The Canadian Journal of Information and Library Science
La Revue canadienne des sciences de l'information et de bibliothéconomie

Machine Translation in Scholarly Publishing
A Scoping Review
Traduction automatique dans la publication savante
Une revue de la littérature

Philips Ayeni, Emanuel Kulczycki et Lynne Bowker

Volume 48, numéro 1, 2025

URI : https://id.erudit.org/iderudit/1118790ar
DOI : https://doi.org/10.5206/cjils-rcsib.v48i1.22296

Aller au sommaire du numéro

Éditeur(s)

Canadian Association for Information Science - Association canadienne des sciences de l'information

ISSN

1195-096X (imprimé)
1920-7239 (numérique)

Découvrir la revue

Citer cet article

Ayeni, P., Kulczycki, E. & Bowker, L. (2025). Machine Translation in Scholarly Publishing: A Scoping Review. *The Canadian Journal of Information and Library Science / La Revue canadienne des sciences de l'information et de bibliothéconomie*, *48*(1), 123–145. https://doi.org/10.5206/cjils-rcsib.v48i1.22296

Résumé de l'article

L'anglais occupe une position centrale dans la publication savante, mais l'utilisation d'une lingua franca pour la publication savante a des conséquences pour les chercheurs, la science et la société. Par exemple, les chercheurs non anglophones peuvent mettre plus de temps à lire et à écrire en anglais et peuvent faire face à davantage de révisions et de rejets de manuscrits, ce qui peut potentiellement entraîner un volume de production de recherche plus faible, pouvant nuire à l'avancement de leur carrière. Dans quelle mesure les outils de traduction automatique (TA) (par exemple, Google Translate) peuvent-ils aider à soutenir un écosystème de publication savante plus multilingue ? Pour le savoir, nous avons entrepris une revue exploratoire de la littérature pour enquêter sur la manière dont les outils de TA sont utilisés pour la publication savante multilingue. Suite à une recherche multilingue dans neuf bases de données bibliographiques, 875 articles ont été récupérés et examinés, et 39 ont été inclus pour une enquête plus approfondie. L'analyse révèle que les outils de TA sont activement développés, testés, appliqués et évalués dans le contexte de la publication savante. Cependant, à l'heure actuelle, ces outils ne déplacent pas l'anglais de sa position centrale ; l'utilisation principale des outils de TA est actuellement de réduire la charge de la publication en anglais pour les chercheurs ayant une maîtrise limitée de l'anglais. Cela suggère que la technologie seule ne peut pas créer ou maintenir un écosystème de publication savante multilingue. Par conséquent, des politiques significatives, en plus d'outils de TA améliorés et de ressources linguistiques, sont nécessaires pour créer un paysage de publication savante plus diversifié et équitable sur le plan linguistique.

The Canadian Journal of Information and Library Science - La Revue canadienne des sciences de l'information et de bibliothéconomie (CJILS-RCSIB)
Vol. 48, No. 1 (2025)
DOI: 10.5206/cjils-rcsib.v48i1.22296
Article

Machine Translation in Scholarly Publishing: A Scoping Review

Philips Ayeni [1], Emanuel Kulczycki [2], and Lynne Bowker [3]
[1]University Library, Lakehead University
[2]Scholarly Communication Research Group, Adam Mickiewicz University
[3]Département de langues, linguistique et traduction, Université Laval

English occupies a central position in scholarly publishing, but using a lingua franca for scholarly publishing has consequences for scholars, science, and society. For instance, non-Anglophone researchers may need longer to read and write in English and may face more manuscript revisions and rejections, potentially leading to a lower volume of research output, which could negatively affect career advancement. To what extent can machine translation (MT) tools (e.g., Google Translate) help to support a more multilingual scholarly publishing ecosystem? To find out, we undertook a scoping review of the literature to investigate how MT tools are being used for multilingual scholarly publishing. Following a multilingual search in nine bibliographic databases, 875 papers were retrieved and screened, and 39 were included for closer investigation. Analysis reveals that MT tools are being actively developed, tested, applied, and evaluated in the context of scholarly publishing. However, at present, these tools are not displacing English from its central position; the main use of MT tools currently is to reduce the burden of publishing in English for scholars with limited English proficiency. This suggests that technology alone cannot create or sustain a multilingual scholarly publishing ecosystem. Hence, meaningful policies, in addition to improved MT tools and language resources, are needed to create a more linguistically diverse and equitable scholarly publishing landscape.

Keywords: machine translation (MT), scholarly publishing, multilingualism, scoping review, linguistic diversity, equity, policy

Introduction

For decades, English has occupied a central position in scholarly publishing. However, this model is coming under pressure as scholars who use other languages are advocating for multilingual scholarly publishing. The use of a lingua franca (i.e., a language used for communication between people who speak different languages) makes sense in principle: everyone who knows the language can participate in the conversation. In practice, however, the use of a lingua franca for scholarly publishing has inequities. For instance, it often requires more time and effort for non-native speakers of the lingua franca to read, write, and edit texts in that language (Amano et al., 2023). In addition, in the case of English, this may likewise give Western viewpoints more visibility and may influence what is studied, which methods are used, where findings are shared, and which communities benefit from the research (Angulo et al., 2021). The movement for multilingual scholarly publishing is gaining traction thanks to groups such as the Helsinki Initiative on Multilingualism in Scholarly Communication (Helsinki Initiative, 2019), UNESCO (2021), the Association francophone pour le savoir (Acfas) (St.-Onge et al., 2021), and Quebec's Commissaire à la langue française (2023). Yet, to create a viable multilingual scholarly publishing ecosystem, scholars need practical support to find, read and/or write scholarly works in other languages. Can machine translation (MT) tools help? To find out, we conducted a scoping review to investigate how MT tools are being used in scholarly publishing.

MT tools are computer tools that undertake the actual task of converting a text from one language (e.g., French) to another (e.g., English). MT tools have existed for decades, but early systems based on linguistics had limited success, in part because grammatical rules and bilingual dictionaries are not sufficient for enabling successful translation (Hutchins & Somers, 1992). When people translate, they make use of other types of knowledge (e.g., pragmatics, cultural knowledge, domain knowledge, real-world knowledge) that are difficult to program into computers (Bowker, 2023). Towards the end of the twentieth century, tool developers began to conceive of a new way of implementing MT – one that focused on the strengths of computers, such as their ability to match patterns and to perform rapid calculations. This

Correspondence concerning methods should be addressed to Philips Ayeni: payeni1@lakeheadu.ca. Questions on the subject matter (i.e., machine translation) should be relayed to Lynne Bowker: lynne.bowker.1@ulaval.ca.

resulted in a series of approaches that can be categorized as data-driven approaches to MT. The first to emerge was statistical MT, which uses probabilistic models to propose translations (Koehn, 2010). When Google Translate was first introduced in 2006, it had a statistical MT architecture, and the quality of the translation surpassed that of tools based on linguistics. In late 2016, a new data-driven approach to MT emerged, known as neural MT, which uses artificial neural networks and machine learning techniques (Forcada, 2017; Pérez-Ortiz et al., 2022). These task-specific neural MT tools were followed in late 2022 by generative AI tools based on Large Language Models (e.g., ChatGPT). LLM tools also use artificial neural networks and machine learning, but they can perform a variety of tasks, such as text generation, text summarization, text simplification, and answering questions, in addition to translation (Moorkens et al., 2024). Both neural MT and LLMs are considered to be state-of-the-art tools, and both have brought dramatic improvements to translation quality, making machine-translated text a viable starting point for many purposes. But are the tools viable for scholarly publishing?

Data-driven tools that use machine learning require massive amounts of training data to learn how to accomplish a given task. To learn how to translate, an AI tool must be pre-trained on billions of pages of high-quality translations (Forcada, 2017; Hughes, 2023). This training material is essentially a set of original texts in language A that are aligned sentence-by-sentence with their translations into language B. In the case of widely spoken languages such as English and French, it is relatively easy to locate a large number of previously translated texts, and so these are referred to as high-resource languages (Pérez-Ortiz et al., 2022). However, for languages that are less widely used, such as Cree or Welsh, it is challenging to find a large and high-quality collection of previously translated texts. This situation is referred to as a low-resource situation (Pérez-Ortiz et al., 2022). The concept of high- and low-resource also applies to subjects and text types. For instance, government administration is commonly discussed, whereas the reproductive challenges of early-generation hatchery coho salmon in the wild are discussed less often. Likewise, some text types are easy to access (e.g., scientific abstracts), while access to others may be limited (e.g., paywalled articles).

The translation quality of a data-driven MT tool is heavily influenced by the training data. The tools tend to work better for widely used languages, subjects, and text types because the training dataset can include more and more relevant material. In contrast, for languages, subjects and text types that are less common, the training dataset will be less robust, leading to lower quality translations. This concept of high-resource versus low-resource situations must be kept in mind when assessing the potential of MT for scholarly publishing. In addition, it is vital to remember that translation can be undertaken for different purposes, such as discovering and reading a text in another language or producing a text in another language for publication. Using a given tool may be more or less successful depending on the purpose of the translation.

Now that we have briefly described the current linguistic landscape in scholarly publishing and introduced the main MT tools that have been used in recent years (i.e., statistical MT, neural MT and LLMs), let's turn our attention more specifically to an investigation into how these tools are being used in the scholarly publishing ecosystem.

Objectives of the study

Our overarching research question of this scoping review is "How are MT tools being used for multilingual scholarly publishing?". We investigate the issue from the perspective of tool users rather than developers. The scoping review considers three main aspects:

1. To what extent are MT tools being used for scholarly publishing?
2. For what purposes within scholarly publishing are MT tools being used?
3. Which specific tools, languages, and text types are being used, and which are not?

To the extent possible based on the information found in the articles, we also make observations on the following aspects: strengths and limitations of the tools; reception of the tools (e.g., user satisfaction); and promising strategies for integrating tools (e.g., ways of working with tools to optimize output).

Methods

We employ a scoping review to synthesize and provide a critical overview of previously published material on our topic. A scoping review "follows a systematic approach to map evidence on a topic and identify main concepts, sources, and knowledge gaps" (Tricco et al., 2018, p. 1). Scoping reviews are useful for bringing together literature in disciplines with emerging evidence to address questions beyond those related to the effectiveness or experience of an intervention (Peters et al., 2015). They help in identifying gaps in the research knowledge base and reporting the types of evidence that address and inform practice in a particular field (Crilly, Jashapara, & Ferlie, 2010; Decaria, Sharp, & Petrella, 2012).

The search was limited to publications appearing between January 2017 and September 2023. The project funding source emphasized that knowledge synthesis projects should focus on contemporary trends. The date of January 2017 was chosen because the underlying architecture used by most MT tools began to change following the emergence of neural MT in November 2016. In the months that followed, many major MT tools, including Google Translate and DeepL, replaced their former statistical MT approach with one based on artificial neural networks. The end date of September 2023 was selected in order to be able to meet the funder's requirement of delivering a preliminary report by the end of 2023.

Limiting scoping reviews and other types of evidence synthesis (e.g., systematic reviews, meta-analyses) to English is problematic because it could result in biased findings and reduce generalizability (Hannah et al., 2024). We conducted our search in English, French, Spanish, and Polish (languages known to team members). Some databases include multilingual metadata, making it possible to retrieve additional articles in other languages. In this project, articles written in Bulgarian, Portuguese, Russian, and Turkish were retrieved via English or Spanish abstracts, and they were translated into English using MT to be evaluated for relevance and inclusion in the synthesis.

Search Process

Searches were conducted in nine academic databases to ensure wide coverage and reduce the risk of overlooking relevant articles: a) Scopus, b) Web of Science core collections, c) ERIC (Education Resources Information Centre), d) MLA (Modern Language Association) International Bibliography database, e) PubMed, f) Dimensions, g) Érudit (for French content), h) Redalyc (for Spanish content), and i) Google Scholar.

Selection of search terms

This review focused on the intersection of two areas: 1) fully automatic MT tools, and 2) scholarly publishing. Table 1 presents the list of English terms and relevant alternative terms that have been used in the literature for these concepts. It was developed by one of the authors (LB), who is a certified translator and has over 25 years of experience working in the domains of MT and scholarly publishing. It was verified by co-authors (PA & EK) who respectively have over 5 and over 15 years' experience working on scholarly publishing and evidence syntheses.

The terms were combined with parentheses and with a combination of the Boolean operator 'OR', which instructs the databases to search for one of the terms, and 'AND', which commands the databases to look for contents where both terms appear in a title, abstract, or list of keywords. The goal was to find the intersection of articles published on machine translation AND scholarly publishing. Figure 1 shows an example of an English-language search query formulated for the Scopus database.

Subsequently, one of the authors (LB) developed a query prompt for ChatGPT-3.5 to facilitate searching in French, Spanish, and Polish. See Appendix A for the ChatGPT prompts, process, and outcomes of query translation.

Selection Process

We developed inclusion and exclusion criteria to guide the selection of relevant works, which included empirical studies that are quantitative in nature (e.g., survey), qualitative studies (e.g., interview-based), experimental research (e.g., comparing the use of MT tools to other translation options), or mixed methods.

Inclusion criteria

Studies have been included for analysis if they meet the inclusion criteria below:

1. Focus on MT tools (e.g., Google Translate).

2. Focus on scholarly publishing, such as through: a)Discovering scholarly literature that was written in another language; b)Reading scholarly literature that was written in another language; c) Writing a scholarly paper in a non-dominant language (e.g., in the author's second or third language).

3. Focus on the use of MT by and for scholars rather than on tool development. An exception might be if the tool is being specifically developed or customized for scholarly publishing (rather than a general-purpose tool).

Exclusion criteria

Studies were excluded from analysis for the following reasons:

1. Tools that do not perform translation but only offer support to professional translators (e.g., electronic dictionaries).

2. MT tools being used by professional translators or translator trainees.

3. MT tools being used for language teaching or in language classrooms.

4. MT tools applied in any context that is not directly related to scholarly publishing (e.g., clinicians using translation tools to communicate with patients).

5. Development of MT tools for any application other than scholarly publishing.

6. Studies where the abbreviation "MT" refers to a concept other than machine translation, and no other search terms related to machine translation were present ("MT" turned out to be a noisy term that returned numerous irrelevant results).

Table 1

Search strategy

Terms relating to machine translation	Terms relating to scholarly publishing
automatic translation* OR automatic translator* OR DeepL OR DeepL Translator OR Google Translate OR Google translator OR machine translation* OR machine translator* OR neural machine translation* OR MT OR online translator* OR post-edit* OR translation engine* OR translation system* OR translation technolog* OR translation tool*	abstract* OR academic abstract* OR academic article* OR academic literature OR academic paper* OR academic publication* OR academic publishing OR academic writing OR journal article* OR journal publication* OR research article* OR research paper* OR research publication* OR science writing OR scientific abstract* OR scientific article* OR scientific literature OR scientific paper* OR scientific publication* OR scientific text* OR scholarly communication* OR scholarly publication* OR scholarly publishing OR scholarly writing OR writing for publication

Figure 1

Search query sample from Scopus database.

TITLE-ABS-KEY("automatic translation*" OR "automatic translator*" OR DeepL OR "DeepL Translator" OR "Google Translate" OR "Google translator" OR "machine translation*" OR "machine translator*" OR "MT" OR "neural machine translation*" OR "online translator*" OR "post-edit*" OR "translation engine*" OR "translation system*" OR "translation technolog*" OR "translation tool*")) AND (TITLE-ABS-KEY("abstract*" OR "academic abstract*" OR "academic article*" OR "academic literature" OR "academic paper*" OR "academic publication*" OR "academic publishing" OR "academic writing" OR "journal article*" OR "journal publication*" OR "research article*" OR "research paper*" OR "research publication*" OR "science writing" OR "scientific abstract*" OR "scientific article*" OR "scientific literature" OR "scientific paper*" OR "scientific publication*" OR "scientific text*" OR "scholarly communication*" OR "scholarly publication*" OR "scholarly publishing" OR "scholarly writing" OR "writing for publication")) AND PUBYEAR > 2017 AND PUBYEAR < 2023

7. Studies where the terms “abstract”, “academic abstract”, or “scientific abstract” appeared only as section headings, and no other search terms related to scholarly publishing were present (these turned out to be noisy terms that returned numerous irrelevant results).

Search and screening outcomes

As previously mentioned, to ensure broad coverage, we searched nine academic databases. Scopus was searched on June 25, 2023, by one of the authors (EK) using the predetermined search queries. The search returned 215 records, which were exported to an Excel spreadsheet for further analysis. The inclusion and exclusion criteria were used as the basis for an initial manual screening of the title, abstract, and keywords (where available) of the records (LB). Following this initial screening, 57 records were saved to the reference management tool Zotero for full text screening. Web of Science (WoS) core collections were searched on June 28, 2023, by one of the authors (PA) and returned 121 records, which were exported to Excel for further analysis. After initial screening of the title and abstract, 20 potentially relevant articles were saved to Zotero for full-text screening. Based on the two initial searches, the query was refined by one of the authors (LB) to eliminate noisy terms (e.g., “MT”, “abstract”) that were returning a high volume of irrelevant content. Articles that had been located in the Scopus and WoS databases based on noisy search terms were discarded for consistency. Next, the PubMed, MLA, and ERIC databases were searched on July 31, 2023, by one of the authors (PA). PubMed returned 13 articles, MLA returned 5 articles, and ERIC returned 2 articles, and these were all exported to Excel for further screening. After initial title, abstract, and keyword screening of the records found, 6 records from PubMed, one record from MLA, and 4 records from ERIC were saved to Zotero for full-text screening.

Furthermore, the Dimensions database was searched on August 8, 2023, by one of the authors (PA), returning 500 records, which were exported to Excel for further screening. After the title, abstract, and keyword screening of the records, 61 records were found to be potentially relevant to our study and were saved to Zotero for full-text screening. To further ensure that we accounted for articles published in languages other than English, we also searched Redalyc (Spanish) and Érudit (French) on August 15, 2023. Using the translation query that one of the authors (LB) developed by prompting ChatGPT-3.5, Redalyc returned four articles. However, after initial title and abstract screening, none of the articles were found to be relevant to our study. Similarly, the search in Érudit using the French search query returned no results. None of the searches using the Polish search query (developed by EK) returned any results. To further ensure that we accounted for content published in open sources that may not be captured in proprietary databases, we searched Google Scholar on September 8, 2023. After screening the titles and abstracts of the first 10 pages of the search results and eliminating duplicate results that had appeared in previous database searches, 15 potentially relevant articles were saved to Zotero for full-text screening.

Full-text screening outcome

Full-text screening was carried out independently by two authors (PA & LB) and recorded in Excel. The reasons for inclusion or exclusion were provided, which then allowed the authors to compare, discuss, and resolve any differences. At the end of the full-text assessment of 160 potentially relevant articles, 121 articles were excluded. Ninety-six articles were excluded because they focused on the use of MT outside of scholarly publishing, while six articles focused on barriers faced by non-Anglophones in scholarly publishing without a mention of MT. Six other articles were review articles and editorials with no empirical data, 10 articles focused on translation-related topics but not specifically on MT tools, and 3 articles were retracted by the publishers. Consequently, 39 articles that met our inclusion criteria were retained for qualitative analysis. The search process and outcome are reported in Figure 2 below, following the Preferred Reporting Items for Systematic reviews and Meta-Analyses - extension for Scoping Review (PRISMA-ScR) checklist (Tricco et al., 2018).

Data extraction and synthesis approach

The coding scheme was pre-developed by one author (LB) who is an expert in MT and scholarly publishing. It was reviewed by other members of the team (PA & EK) who have experience in evidence synthesis and scholarly publishing. One of the authors (PA) then applied the coding scheme in NVivo and carried out qualitative content analysis of the 39 included studies. This was done by synthesizing the findings from the included studies to uncover how translation tools have been used for scholarly publishing. The qualitative content analysis also revealed gaps in existing studies, leading to recommendations for future studies. The pre-developed codes included four main themes, with sub-themes and categories. The main themes are 1) purpose of the study, 2) tools, 3) languages, and 4) text types.

The content analysis of the included articles was done by highlighting sentences from articles that fit within the pre-developed codes. When the texts did not fit, new codes were developed, leading to the modification and refinement of the pre-developed codes to include two additional themes: 5) evaluation methods and 6) evaluation outcomes. See Appendix B for a detailed description of the themes in the codebook imported from NVivo (v.14). To check the accuracy of the initial coding, a second author (LB) randomly selected five articles from the included studies and independently coded them using the revised coding scheme. Afterwards, their codes were compared, and the authors discussed and resolved some minor discrepancies to achieve inter-annotator agreement. Changes that resulted from resolving the discrepancies were applied to all instances of coding the phenomenon in question.

Finally, a frequency tool incorporated into NVivo (v.14) was used to generate a list of the 100 words that appear most frequently in the corpus of the 39 included studies.

Results

This section presents the findings from the analysis of the 39 included studies, beginning with some general characteristics of these studies as summarized in Appendix C. Next, it discusses the six themes that emerged from the qualitative coding and analysis: 1) purpose of the study, 2) tools, 3) languages, 4) text types, 5) evaluation methods, and 6) evaluation outcomes.

Study characteristics

As presented in Figure 3, five studies (13%) were conducted by first authors affiliated in Japan, three studies (8%) each in Canada, Taiwan, the UK, and the USA, followed by two studies (5%) each in China, Brazil, Germany, Ireland, and Russia. One study was conducted in each of the other countries. For one study, it was not possible to confirm the location definitively since the authors declared an affiliation with an international organization (Amazon) without specifying the location. Several projects were conducted collaboratively by researchers in different locations, but only the country of the first author's affiliation has been captured for this high-level portrait. While English-speaking countries are represented, a greater volume of research on this topic is taking place in countries where English is not the (only) national language.

Twenty studies were published as journal articles (51%), followed by 15 conference papers (39%). Others were published as book chapters and workshop reports. In terms of methods used, 31 studies (80%) used experiments. Four studies (10%) used mixed methods approaches by combining surveys, interviews, and analysis of written texts produced with MT tools. A few studies used pilot studies and workshops to produce or evaluate scholarly publications using MT tools.

As noted previously, we used NVivo (v.14) to generate a list of the 100 most frequent words that appeared in the 39 studies

Figure 2

Search process and outcome flowchart.

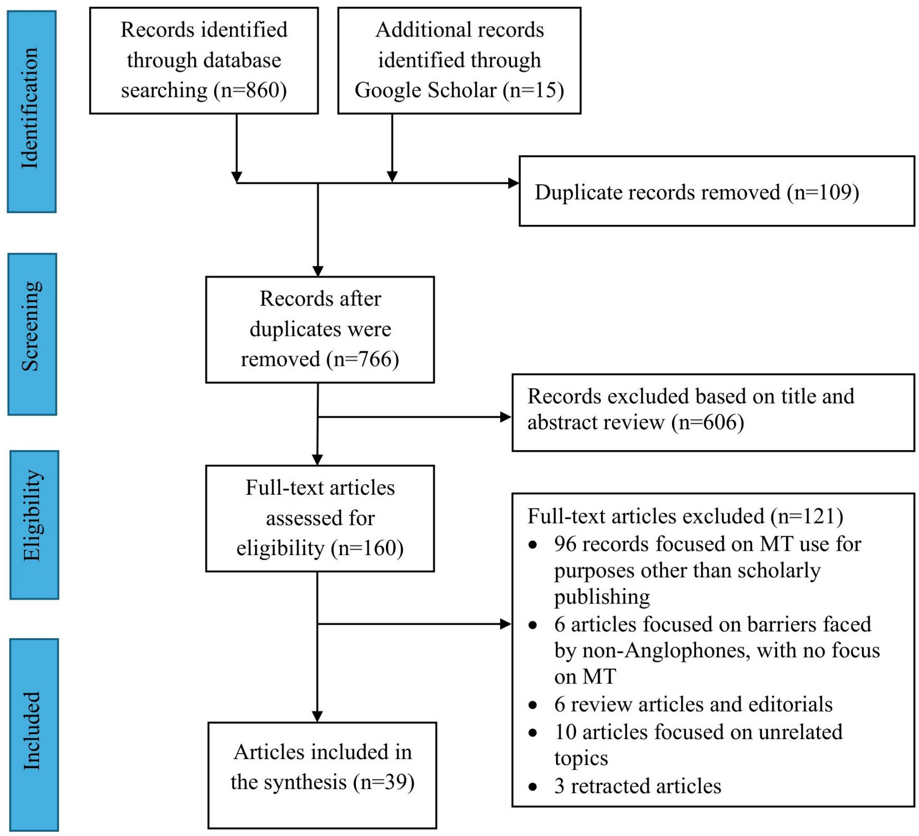
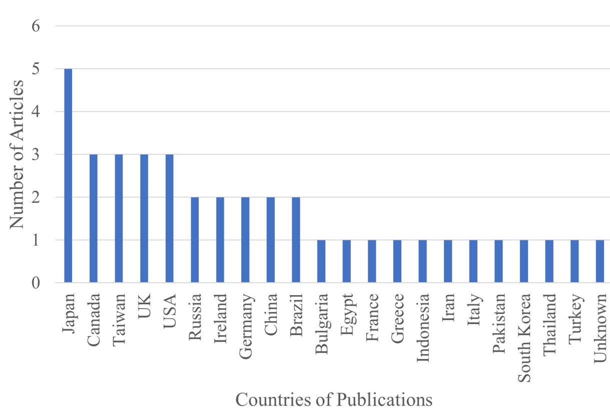


Figure 3

Countries of publication of the included studies (n = 22).



(see Figure 4). “Translators” is the most frequent word, while “English” comes in second place. Six other languages appear further down the list: Japanese in 18th place, Chinese in 23rd place, French in 31st place, German in 59th place, Spanish in 62nd place, and Italian in 81st place. Meanwhile, the word “English” appears 202 times in the 39 included studies — over four times more often than the second-most frequently mentioned “Japanese”, which occurs just 44 times in total.

Purpose of the studies

Of the 39 included studies, 16 (41%) evaluated translation quality by checking the performance of the MT tools and the quality of machine-translated texts (e.g., Bawden et al., 2020; Bowker, 2019; Daniele, 2019; Kostadinova, 2019). Twelve studies (31%) focused on developing MT tools for scholarly publishing (e.g., Chang et. al., 2020; Nayak, et al., 2020; Rousssis et al., 2022; Sel & Hanbay, 2022). For instance, Roussis et al. (2022) focused on the development of an MT tool based on SciPar, a multilingual parallel corpus of thesis and dissertation abstracts with 9.17 million sentence pairs in 31 language pairs. The MT tool was developed to facilitate the translation of scientific text to and from these languages

Figure 4

Word cloud of the 100 most frequent words in the included studies.

to improve equitable access to scientific knowledge and accelerate research. Ten (26%) of the studies focused on using MT tools to produce scholarly content (e.g., Bowker, 2019, 2020; O'Brien et al., 2018; Sun & Yang, 2023). In contrast, just three (8%) studies explore how MT tools can be used for discovering scholarly content (Bowker, 2018; Dobrynina, 2021; Esmailpour et al., 2020).

Tools

Findings showed that, overwhelmingly, the main type of MT tool in use in the scholarly publishing community during the study period was neural MT. Only three studies involved statistical SMT (e.g., Neves et al., 2018; Soares et al., 2018), which is an architecture that has been superseded by NMT since late 2016 (Koehn, 2020). Meanwhile, only one study (Takeshita et al., 2022) incorporated research involving LLMs, which were introduced in late 2022.

The most frequently employed tool is Google Translate, used in 18 (45%) of the included studies (e.g., Chang et al., 2020; Daniele, 2019; Esmailpour et al., 2020; Sun et al., 2022). The second-most popular tool in the studies is DeepL Translator (e.g., Takeshita et al., 2022; Wahab et al., 2020), which accounts for four (10%) of the included studies. Few studies investigated other tools, such as Microsoft Translator, Baidu, or Amazon Translate (e.g., Kostadinova 2019), although it must be noted that these tool names were not explicitly used as search terms, whereas Google Translate and DeepL were. Meanwhile, nine (23%) studies focused on comparing two or more translation tools (e.g., Bowker, 2018; Soares et al., 2021). While the majority of studies focus on using existing tools, 12 (31%) research teams set out to develop custom-built prototype MT tools designed specifically to translate scholarly texts (e.g., Xie et al., 2020; Xu et al., 2021).

While MT is the main technology of interest to the scholarly community, a few other types of tools are investigated. For instance, Zhang & Misra (2023) examine the integration of MT with cross-lingual information retrieval (CLIR) tools, which allow users to search in one language and retrieve texts written in another language. In addition, Zomer & Frankenberg-Garcia (2021) focus on language checkers to improve input quality, which can then improve the quality of MT output.

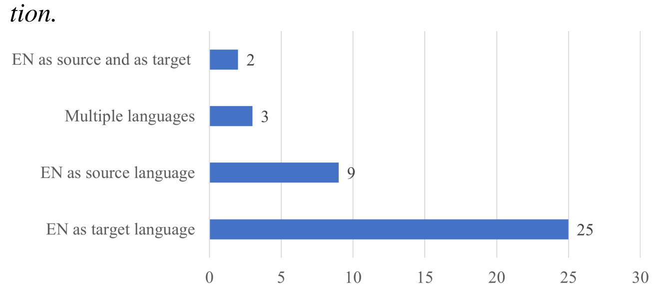
Languages

As noted in the introduction, English is a key language for scholarly publishing, and many translation tools use a data-driven approach. For high-resource languages and language pairs, it is straightforward to collect good-quality training data, and the resulting tools perform relatively well. However, for low-resource languages, it is challenging to gather sufficient high-quality training data, and tool performance is likely to be lower.

One noteworthy observation is that all 39 of the studies included in this review include English as one of the working languages for translation (see Appendix C). More specifically, Figure 5 illustrates that in 25 studies (64%), English is the target language (i.e., texts are being translated into English), while in only nine (23%) studies, English is the source language (i.e., texts are being translated out of English). Two (5%) additional studies are bidirectional, and the remaining three (8%) studies use multiple language pairs, including some that involve English. The high proportion of studies that use English as the target language confirms that translation tools are not currently being used to displace English as the central language of scholarly publishing. Instead, the main application for MT tools seems to be to reduce the burden of preparing English-language publications, which does not contribute to creating a genuinely multilingual scholarly publishing ecosystem.

Figure 5

Distribution of studies involving English by translation direction.



With regard to other languages, 25 (64%) of the included studies focused on high-resource language pairs, with the most frequently used being English-Japanese or Japanese-English (e.g., Takakusagi et al., 2021; Takeshita et al., 2022), and English-Chinese or Chinese-English (Bawden et al., 2020; Sun & Yang, 2023). Other high-resource languages featured in the studies include French, Portuguese, Spanish, and German (e.g., Bawden et al., 2020; Neves al., 2018; Roussis et al., 2022; Soares et al., 2019; Xu et al., 2023).

Meanwhile, only 10 (26%) studies included at least one low-resource language, such as Basque (Nayak et al., 2019), Bulgarian (Kostadinova, 2019), Indonesian (Winiharti et al., 2021), Persian (Esmailpour et al, 2020), and Thai (Tongpoon-Patanasorn & Griffith, 2020). Given the tendency for low-resource languages to produce lower quality translations, it is not surprising that some studies using low-resource languages focus on investigating how accurately and naturally MT tools can translate academic texts written in those languages into English (Tongpoon-Patanasorn & Griffith, 2020; Winiharti et al., 2021). Meanwhile, some others investigate techniques for producing translation-friendly texts that can be more easily machine-translated into English (e.g., Esmailpour et al, 2020; Kostadinova, 2019).

Text types

Scholarly publishing covers multiple text types. As explained in the introduction, data-driven technologies require a large volume of data, but it must also be the right kind of data. For instance, the characteristics of an abstract are not the same as those found on a PowerPoint slide. To obtain a high-quality translation, the characteristics of the text to be translated must be well represented in the training data.

In the studies included in this review, 12 (31%) explored how MT could be used to help produce written texts such as articles (e.g., Esmailpour et al., 2020; O'Brien et al., 2018; Sun et al., 2022). However, 9 studies involved using tools that had been trained using only abstracts (23%) (e.g., Daniele, 2019; Nayak et al., 2020; Xu et al., 2021), while 4 (10%) used only theses/dissertations (e.g., Chang et al., 2020: Sel & Hanbay, 2022; Soares et al., 2018).

Evaluation methods

As noted previously, scholars are actively evaluating the quality of machine-translated texts. This can be done using automatic metrics, which have the advantage of being fast, convenient, and cheap, or via manual evaluation, which can be more accurate, but is also more costly and time-consuming. Of the 27 included studies that dealt with an aspect of translation quality assessment, 18 (67%) used automatic metrics, while 15 (56%) employed manual evaluation. A few incorporated both (e.g., Matsumura et al., 2018; Xie et al., 2020), which explains why these numbers add up to more than 27 (100%).

Among those studies that use automatic metrics, the most popular is the BLEU score, which is used in 13 (72%) cases (e.g., Mino et al., 2021; Morishota et al., 2019; Nayak et al., 2020). Other studies employed some less common automatic metrics such as ROUGE (Yamamoto et al., 2021) or METEOR (Sel & Hanbay, 2022). Although BLEU is a very popular evaluation tool, it has some known shortcomings (Callison-Burch et al., 2006), and it was not developed for NMT tools. Other metrics have been specifically designed for NMT (e.g., BERTscore, BLEURT), although these are not problem-free (e.g., Hanna and Bojar (2021) found BERTscore to be very sensitive to text style). However, none of the retained studies employed the NMT-oriented metrics, which could point to low awareness and to the importance of making these metrics tools accessible to broader research communities.

Manual evaluation is used in 15 (56%) studies, meaning that subject experts and/or human translators are engaged to check the translations. Some studies employed native speakers of the target language who had good knowledge of the source language (e.g., Dobrynina, 2021; Esmailpour et al., 2020), evaluating the translated texts according to international academic and publication conventions (e.g., Castilho et al., 2018). For instance, Dobrynina (2021) conducted a manual evaluation of MT output by checking the accuracy of translation, consistency of translation, and compliance with the rules of English grammar.

Evaluation outcomes

This section discusses the evaluation outcomes of translation tools as presented in the included studies. This includes the strengths and limitations of the tools, as well as their reception by users. The section ends with suggestions and strategies for integrating the tools in the scholarly publishing system.

Strengths and limitations of the tools

Providing an overall assessment of strengths and limitations of translation tools is challenging owing to the way that data-driven tools are influenced by high- and low-resource situations as these pertain to different languages, text types, and subjects. In other words, a translation tool that works well for one language combination or subject matter may perform poorly for another. In addition, as noted in the introduction, translation can be undertaken for different purposes, and MT may be more suitable for some use cases and less suitable for others. Therefore, it is not surprising to see a wide range of experiences reflected in the studies included in this review.

Some studies reported a good translation performance by Google Translate for translating scientific articles and abstracts (Daniele, 2019; Yamamoto et al., 2021; Zhovtova et al., 2020). For instance, Daniele (2019) found the performance of Google Translate to be fairly effective because of the intrinsic characteristics of medical language and lexical items used in medical abstracts. In contrast, a number of studies report that translation errors are still common, including spelling errors (Bowker, 2018), lexical errors (Sun et al., 2022), various types of grammatical errors (e.g., agreements, tenses) (Daniele, 2019; Sun et al., 2022), mechanics and punctuation errors (Tongpoon-Patanasorn & Griffith, 2020), and stylistic issues (e.g., overuse of passive voice) (Dobrynina, 2021). In general, studies reporting higher performance tended to be those involving high-resource languages (e.g., Zhivotova et al., 2020; Matsumura et al., 2018; Neves et al., 2018). However, it is important to note that translation tools operating with high-resource languages are not error-free, as reported by researchers working with French (Bowker, 2018), Italian (Daniele, 2019), and Japanese (Yamamoto et al., 2021).

The volatility and unpredictability of data-driven tools have been observed elsewhere (e.g., Fadaee & Monz, 2020), and it was evident in some of the included studies that tracked system performance over time. The tools continue to learn when their training data is updated, but it is important to recognize that, as part of the learning process, the tool can acquire bad strategies as well as good strategies (especially if the training data is not a good match to the task at hand). Consequently, some changes in system performance may result in improvements, while others may be setbacks. For instance, in a comparative study, Neves et al. (2018) noted that the quality of translation into English had improved over time, while the performance into French had plateaued, and translation into Romanian revealed new problems not previously observed. This may indicate that the gap between the quality of translation tools available for high- and low-resource languages will widen since the volume of text produced in high-resource languages is growing more quickly. For studies comparing a custom-built prototype against an existing general-purpose translation tool, the results suggest that domain-adapted MT tools perform better when translating specialized texts in these domains (e.g., Sel & Hanbay, 2022; Matsumura et al., 2018).

Reception of the tools

There is a noticeable interest in and some positive attitudes towards translation tools in scholarly publishing. Kim and LaBianca (2018) found that students and faculty have positive attitudes towards the ethical use of MT tools for academic writing, particularly at the word and phrase levels, as well as for translating sentences and paragraphs. Sun and Yang (2023) reported that participants' writing quality improved as a result of translation-friendly writing strategies and also affirmed the legitimacy of using MT as an aid for writing scholarly texts in another language. Zou et al. (2023) found that Google Translate is the most important language tool that supports the academic writing process of one of their participants, who was unable to write directly in English. However, some scholars continue to have reservations about quality (e.g., Dobrynina, 2021), while others note that the decision to use a translation tool can depend on a number of factors, such as the specific task at hand (e.g., understanding vs producing a text) (e.g., O'Brien et al., 2018; Zulfiqar et al., 2018).

Strategies for integrating the tools

The studies included in the review reveal that it can be useful to develop strategies for integrating translation tools so that researchers can work in a way that improves translation quality. For instance, Bowker (2019, p. 619) submits that "while MT works well for reading existing literature, it is less immediately successful as a writing aid and so more emphasis should be placed on both pre- and post-editing techniques, which may differ from one language to the next". Similarly, Zhonotova et al. (2019) found that pre-editing a source text helps to improve MT quality. Sun & Yang (2023) report that back-translation (i.e., writing a text in language A, using MT to translate it into language B, and then using MT again to translate it back into language A to compare against the original text) is regularly used by participants to maximize the benefits of the translation tools. They suggest that it is important to develop students' abilities to write for MT. Meanwhile, for publication purposes, it is important not to rely solely on the tools but to verify and post-edit the translated texts (e.g., Dobrynina, 2021). Overall, the authors of the studies included in this review emphasize that there is a need for awareness, training, and MT literacy among scholars who employ translation tools to discover, to consult, and especially to write scholarly publications. For example, Bowker (2020) recommends helping translation tool users, many of whom have no background in translation, to improve their MT literacy through the provision of training or guidelines on the effective use of MT tools.

Discussion

The overarching goal of this scoping review was to gain a better understanding of how translation tools are being used in the context of scholarly publishing. One key observation is that the field is evolving rapidly, as evidenced by the fact that only a few studies focused on statistical MT – a paradigm that dominated the field for approximately 15 years until neural MT was introduced in November 2016 (Forcada, 2017). Moreover, although ChatGPT and other generative AI tools only became accessible in December 2022, some emergent research on this topic is also represented in our review. The rapid pace of change presents challenges for policymakers, who must respond quickly to support researchers.

Our review reveals that MT tools are indeed being used in the scholarly community, and that the most commonly used commercial tool is Google Translate, followed by DeepL. The high use of Google Translate may be explained by its free and easily accessible online interface and its broad linguistic coverage (130 languages during the review period) (Winiharti, Syihabuddin, & Sudana, 2021). However, these commercial tools are general-purpose MT tools (e.g., Google Translate), which are typically and most successfully used to translate between high-resource languages and on more general topics. The fact that nearly one-third of the studies focused on designing or customizing prototype MT tools specifically for scholarly texts suggests that general-purpose tools are not currently able to satisfy user needs related to scholarly publishing.

With regard to languages, English and a handful of other high-resource languages (e.g., Chinese, French, Spanish, German) are attracting more attention, while less research focuses on low-resource languages. Our review provides additional evidence that English continues to occupy the central position in scholarly publishing. In the word cloud featured in Figure 3, English is mentioned more times than all the other languages combined. In addition, all of the studies included in our review had English as one of the languages under investigation, and the majority focused on how MT tools can be used to help scholars with limited English proficiency to write in English. Although MT tools have the potential to help level the playing field by allowing users to discover and access works written in other languages, this is not presently a main focus of MT use in the scholarly community.

Implications

While it is encouraging to see that some researchers are investigating low-resource languages, most attention is focused on high-resource languages. This raises a potential concern that the gap between central and (semi)peripheral languages will be exacerbated since the tools to support the former group are already good and are getting better as more attention is being paid to improving them. Meanwhile, tools to support the latter group are developing at a slower pace and are producing lower quality results, which risks widening the gap between the central and (semi)peripheral languages.

Even though MT tools can translate in both directions and could, in principle, be used to support a more multilingual scholarly publishing ecosystem, this does not seem to be what is happening. This in turn, suggests that the mere availability of MT is not sufficient to create or sustain a multilingual scholarly publishing ecosystem. As emphasized by Steigerwald et al. (2022), tool use must be supported by meaningful policies. Currently, the dominant use case for MT in the context of scholarly publishing is to help non-Anglophone scholars publish in English. As a result, the responsibility for multilingualism in scholarly publishing continues to rest on the shoulders of these non-Anglophone scholars while English continues to occupy a central position in the scholarly publishing ecosystem. As argued by Pölönen et al. (2021), policies are needed to incentivize and value research published in other languages. In this way, MT tools could be leveraged for discovering, accessing, and engaging with research that has been written in other languages. These applications, which focus on using MT tools to assimilate information, are generally considered to be a more effective use of this technology than using it to translate a text intended for dissemination or publication (e.g., Hutchins & Somers, 1992; Zulfiquar et al., 2018; Bowker & Buitrago Ciro, 2019).

While MT tools have not yet displaced English from its central position in scholarly publishing, this review nonetheless reveals an emerging appetite for publishing and consulting works in languages other than English. Therefore, there is an opportunity for funding agencies, journals and institutions to develop policies, offer support, and implement value structures to facilitate and appropriately recognize scholarly publishing not only in English but in other languages.

In addition, given that scholars are actively using MT tools for scholarly publishing, journals and funding agencies should develop nuanced policies that take this into account. The tools are not perfect, and there is a need for guidance and policies about responsible and transparent use of translation tools for linguistic support in scholarly publishing. MT users must improve their MT literacy, such as by understanding the potential for data bias, recognizing the limitations of automatic evaluation metrics, having realistic expectations regarding tool capabilities, and knowing how to optimize the tools through improved human-computer interaction (e.g., pre- and post-editing) and identification of good use cases (Bowker, 2020). Training programs or other forms of guidance are needed to support the development of MT literacy skills among scholars.

Finally, given that data-driven tool performance depends not only on having a large quantity of training data but also on having the right kind of training data, there is a role for open access. General-purpose MT tools may not meet specialized needs such as scholarly publishing, so tool (or training dataset) customization is needed. At present, some data-driven tools that are being developed to translate scholarly texts may rely on abstracts or theses/dissertations for training data (Neves et al. 2018; Fiorini 2023). Accessibility of texts for use as training data is undoubtedly a factor: full-text articles are often paywalled, while abstracts and theses/dissertations are more accessible. This could potentially affect translation quality since there could be some degree of mismatch between the characteristics of the texts used to train the translation tools and the characteristics of the texts that scholars wish to translate. A recent deal in which the commercial academic publisher Taylor & Francis sold their paywalled academic content to Microsoft underscores the value of this type of data for machine learning applications (Potter 2024). It also speaks to the need for open LLM-based tools as alternatives to commercial AI tools, but these open tools can only be created if relevant training data is openly available.

Therefore, policies that meaningfully support open access and multilingual scholarly publishing could help to ensure that a greater volume and more diverse range of academic text types and languages are available for inclusion in high-quality open training datasets. Of course, this is a rapidly evolving situation, with new tools (e.g. DeepSeek) and practices emerging, which add to the challenge of developing evidence-based policies and initiatives.

Limitations of the study

One limitation of this study is that the search was conducted in only four languages – English, French, Polish, and Spanish – which correspond to the working languages of the review team. As acknowledged by Niemann Rasmussen and Montgomery (2018), searches that are not multilingual risk overlooking relevant studies, but it can be challenging to implement searches in all languages without significant resources. However, implementing the search in additional languages is worthwhile: we retrieved articles written in other languages because these works contained multilingual metadata (e.g., a Portuguese article was retrieved because it contained metadata in Spanish, which was one of our search languages). MT was then used to translate the additional articles into English for full-text screening. As attested by Zulfiquar et al. (2018), Bowker (2019), and others, while not perfect, MT can be very useful for allowing a reader who is already familiar with a topic to follow the essential points of an article.

Another limitation of our search is that, owing to external constraints, the end date of our search period was September 2023, meaning that translation tools based on large language models are not well covered in this review. Given publication lag, some studies conducted before the introduction of NMT may have been published later. Meanwhile, studies on LLMs conducted in late 2022 or early 2023 were not likely to have been published by the time that data collection for this scoping review ended in September 2023. In addition, LLM-related terms were not explicitly included in the search terms. This is an area undergoing rapid development, and it will be impossible for any review to be completely up to date. However, even understanding the state of a field at a given point in time is useful.

Conclusion

With this scoping review, we set out to obtain a better understanding of whether, why, and how MT is being used in scholarly publishing. After building a query and using it to search nine academic databases in four languages (English, French, Spanish, and Polish), we retrieved a total of 875 works published between January 2017 and September 2023. It is relevant to note that although we conducted our search in four languages and in repositories that contain multilingual content, the overwhelming majority of items retrieved were published in English, which further confirms the central role that this language plays in scholarly publishing. These 875 works were then screened according to a series of inclusion and exclusion criteria intended to ensure that the focus of the study remained firmly at the intersection of MT and scholarly publishing. Thirty-nine works were retained for closer inspection, and these were coded according to a set of pre-determined codes, as well as some additional codes for themes that emerged during the close reading of the texts.

Globally, the studies included in the review reveal that MT is actively used for scholarly publishing, but mainly in support of the continued use of English as a central language for scholarly publishing. Despite its potential for helping to shift this largely monolingual English space towards a more linguistically diverse ecosystem, relatively few studies focus on MT for other languages, and even fewer on low resource languages.

A greater effort is therefore needed to disrupt the centrality of English in the scholarly publishing ecosystem, including the development of language resources for less widely used languages, as well as MT tools customized for the type of specialized translation needed for scholarly publishing. However, as this review demonstrates, the solutions required are not solely technological in nature. Current MT tools, while not perfect, can carry out translation in multiple directions and can support tasks such as discovering and reading research that has been written in other languages. Nevertheless, at present, these tools are mainly being used to translate out of other languages into English, thus reinforcing rather than diversifying this monolingual ecosystem. Providing a detailed and evidence-based response to this question is beyond the scope of this review, although we encourage policy makers to reflect on how practices such as metrics-based research assessments—which prioritize internationalization, citation, impact factor and rankings—put pressure on scholars to participate in scholarly publishing primarily through the medium of English. Following on from our review, we can surmise that while MT tools can be part of the solution, meaningful policies aimed at valuing and incentivizing publication in other languages are needed to achieve and sustain a multilingual scholarly publishing ecosystem.

Acknowledgements

This research is funded by a Knowledge Synthesis Grant (Shifting the Dynamics of Privilege and Marginalization) awarded by the Social Sciences and Humanities Research Council of Canada in partnership with Genome Canada (Grant number 872-2022-1003; Principal Investigator: Lynne Bowker). At the time of data collection, Philips Ayeni and Lynne Bowker were employed at the University of Ottawa, although their respective affiliations changed during the later part of this study.

References

Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., González-Trujillo, J. D., Montaño-Centellas, F., Paudel, K., White, R. L., & Veríssimo, D. (2023). The manifold costs of being a non-native English speaker in science. *PLoS Biology* 27(1): 21(7): e3002184 https://doi.org/10.1371/journal.pbio.3002184

Angulo, E., Diagne, C., Ballesteros-Mejia, L., Adamjy, T., Ahmed, D. A., Akulov, E., Banerjee, A. K., Capinha, C., Dia, C. A. K. M., Dobigny, G., Duboscq-Carra, V. G., Golivets, M., Haubrock, P. J., Heringer, G., Kirichenko, N., Kourantidou, M., Liu, C., Nuñez, M., Renault, D., Roiz, D., Taheri, A., Verbrugge, L. N. H., Watari, Y., Xiong, W., & Courchamp, F. (2021). Non-English languages enrich scientific knowledge: The example of economic costs of biological invasions. *Science of the Total Environment* 775: 144441. https://doi.org/10.1016/j.scitotenv.2020.144441

Bawden, R., Di Nunzio, G. M., Grozea, C., Jauregi Unanue, I., Jimeno Yepes, A., Mah, N., Martinez, D., Névéol, A., Neves, M., Oronoz, M., Perez-de-Viñaspre, O., Piccardi, M., Roller, R., Siu, A., Thomas, P., Vezzani, F., Vicente Navarro, M., Wiemann, D., & Yeganova, L. (2020). Findings of the WMT 2020 Biomedical Translation Shared Task: Basque, Italian and Russian as New Additional Languages. *Proceedings of the Fifth Conference on Machine Translation*, 660–687. https://aclanthology.org/2020.wmt-1.76

Bowker, L. (2023). *De-mystifying translation*. London: Routledge.

Bowker, L. (2020). Chinese speakers' use of machine translation as an aid for scholarly writing in English: A review of the literature and a report on a pilot workshop on machine translation literacy. *Asia Pacific Translation and Intercultural Studies*, *7*(3), 288–298. https://doi.org/10.1080/23306343.2020.1805843

Bowker, L. (2019). Machine translation literacy: Academic libraries' role. *Proceedings of the Association for Information Science and Technology*, 56: 618-619. https://doi.org/10.1002/pra2.108

Bowker, L. (2018). Machine translation and author keywords: A viable search strategy for scholars with limited English proficiency? *Advances in Classification Research Online*, 13–16. https://doi.org/10.7152/acro.v29i1.15455

Bowker, L., & Buitrago-Ciro, J. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Bingley: Emerald.

Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research. 11th Conference of the European Chapter of the Association for Computational Linguistics, 249-256. https://aclanthology.org/E06-1032/

Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to Human and Machine Translation Quality Assessment. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 9–38). Springer International Publishing. https://doi.org/10.1007/978-3-319-91241-7_2

Chang, C.-M., Chang, C.-H., & Hwang, S.-Y. (2020). Employing word mover's distance for cross-lingual plagiarized text detection. *Proc Assoc Inf Sci Technol. 2020;57:e229* https://doi.org/10.1002/pra2.229

Commissaire à la langue française. (2023). Le français, langue du savoir? Pour une approche structurée de l'usage de la traduction automatique dans le milieu scientifique. https://commissairelanguefrancaise.quebec/publications/avis/francais-traduction-milieu-scientifique.pdf

Crilly, T., Jashapara, A., Ferlie, E. (2010). Research utilisation and knowledge mobilisation: a scoping review of the literature. *National Institute for Health Research Service Delivery and Organisation programme*.

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

Decaria, J., Sharp, C., & Petrella, R. (2012). Scoping review report: obesity in older adults. *International Journal of Obesity 36*,1141–1150.

Dobrynina, O. L. (2021). Academic writing for publication purposes and machine translation: is the symbiosis possible? *Www. Vovr. Elpub. Ru; Www. Vovr. Ru 1992* , *30*(12), 88.

Esmailpour, R., Ebrahimy, S., Fakhrahmad, S. M., Mohammadi, M., & Abbaspour, J. (2020). Developing an effective scheme for translation and expansion of Persian user queries. *Digital Scholarship in the Humanities*, *35*(3), 493–506. https://doi.org/10.1093/llc/fqz041

Fadaee, M., & Monz, C. (2022). The unreasonable volatility of neural machine translation models. *Proceedings of the 4th Workshop on Neural Generation and Translation (WNGT 2020)*, 88–96 Online, July 10, 2020. https://aclanthology.org/2020.ngt-1.10.pdf

Fiorini, S. (2023). Exploratory studies for the creation of a technology-aided collaborative translation service in open scholarly communication. Zenodo. https://doi.org/10.5281/zenodo.10972986

Forcada, M.L. (2017). Making sense of neural machine translation. *Translation Spaces* 6(2) : 291-309.

Hanna, M., and Bojar, O. (2021). A fine-grained analysis of BERTScore. *Proceedings of the Sixth Conference on Machine Translation* (WMT), 507–517. https://aclanthology.org/2021.wmt-1.59/

Hannah, K., Haddaway, N. R., Fuller, R. A., & Amano, T. (2024) Language inclusion in ecological systematic reviews and maps: Barriers and perspectives. Research Synthesis Methods 15(3): 466-482. https://doi.org/10.1002/jrsm.1699

Hawker, S., Payne, S., Kerr, C., Hardey, M., & Powell, J. (2002). Appraising the evidence: reviewing disparate data systematically. *Qualitative Health Research*, *12*(9), 1284-1299. https://doi.org/10.1177/1049732302238251

Helsinki Initiative. (2019). Helsinki Initiative on Multilingualism in Scholarly Communication. Helsinki: Federation of Finnish Learned Societies, Committee for Public Information, Finnish Association for Scholarly Publishing, Universities Norway & European Network for Research Evaluation in the Social Sciences and the Humanities. https://doi.org/10.6084/m9.figshare.7887059

Hughes, Alex. (2023, 25 September). ChatGPT: Everything you need to know about OpenAI's GPT-4 tool. *BBC Science Focus*. https://www.sciencefocus.com/future-technology/gpt-3

Hutchins, W. J., & Somers, H. L. (1992). *An Introduction to Machine Translation*. London: Academic Press.

Kim, E.-Y. J., & LaBianca, A. S. (2018). Ethics in Academic Writing Help for International Students in Higher Education: Perceptions of Faculty and Students. *Journal of Academic Ethics*, *16*(1), 39–59. https://doi.org/10.1007/s10805-017-9299-5

Koehn, P. (2020). *Neural machine translation*. Cambridge: Cambridge University Press.

Koehn, P. (2010). *Statistical machine translation*. Cambridge: Cambridge University Press.

Kostadinova, D. (2019). On some particular cases of translation and self-translation. *- Orbis Linguarum*, *17*(2), 86–92.

Lin, L. H. F., & Morrison, B. (2021). Challenges in academic writing: Perspectives of Engineering faculty and L2 postgraduate research students. *English for Specific Purposes*, *63*, 59–70. https://doi.org/10.1016/j.esp.2021.03.004

Matsumura, Y., Katsumata, S., & Komachi, M. (2018). TMU Japanese-English neural machine translation system using Generative Adversarial Network for WAT 2018. *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation: 5th Workshop on Asian Translation: 5th Workshop on Asian Translation*. PACLIC 2018, Hong Kong. https://aclanthology.org/Y18-3014

Mino, H., Kinugawa, K., Ito, H., Goto, I., Yamada, I., & Tokunaga, T. (2021). NHK's lexically-constrained neural machine translation at WAT 2021. *Proceedings of the 8th Workshop on Asian Translation (WAT2021)*, 46–52. https://doi.org/10.18653/v1/2021.wat-1.2

Moorkens, J., Way, A., & Lankford, S. (2024). *Automating Translation*. London: Routledge.

Morishita, M., Suzuki, J., & Nagata, M. (2019). NTT Neural Machine Translation Systems at WAT 2019. *Proceedings of the 6th Workshop on Asian Translation*, 99–105. https://doi.org/10.18653/v1/D19-5211

Nayak, P., Haque, R., & Way, A. (2020, November). *The ADAPT's submissions to the WMT20 biomedical translation task*. The Fifth Conference on Machine Translation (The Biomedical Shared Task), Dominican Republic (Online). https://aclanthology.org/2020.wmt-1.91.pdf

Neves, M., Jimeno Yepes, A., Névéol, A., Grozea, C., Siu, A., Kittner, M., & Verspoor, K. (2018). Findings of the WMT 2018 Biomedical Translation Shared Task: Evaluation on Medline test sets. *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, 324–339. https://doi.org/10.18653/v1/W18-6403

Neimann Rasmussen, L., & Montgomery, P. (2018). The prevalence of and factors associated with inclusion of non-English language studies in Campbell systematic reviews: A survey and meta-epidemiological study. *Systematic Reviews*, *7*(1), 129. https://doi.org/10.1186/s13643-018-0786-6

O'Brien, S., Simard, M., & Goulet, M.-J. (2018). Machine translation and self-post-editing for academic writing support: Quality Explorations. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 237–262). Springer International Publishing. https://doi.org/10.1007/978-3-319-91241-7_11Peters, M. D., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation*, *13*(3), 141-146.

Pérez-Ortiz, J.-A., Forcada, M. L., & Sánchez-Martínez, F. (2022). How neural machine translation works. In D. Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 141-164. Berlin: Language Science Press. http://doi.org/10.5281/zenodo.6760020

Pölönen, J., Kulczycki, E., Mustajok, H., and Røeggen, V. (2021). Multilingualism is integral to accessibility and should be part of European research assessment reform. *LSE Impact Blog*, December 7. https://blogs.lse.ac.uk/impactofsocialsciences/2021/12/07/multilingualism-is-integral-to-accessibility-and-should-be-part-of-european-research-assessment-reform/

Potter, W. (2024, July 23). An academic publisher has struck an AI deal with Microsoft—without their authors' knowledge. *The Conversation*. https://theconversation.com/an-academic-publisher-has-struck-an-ai-data-deal-with-microsoft-without-their-authors-knowledge-235203

Roussis, D., Papavassiliou, V., Prokopidis, P., Piperidis, S., & Katsouros, V. (2022). SciPar: A Collection of Parallel Corpora from Scientific Abstracts. *Proceedings of the Thirteenth Language Resources and Evaluation Con-*

ference, 2652–2657. https://aclanthology.org/2022.lrec-1.284

Sel, İ., & Hanbay, D. (2022). Fully attentional network for low-resource academic machine translation and post editing. *Applied Sciences*, *12*(22), Article 22. https://doi.org/10.3390/app122211456

Soares, F., Moreira, V. P., & Becker, K. (2019). *A large Parallel Corpus of full-text scientific articles* (arXiv:1905.01852). arXiv. https://doi.org/10.48550/arXiv.1905.01852

Soares, F., Tateisi, Y., Takatsuki, T., & Yamaguchi, A. (2021). O-JMeSH: Creating a bilingual English-Japanese controlled vocabulary of MeSH UIDs through machine translation and mutual information. *Genomics & Informatics*, *19*(3), e26. https://doi.org/10.5808/gi.21014

Soares, F., Yamashita, G. H., & Anzanello, M. J. (2018). A Parallel Corpus of theses and dissertations abstracts. In A. Villavicencio, V. Moreira, A. Abad, H. Caseli, P. Gamallo, C. Ramisch, H. Gonçalo Oliveira, & G. H. Paetzold (Eds.), *Computational Processing of the Portuguese Language* (pp. 345–352). Springer International Publishing. https://doi.org/10.1007/978-3-319-99722-3_35

Steigerwald, E., Ramírez-Castañeda, V., Brandt, D. Y. C., Báldi, A., Shapiro, J. T., Bowker, L., & Tarvin, R. D. (2022). Overcoming Language Barriers in Academia: Machine Translation Tools and a Vision for a Multilingual Future. *Bioscience* 72 (10): 988–98. https://doi.org/10.1093/biosci/biac062.

St-Onge, S., Forgues, É., Larivière, V., Riddles, A. & Volkanova, V. (2021). *Portrait et défis de la recherche en français en contexte minoritaire au Canada*. Acfas. https://www.acfas.ca/sites/default/files/documents_utiles/rapport_francophonie_final_1.pdf

Sun, Y.-C., & Yang, F.-Y. (2023). Exploring the process and strategies of Chinese–English abstract writing using machine translation tools. *Journal of Scholarly Publishing*, *54*(2), 260–289. https://doi.org/10.3138/jsp-2022-0039

Sun, Y.-C., Yang, F.-Y., & Liu, H.-J. (2022). Exploring Google Translate-friendly strategies for optimizing the quality of Google Translate in academic writing contexts. *SN Social Sciences*, *2*(8), 147. https://doi.org/10.1007/s43545-022-00455-z

Takakusagi, Y., Oike, T., Shirai, K., Sato, H., Kano, K., Shima, S., Tsuchida, K., Mizoguchi, N., Serizawa, I., Yoshida, D., Kamada, T., & Katoh, H. (2021). Validation of the reliability of machine translation for a medical article from Japanese to English using DeepL Translator. *Cureus*. https://doi.org/10.7759/cureus.17778

Takeshita, S., Green, T., Friedrich, N., Eckert, K., & Ponzetto, S. P. (2022). X-SCITLDR: Cross-lingual extreme summarization of scholarly documents. *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries*, 1–12. https://doi.org/10.1145/3529372.3530938

Tehseen, I., Tahir, G. R., Shakeel, K., & Ali, M. (2018). Corpus based machine translation for scientific text. In L. Iliadis, I. Maglogiannis, & V. Plagianakos (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 196–206). Springer International Publishing. https://doi.org/10.1007/978-3-319-92007-8_17

Thomas, J and Harden, A (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology 8*(1), 1–10.

Tongpoon-Patanasorn, A., & Griffith, K. (2020). Google Translate and translation quality: A Case of translating academic abstracts from Thai to English. *PASAA: Journal of Language Teaching and Learning in Thailand*, *60*, 134–163.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, *169*(7), 467-473.

UNESCO. (2021). Recommendation on Open Science. https://unesdoc.unesco.org/ark:/48223/pf0000378841

Wahab, M. F., Zulfiqar, S., Sarwar, M. I., & Lieberwirth, I. (2020). Facile solutions to the problems associated with Chemical information and mathematical symbolism while using machine translation tools. *Journal of Chemical Information and Modeling*, *60*(7), 3423–3430. https://doi.org/10.1021/acs.jcim.0c00274

Windsor, L. C., Cupit, J. G., & Windsor, A. J. (2019). Automated content analysis across six languages. *PLOS ONE*, *14*(11), e0224425. https://doi.org/10.1371/journal.pone.0224425

Winiharti, M., & Sudana, D. (2021). The English Google translation of Indonesian lecturer's academic writing: A preliminary study. *Journal of Language and Linguistic Studies*, *17*(2), 706–719. https://doi.org/10.3316/informit.215875937433823

Xie, Q., Zhang, X., Ding, Y., & Song, M. (2020). Monolingual and multilingual topic analysis using LDA and BERT embeddings. *Journal of Informetrics*, *14*(3), 101055. https://doi.org/10.1016/j.joi.2020.101055

Xu, J., Abdul Rauf, S., Pham, M. Q., & Yvon, F. (2021). LISN @ WMT 2021. *6th Conference on Statistical Machine Translation*. https://hal.science/hal-03430610

Yamamoto, S., Suzuki, R., Fukusato, T., Kataoka, H., & Morishima, S. (2021). A case study on user evaluation of scientific publication summarization by Japanese Students. *Applied Sciences*, *11*(14), Article 14. https://doi.org/10.3390/app11146287

Zhang, B., & Misra, A. (2023). *Machine translation impact in e-commerce multilingual search* (arXiv:2302.00119). arXiv. https://doi.org/10.48550/arXiv.2302.00119

Zhivotova, A. A., Berdonosov, V. D., & Redkolis, E. V. (2020). Improving the quality of scientific articles Machine Translation while writing original text. *2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, 1–4. https://doi.org/10.1109/FarEastCon50210.2020.9271442

Zomer, G., & Frankenberg-Garcia, A. (2021). Beyond grammatical error correction: Improving L1-influenced research writing in English using pre-trained encoder-decoder models. *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2534–2540.

Zou, C., Gong, W., & Li, P. (2023). Using online machine translation in international scholarly writing and publishing: A longitudinal case of a Chinese engineering scholar. *Learned Publishing*, *n/a*(n/a). https://doi.org/10.1002/leap.1565

Zulfiqar, S., Wahab, M. F., Sarwar, M. I., & Lieberwirth, I. (2018). Is Machine Translation a reliable tool for reading German scientific databases and research articles? *Journal of Chemical Information and Modeling*, *58*(11), 2214–2223. https://doi.org/10.1021/acs.jcim.8b00534

Appendix A. Appendix A. ChatGPT prompts, process, and outcomes of query translation

Prompts for ChatGPT

- You are an academic database query tool.
- Reconstruct the following query for the Web of Science database but substitute the English keywords with their French equivalents.
- (TS=("automatic translation*" OR "automatic translator*" OR DeepL OR "DeepL Translator" OR "Google Translate" OR "Google translator" OR "machine translation*" OR "machine translator*" OR "MT" OR "neural machine translation*" OR "online translator*" OR "post-edit*" OR "translation engine*" OR "translation system*" OR "translation technolog*" OR "translation tool*")) AND (TS=("abstract*" OR "academic abstract*" OR "academic article*" OR "academic literature" OR "academic paper*" OR "academic publication*" OR "academic publishing" OR "academic writing" OR "journal article*" OR "journal publication*" OR "research article*" OR "research paper*" OR "research publication*" OR "science writing" OR "scientific abstract*" OR "scientific article*" OR "scientific literature" OR "scientific paper*" OR "scientific publication*" OR "scientific text*" OR "scholarly communication" OR "scholarly publication*" OR "scholarly publishing" OR "scholarly writing" OR "writing for publication")) AND PY=(2017-2023)

LB then edited the resulting query from ChatGPT (e.g., to add asterisks to account for agreements in French, or to add additional terms that are relevant for French). She then used the corrected French WoS search query as input to get the French-language Scopus query by continuing the chain of prompts.

- Reformulate the following Web of Science query for Scopus.
- (TS=("traduction automatique" OR "traducteur* automatique*" OR DeepL OR "Traducteur DeepL" OR "Google Traduction" OR "traducteur Google" OR "traduction automatique neuronale" OR "traducteur* en ligne" OR "post*édit*" OR "moteur* de traduction" OR "système* de traduction" OR "technologie* de traduction" OR "technologie* de la traduction" OR "outil* de traduction")) AND (TS=("résumé* académique*" OR "article* académique*" OR "littérature académique" OR "article* académique*" OR "publication* académique*" OR "publication* universitaire*" OR "rédaction académique" OR "article* de revue*" OR "publication* de revue*" OR OR "article* de recherche*" OR "article* scientifique*" OR "publication* scientifique*" OR "écriture* scientifique*" OR "résumé* scientifique*" OR "article* scientifique*" OR "littérature scientifique" OR "article* universitaire*" OR "article* savant*" OR "communication* savante*" OR "publication* savante*" OR "édition savante" OR "rédaction savante" OR "écriture pour la publication")) AND PY=(2017-2023)

The same process was then repeated for Spanish and Polish. Note that EK is a native speaker of Polish and also fluent in English and Spanish. Meanwhile, as a certified translator, LB knows both French and Spanish well, in addition to being a native English speaker. Therefore, we feel comfortable that these queries are correct and well formed. The translations were used to search other databases that are in languages other than English. For instance, the French translation of the search query was employed to search the Érudit database, while the Spanish translation was employed to search the Redalyc database.

Appendix B. Codebook of thematic analysis of included studies

Themes and sub-themes	Description	Files	References
1. Purpose of the study	This theme discusses the purpose and main focus of the included studies.	38	55
a. Compiling resources for tools to use (e.g., corpora of scholarly texts)	Compiling resources such as the large corpora of source- and target-language texts needed to feed data-driven translation tools.	0	0
b. Developing translation tools	Developing translation tools specifically for scholarly communication.	13	15
c. Discovering scholarly content	Using translation tools to discover scholarly content (e.g., searching for articles in other languages).	3	3
d. Consuming scholarly content	Using translation tools to consume scholarly content (e.g., reading articles that were originally written in a language that you don't understand well; following a conference presentation in a language that you don't understand well (e.g., via subtitles)).	1	3
e. Producing scholarly content	Using translation tools to help produce scholarly content (e.g., writing articles or preparing conference presentations in a language that is not your dominant language).	10	13
f. Evaluating translation quality	Evaluating the quality of texts that have been translated automatically by translation tools.	16	21
2. Tools	This theme discusses the type of tool that is the focus of the included studies.	34	42
a. Rule-based machine translation (RBMT) tool	Machine translation tool with an underlying architecture based on linguistic grammars and bilingual dictionaries.	0	0
b. Statistical machine translation (SMT) tool	Machine translation tool with an underlying architecture based on statistical processing (e.g., versions of Google Translate used prior to 2017).	3	3
c. Neural machine translation (NMT) tool	Machine translation tool with an underlying architecture based on neural networks and machine learning (e.g., DeepL Translator, versions of Google Translate produced from 2017 on).	33	38
d. Large Language Model (LLM)	Probabilistic model of language based on a massive corpus of texts used to train the model (e.g., ChatGPT, GPT-4, Bard, Bing AI chatbot)	1	1
e. Terminology extraction tool	Tool that attempts to automatically identify and extract specialized terms from a text.	0	0
f. Cross-language information retrieval tool	Tool that allows a user to enter a search term in one language and retrieve texts written in another language.	1	2
3. Languages	This theme focuses on the languages and language resources (e.g., corpora) used in the research project. Because we are dealing with translation, there are typically two languages involved, although some projects may involve more than two.	32	36

a. High-resource language or language pair	Languages that are widely used or between which there is a lot of translation activity are described as high-resource because it is easy to build a large high-quality bilingual training corpus for the tool to learn from. English, French and Spanish are examples of high-resource languages; the combination of English-French is a high-resource language pair.	25	29
b. Low-resource language or language pair	Languages that are less widely used or between which there is a limited amount of translation activity are described as low-resource because it is more challenging to build a large high-quality bilingual training corpus for the tool to learn from. Ukrainian, Welsh and Greek are examples of low-resource languages; the combination Ukrainian-Welsh is a low-resource language pair. Sometimes there could be one high-resource and one low-resource language in the pair (e.g., English-Welsh), and this is generally still described as a low-resource situation overall.	7	7
4. Text types	These are the types of texts that the researchers are working with (e.g., a machine translation tool that uses a training corpus of dissertations, or a machine translation tool intended to translate scientific abstracts).	31	42
a. Scientific abstract	Short summary of a research article that is part of the metadata for the article (along with the title and keywords).	9	12
b. Scholarly article or chapter	A full-text research paper of approximately 10,000 words published in a scholarly journal or edited volume.	13	15
c. Book	Monograph of approximately 50,000 to 200,000 words.	0	0
d. Thesis or dissertation	Scholarly work of approximately 50,000 to 100,000 words submitted in partial fulfilment for a graduate research degree.	4	4
e. Conference presentation	Write-up of a presentation delivered at an academic conference.	0	0
f. Popularized text (e.g., plain language summary)	Description of research that is written in language that is accessible to non-experts.	0	0
g. Keywords (index terms)	List of approximately 5 to 10 terms that describe the content of scholarly work that form part of the work's metadata and are used to index the work.	2	3
h. Scientific terms	Individual specialized terms appearing in a list of keywords or in a scholarly text.	1	2
i. Coursework	Texts produced as part of an academic course (e.g., essays or research papers for graduate courses).	1	1
g. Scientific/technical texts	Scholarly texts other than articles or chapters that contain scientific or technical content (e.g., technical reports).	5	5
5. Evaluation methods	This theme describes the approach that authors used to evaluate the quality of texts translated by a translation tool.	27	40
a. Automatic evaluation	Using one or more automated metrics (e.g., BLEU, METEOR) that have been previously validated.	18	24
b. Manual evaluation	Employing subject experts or human translators.	15	16
6. Evaluation outcomes	This theme contains the main findings of the article, or the outcome of the translation quality assessment.	37	60
a. Strengths of the tools	Positive evaluation of the translation tools used in the included studies.	18	27

b. Limitations of the tools	Drawbacks and limitations of the translation tool used for translating scholarly texts in included studies.	10	14
c. Reception of the tools	User satisfaction with regard to the translation tool in the included studies.	6	9
d. Strategies for integrating the tools	Strategies for integrating and optimizing translation tools in scholarly communication.	7	10

Appendix C. Characteristics of included studies

First Author	Year	Country	Publication	Methods	Language(s)	Text type(s)	Translation tools
Bawden, R.	2020	UK	Conference paper	Mixed-methods	English (source), target languages: Basque, Chinese, French, German, Italian, Portuguese, Russian, Spanish	Scientific abstracts from Medline database, and Basque medical journal	Custom-built prototypes
Bowker, L.	2020	Canada	Journal article	Workshop	Chinese, English	Not reported	Free online MT systems
Bowker, L.	2019	Canada	Conference paper	Workshop	Any (not specified)	Scientific articles	Any free online MT systems
Bowker, L.	2018	Canada	Conference paper	Experiment	French, Spanish, English	71 Scientific keywords and 43 French keywords from the library and information science domain	Google Translate (NMT version), DeepL Translator
Chang, C.-M.	2020	Taiwan	Conference paper	Experiment	English, Chinese	Scientific abstracts and theses/dissertations	Google Translate
Daniele, F.	2019	Italy	Journal article	Experiment	English, Italian	111 scientific abstracts from PubMed	Google Translate
Dobrynina, O. L.	2021	Russia	Journal article	Mixed-methods	Russian, English	Scientific texts	Custom-built prototype - AiGobex
Esmailpour, R.	2020	Iran	Journal article	Experiment	English, Farsi	Bibliographic data from Persian journals that also have English metadata	Google Translate
Kim, E.-Y. J.	2018	USA	Journal article	Survey of 160 participants	Various (24 represented in student sample)	Scientific thesis/research paper	Not specified (just “machine translation”)
Kostadinova, D.	2019	Bulgaria	Journal article	Experiment	Bulgarian, English	Scientific texts	Microsoft Translator, Google Translate

Lin, L. H. F.	2021	China	Journal article	Survey of 110 participants	Chinese, English	Scientific papers in Engineering	Google Translate
Matsumura, Y.	2018	Japan	Conference paper	Experiment	English, Japanese	Scientific texts	Custom-built NMT prototype tool
Mino, H.	2021	Japan	Workshop report	Experiment	English, Japanese	Scientific abstract	Custom-built NMT prototype tool
Morishita, M.	2019	Japan	Workshop report	Experiment	English, Japanese	Scientific papers	Custom-built prototype
Nayak, P.	2019	Ireland	Conference paper	Experiment	English to Basque	Biomedical texts	Custom-built prototype
Neves, M.	2018	Germany	Conference paper	Experiment	English (source), target languages: Chinese, French, German, Portuguese, Romanian, Spanish	Scientific abstracts from Medline and EDP database	Moses SMT, OpenNMT, and various custom-built prototypes
O'Brien, S.	2018	Ireland	Book chapter	Experiment	Source languages: Arabic, Chinese, French, German, Romanian, Spanish; target language: English	Academic abstracts from various fields (biotechnology, engineering, chemistry, geology, marketing, psychology, social sciences)	Google Translate (NMT version)
Roussis, D.	2022	Greece	Conference paper	Experiment	31 language pairs	Scientific abstracts from theses and dissertations in 86 European repositories	Data-driven MT
Sel, İ.	2022	Turkey	Journal article	Experiment	Turkish, English	Scientific abstracts from 245,100 theses in the Turkish CoHe thesis database.	Google Translate
Soares, F.	2019	Brazil	Conference paper	Experiment	English, Spanish and Portuguese	Scientific articles from SciELO	Moses SMT system

Soares, F.	2021	UK	Journal article	Experiment	English, Japanese	Scientific articles on COVID-19	Google Translate, Microsoft Bing Translator
Soares, F.	2018	Brazil	Conference paper	Experiment	English and Portuguese	Scientific abstracts from theses and dissertations in CAPES TDC (Thesis and Dissertation Catalogue)	Moses SMT system, OpenNMT system, Google Translate
Sun, Y.-C.	2022	Taiwan	Journal article	Experiment	Chinese, English	Scientific abstracts of scholarly articles in the domain of language teaching and learning.	Google Translate
Sun, Y.-C.	2023	Taiwan	Journal article	Mixed-methods	Chinese English	Academic abstracts	Google Translate
Takakusagi, Y	2021	Japan	Journal article	Experiment	English and Japanese	Scientific text	DeepL Translator
Takeshita, S.	2022	Germany	Conference paper	Experiment	English (source); target: German, Italian, Chinese, Japanese	Scientific articles	DeepL Translator, LLM (BART)
Tehseen, I.	2018	Pakistan	Book chapter	Experiment	English, Urdu	Scientific terminology of the field of computer science.	Custom-built term tagger and term translator, and Google Translate
Tongpoon-Patanasorn, A.	2020	Thailand	Journal article	Experiment	Thai, English	54 Scientific abstracts from Thai-language journals in 8 disciplines in the humanities and social sciences.	Google Translate (NMT version)
Wahab, M. F.	2020	USA	Journal article	Experiment	English, French, German	Scientific articles in the domain of chemistry.	Google Translate, DeepL Translator
Windsor, L. C.	2019	USA	Journal article	Experiment	English, French, German, Russian, Arabic, Chinese	MultiUN Corpus (United Nations documents)	Google Translate SMT version

Winiharti, M.	2021	Indonesia	Journal article	Experiment	Indonesian to English	3 Scientific articles in Japanese, management, math	Google Translate
Xie, Q.	2020	South Korea	Journal article	Experiment	English, Chinese	Scientific articles	Custom-built prototype MT tool
Xu, J.	2021	France	Conference paper	Experiment	English and French	Scientific abstracts in the domain of biomedicine.	Custom-built prototype MT tool
Yamamoto, S.	2021	Japan	Journal article	Experiment	English to Japanese	Summaries of Scientific articles	Google Translate
Zhang, B.	2023	Unknown	Conference paper	Experiment	English, Spanish, Portuguese, French, Korean, Malayam, German, Japanese, Dutch, Turkish, Kannada	Historical search queries in the field of e-Commerce	Custom built prototype
Zhivotova, A. A.	2020	Russia	Conference paper	Experiment	Russian, English	Scientific abstracts on articles about unmanned aerial vehicles	Google Translate, DeepL Translator, Amazon Translate
Zomer, G.	2021	UK	Conference paper	Experiment	English, Spanish, Portuguese	Scholarly publications	Custom-built language checker
Zou, C.	2023	China	Journal article	Case study	Chinese to English	Academic articles on engineering.	Google Translate, also mentions Baidu Translate
Zulfiqar, S.	2018	Egypt	Journal article	Experiment	German, English	Scientific texts on various subfields of chemistry from German-language academic databases.	Google Translate (NMT version) and DeepL Translator