ی ماشینی با معرفی روش تغذیه اطلاعات

زبانی به زبان سازگار با ماشین به کمک سیستم‌های کدگذاری ریاضی آغاز شد (Dew et al. 2018). همانگونه که مطرح شد، محققین در صدد ارائه‌ی مدلی براساس روش‌های آماری بودند که در آن، ترجمه‌ی کلمه به کلمه با دقت و صحت بیشتری انجام شود (Chakraverty et al. 2019). به آرامی اما مطمئناً، ایده‌ی ترجمه‌ی کلمه به کلمه، به ترجمه‌ی مبتنی بر عبارت و مبتنی بر نحو توسعه یافت (Umeozor, 2020). در این میان، تلاش‌های زیادی برای منطقی بودن ترجمه‌های خودکار صورت گرفت (Hilbert, 2021). با توجه به آن، محققان سعی کردند کیفیت ترجمه‌ها را با گنجاندن نقش بافت در دستیابی به یک ترجمه‌ی ماشینی بدون نقص به ویژه در بازیابی اطلاعات معنایی عبارات و اصطلاحات، بهبود بخشند (Khullar, 2021). همچنین، محققان با در نظر گرفتن مفهوم کیفیت در ترجمه‌ی ماشینی، روش‌هایی مانند محاسبات مبتنی بر پیکره، یادگیری عمیق، شبکه تکرار مکرر، آموزش سطح توالی برای ترجمه‌های ماشینی عصبی و غیره را برای ارتقای کارایی، وابستگی به خود و اصلاح خودکار در ترجمه‌ی ماشینی پیشنهاد کردند. سیستم‌هایی که ترجمه را با سرعت و دقت بالا در ترجمه‌ی ماشینی عصبی افزایش می‌دهند (Shao et al., 2021).

در این پژوهش، با بهره‌مندی از روش RPYS سعی شد سیر تکاملی حوزه ترجمه‌ی ماشینی شناسایی شود. در پژوهش‌های آتی می‌توان از روش‌های دیگر علم‌سنجی از قبیل تحلیل هم‌استنادی، هم‌واژگانی، هم‌نویسندگی جهت شناسایی ساختارهای فکری حاکم بر حوزه ترجمه‌ی ماشینی و کسب بینش جامع در این حوزه جالب استفاده نمود.

## تعارض منافع

هیچ گونه تعارض منافع از سوی نویسندگان گزارش نشده است.

## References

- Ballandonne, M. (2019). The historical roots (1880–1950) of recent contributions (2000–2017) to ecological economics: Insights from reference publication year spectroscopy. *Journal of Economic Methodology*, 26(4), 307-326. DOI: 10.1080/1350178X.2018.1554227
- Bar-Hillel, Y. (1960). The present status of automatic translation of languages. *Advances in computers*, 1, 91-163. [https://doi.org/10.1016/S0065-2458\(08\)60607-5](https://doi.org/10.1016/S0065-2458(08)60607-5)
- Bolte, J., & Pauwels, E. (2020). *A mathematical model for automatic differentiation in machine learning*. arXiv preprint arXiv:2006.02080. <https://doi.org/10.48550/arXiv.2006.02080>
- Bornmann, L., & Marx, W. (2014). The wisdom of citing scientists. *Journal of the Association for Information Science and Technology*, 65(6), 1288-1292. <https://doi.org/10.1002/asi.23100>
- Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J., Mercer, R. L., & Roossin, P. S. (1990). A statistical approach to machine translation. *Computational Linguistics*, 16(2), 79-85. <https://aclanthology.org/J90-2002.pdf>
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2), 263-311. <https://aclanthology.org/J93-2003.pdf>
- Carl, M., & Báez, M. C. T. (2019). Machine translation errors and the translation process: A study across different languages. *Journal of Specialised Translation*, 31, 107-132. [https://www.researchgate.net/publication/335920678\\_Machine\\_translation\\_errors\\_and\\_the\\_translation\\_process\\_a\\_study\\_across\\_different\\_languages](https://www.researchgate.net/publication/335920678_Machine_translation_errors_and_the_translation_process_a_study_across_different_languages).



- Chakraverty, S., Sahoo, D. M., & Mahato, N. R. (2019). Mcculloch–Pitts neural network model. In *Concepts of Soft Computing* (167-173). DOI:10.1007/978-981-13-7430-2\_11
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Comins, J. A., & Leydesdorff, L. (2016). Identification of long-term concept-symbols among citations: Can documents be clustered in terms of common intellectual histories? arXiv preprint arXiv:1601.00288. <https://doi.org/10.48550/arXiv.1601.00288>
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
- Dew, K. N., Turner, A. M., Choi, Y. K., Bosold, A., & Kirchhoff, K. (2018). Development of machine translation technology for assisting health communication: A systematic review. *Journal of biomedical informatics*, 85, 56-67. <https://doi.org/10.1016/j.jbi.2018.07.018>
- Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. *Aslib Journal of Information Management*, 72(3), 305-319. DOI:10.1108/AJIM-06-2019-0142
- Gile, D. (2015). Analyzing translation studies with scientometric data: From CIRIN to citation analysis. *Perspectives*, 23(2), 1-9. DOI:10.1080/0907676X.2014.972418
- Gupta, B. M., & Dhawan, S. M. (2019). Machine Translation Research: A Scientometric Assessment of Global Publications Output during 2007-16. *DESIDOC Journal of Library & Information Technology*, 39(1), 31-38. DOI:10.14429/djlit.39.1.13558
- Habibi, R., Mokhtarpour, R., & Khasseh, A. A. (2018). Analysis of Evolutionary Trends in Global Entrepreneurship Research using Scientometric Techniques. *Journal of Entrepreneurship Development*, 10(4), 575-594. <https://doi.org/10.22059/jed.2018.246176.652>
- Hardiyanti, M. (2021). Identifying The Common Type of Spelling Error by Leveraging Levenshtein Distance and N-gram. *Scientific Journal of Informatics*, 8(1), 71-75. DOI:10.15294/sji.v8i1.29273
- Heidarimoghadam, R., Khasseh, A. A., Vakilmofrad, H., Fattahi, A., & Amiri, M. R. (2021). Identification and Analysis of the Historical Origins of Ergonomics by Referenced Publication Year's Spectroscopy. *Iranian Journal of Ergonomics*, 9 (2), 42-57. DOI:10.30699/jergon.9.2.42. (In Persian)
- Hilbert, M. (2021). Information Theory for Human and Social Processes. *Multidisciplinary Digital Publishing Institute*, 23(1), 9. DOI:10.3390/e23010009
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. doi: <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hutchins, W. J. (1986). *Machine translation: past, present, future*. Ellis Horwood Chichester. <https://mt-archive.net/70/CL-1988-Kittredge.pdf>
- Khasseh, A. A., & Mokhtarpour, R. (2016). Tracing the historical origins of knowledge management issues through referenced publication years spectroscopy (RPYS). *Journal of Knowledge Management*, 20(6), 1393-1404. DOI:10.1108/JKM-01-2016-0019. (In Persian)
- Khasseh, A. A., Asghariyan, N., Tajedini, O., Moosavi, A., & Ghazizadeh, H. (2019). Identification and Analysis of the His-torical Origins of Occupational Therapy by Referenced Publication Years Spectroscopy. *Scientometrics Research Journal*, 5(9), 161-184. <https://doi.org/10.22070/rsci.2019.3710.1230>. (In Persian)

- Khullar, P. (2021). Are Ellipses important for Machine Translation?. *Computational Linguistics*, 47(4), 1-10. DOI:10.1162/coli\_a\_00414
- Koehn, P. (2010). *Statistical Machine Translation*. Netherlands: Cambridge University Press. [https://www.google.com/books/edition/Statistical\\_Machine\\_Translation/4v\\_Cx1wIMLkC?hl=en](https://www.google.com/books/edition/Statistical_Machine_Translation/4v_Cx1wIMLkC?hl=en)
- Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., & Herbst, E. (2007). *Moses: Open source toolkit for statistical machine translation*. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, (177-180). Association for Computational Linguistics. <http://www.aclweb.org/anthology/P07-2045>
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1), 79-86. DOI:10.1214/AOMS/1177729694
- Kwon, J., Ho, N., & Caramanis, C. (2021). On the minimax optimality of the EM algorithm for learning two-component mixed linear regression. *International Conference on Artificial Intelligence and Statistics, PMLR*, (130), 1405-1413. <https://proceedings.mlr.press/v130/kwon21b.html>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 33(1), 159-74. PMID: 843571. <https://pubmed.ncbi.nlm.nih.gov/843571/>
- Lesk, M. (1986). *Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone*. Proceedings of the 5th annual international conference on Systems documentation, Toronto, Ontario, Canada. <https://doi.org/318723.31872810.1145>
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, 10, 707-710. Corpus ID: 60827152
- Leydesdorff, L., Bornmann, L., Marx, W., & Milojević, S. (2014). Referenced Publication Years Spectroscopy applied to iMetrics: Scientometrics, Journal of Informetrics, and a relevant subset of JASIST. *Journal of Informetrics*, 8(1), 162-174. <https://doi.org/10.1016/j.joi.2013.11.006>
- Locke, W. N., & Booth, A. D. (1956). Machine translation of languages. *American Documentation*, 7(2), 135-136. DOI: 10.1002/asi.5090070209
- Manz, O. (2021). Entropy Coding. In *Well Packed–Not a Bit Too Much* (17-27). DOI:10.1007/978-3-658-34737-6\_4
- Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751-764. DOI:10.1002/asi.23089
- Mokhtarpour, R., Khasseh, A. (2017). Tracing the Historical Origins of Research Methodology Issues through Referenced Publication Years Spectroscopy (RPYS). *Journal of Studies in Library and Information Science*, 9(20), 43-58. <https://doi.org/10.22055/slis.2017.13186>. (In Persian)
- Mousavi Chelak A., Khasseh A. A., Soheili F. (2018). Exploring the evolution of reference services literature using Reference Publication Year Spectroscopy (RPYS). *Research on Information Science & Public Libraries*, 24(1), 103-124. <http://publij.ir/article-1-1707-en.html>. (In Persian)
- Och, F. J., & Ney, H. (2000). *Improved statistical alignment models*. Proceedings of the 38th annual meeting of the association for computational linguistics. <https://doi.org/10.3115/1075218.107527>
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). *Bleu: A method for automatic evaluation of machine translation*. Proceedings of the 40th annual meeting of the Association for Computational Linguistics. DOI:10.3115/1073083.1073135

- Petrilli, S. (2021). *Translation translation*. BRILL. <https://brill.com/display/title/27737?language=en>
- Reifler, E. (1950). *Studies in Mechanical Translation, n 1, MT*. Mimeographed.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533-536. <http://dx.doi.org/10.1038/323533a0>
- Saha, S., & Jose, J. (2020). Shannon entropy as a predictor of avoided crossing in confined atoms. *International Journal of Quantum Chemistry*, 120(22), 26374. DOI:10.1002/qua.26374
- Salton, G., Wong, A., & Yang, C.-S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620. <https://doi.org/10.1145/361219.361220>
- Scheidsteger, T., & Haunschild, R. (2020). Telling the story of solar energy meteorology into the satellite era by applying (co-citation) reference publication year spectroscopy. *Scientometrics*, 125(2), 1159-1177. DOI: 10.1007/s11192-020-03597-0
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423. <https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>
- Shao, C., Feng, Y., Zhang, J., Meng, F., & Zhou, J. (2021). Sequence-Level Training for Non-Autoregressive Neural Machine Translation. *arXiv preprint arXiv:2106.08122*. <https://doi.org/10.48550/arXiv.2106.08122>
- Soheili F., Khasseh A. A. (2015). Historical Origins of Information Behavior Research by Reference Publication Year Spectroscopy. *Journal of Information Processing and Management (JIPM)*, 31(1), 3-26. 10.35050/JIPM010.2015.001. (In Persian)
- Stolcke, A. (2002). *SRILM-an extensible language modeling toolkit*. Proceedings of International Conference on Spoken Language Processing, Denver, 901-904. <https://www.scrip.org/reference/referencespapers?referenceid=2664291>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to sequence learning with neural networks*. Proceedings of the 27th International Conference on Neural Information Processing Systems. 2, Montreal, Canada. <https://dl.acm.org/doi/10.5555/2969033.2969173>
- Ulitkin, I., Filippova, I., Ivanova, N., & Poroykov, A. (2021). Automatic evaluation of the quality of machine translation of a scientific text: the results of a five-year-long experiment. E3S Web of Conferences. DOI:10.1051/e3sconf/202128408001
- Umeozor, S. N. (2020). Information Retrieval: A Communication Process in the 21st Century Library. *International Journal of Knowledge Content Development & Technology*, 10(2), 7-18. <https://journals.sfu.ca/ijkcdt/index.php/ijkcdt/article/view/339>
- Vieira, L. N., O'Hagan, M., & O'Sullivan, C. (2021). Understanding the societal impacts of machine translation: A critical review of the literature on medical and legal use cases. *Information, Communication & Society*, 24(11), 1515-1532. <https://doi.org/10.1080/1369118X.2020.1776370>
- Wray, K. B., & Bornmann, L. (2014). Philosophy of science viewed through the lens of Referenced Publication Years Spectroscopy (RPYS). *Scientometrics*, 102(3), 1987-1996. DOI:10.1007/s11192-014-1465-6
- Wu, Y., & Xu, R. (2011). The Application of Chomsky's Syntactic Theory in Translation Study. *Journal of Language Teaching and Research*, 2(2), 396. DOI:10.4304/jltr.2.2.396-399
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., & Dean, J. (2016). Google's neural machine translation system: Bridging the

عزتی لارسری و خاصه: شناسایی ریشه‌های تاریخی حوزه‌ی ترجمه ماشینی: یک تحلیل ...

gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.  
<https://doi.org/10.48550/arXiv.1609.08144>

Yao, Q., Li, X., Luo, F., Yang, L., Liu, C., & Sun, J. (2019). The historical roots and seminal research on health equity: A referenced publication year spectroscopy (RPYS) analysis. *International Journal for Equity in Health*, 18(1), 1-15. <https://equityhealthj.biomedcentral.com/articles/10.1186/s12939-019-1058-3>



**Copyrights**

© 2024, by the authors. Licensee SCU, Ahvaz, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>)