uOttawa

UAM
ADAM MICKIEWICZ UNIVERSITY POZNAŃ

Linguistic privilege and marginalization in scholarly communication: Understanding the role of new language technologies for shifting language dynamics

Final report, 17 December 2023

https://doi.org/10.20381/858s-q632

Lynne Bowker, Professor, School of Translation and Interpretation, University of Ottawa, Canada (lbowker@uottawa.ca)

Philips Ayeni, Postdoctoral Researcher, School of Information Studies, University of Ottawa, Canada (payeni@uottawa.ca)

Emanuel Kulczycki, Professor, Scholarly Communication Research Group, Adam Mickiewicz University, Poland (emanuel.kulczycki@amu.edu.pl)

This project is funded by the Social Sciences and Humanities Research Council in partnership with Genome Canada. Ce projet est financé par le Conseil de recherches en sciences humaines en partenariat avec Génome Canada.

Social Sciences and Humanities Research Council of Canada | Conseil de recherches en sciences humaines du Canada

Canada

GenomeCanada

Table of Contents

Executive Summary iv

1. Background 1

    1.1 Scholarly communication 1

    1.2 Translation technologies 1

2. Objectives 2

3. Methods 3

    3.1 Search Process 3

        3.1.1. Selection of search terms 4

    3.2 Selection Process 5

        3.2.1 Inclusion criteria 5

        3.2.2 Exclusion criteria 6

    3.3 Search and screening outcomes 6

    3.4 Full-text screening outcome 7

    3.5 Data extraction and coding approach 9

    3.6 Qualitative analysis of included studies 9

4. Results 17

    4.1 Study characteristics 17

    4.2 Purpose of the study 18

    4.3 Tools 18

    4.4 Languages 20

    4.5 Text types 21

    4.6 Evaluation methods 22

    4.7 Evaluation outcomes 23

        4.7.1 Strengths and limitations of the tools 23

        4.7.2 Reception of the tools 25

4.7.3 Strategies for integrating the tools 25
5. Implications..... 26
5.1 Policy implications..... 26
5.2 Practical implications..... 27
6. Conclusions..... 28
6.1 Future areas of research 29
7. Knowledge Mobilization Activities 30
7.1 Researchers 30
7.2 Editors, publishers and policy makers 31
7.3 Wider society 31
Bibliography 32
Appendix A: ChatGPT prompts, process, and outcomes of query translation..... 37
Appendix B: Codebook of thematic analysis of included studies..... 38
Appendix C: Distribution of articles by language and translation direction..... 41

List of Tables

Table 1: Search strategy 4
Table 2: Characteristics of included studies..... 11

List of Figures

Figure 1: Search query sample from Scopus database..... 5
Figure 2: PRISMA protocol of search process and outcomes..... 8
Figure 3: Word cloud of the 100 most frequent words in the included studies..... 10
Figure 4: Countries of publication of the included studies 17
Figure 5: Distribution of studies involving English by translation direction..... 21

Executive Summary

Background: the issue

English plays the key role in scholarly communication and resource distribution, but using a single language for research 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. This may result in a lower volume of research output, which could negatively affect career advancement. Meanwhile, native English-speaking scholars, who come mainly from Western cultures, will have more visibility and power, and this could influence which subjects are investigated and which communities benefit from the results. Their higher visibility and volume of output may also improve career advancement opportunities for English speakers, such as prestigious appointments as journal editors, where they may reinforce Western viewpoints. To create a more equitable and linguistically diverse scholarly communication ecosystem, support is needed to ensure that scholars can contribute research in different languages and still be able to discover and read each other's work.

Objectives

Language technologies such as machine translation (MT) tools (e.g., Google Translate) or the recent AI tools based on large language models (LLMs) (e.g., ChatGPT) could potentially play a role, but little is understood about whether, how and to what effect these tools are used by scholars. The overarching objective of this systematic review is to understand more about how translation tools are being used for scholarly communication. More specifically, this review will reveal to what extent and how translation tools are being used, as well as which tools, languages, and text types are in play – and which are not. To the extent that it is possible based on the studies included in the review, we also aim to learn more about the strengths and limitations of different tools, user's attitudes towards them, and strategies for integrating them more effectively into scholarly communication.

Results

Translation tools are actively being used in the context of scholarly communication, but their use favours some text types, languages, and purposes. For instance, the most commonly used type of translation tool is neural machine translation, while tools such as cross-lingual information retrieval tools, speech processing tools, subtitling tools or terminology extractors receive scant attention. Translation tools are used mainly to translate texts from other languages into English with a view to publishing these texts in English. There is little evidence that translation tools are used to search for or read texts in languages other than English. Moreover, text types such as slides, posters, or popularized texts are not discussed in the literature, nor are potential ways of leveraging translation tools to support conference presentations, interviews, or discussions (e.g., by using translated subtitles or translated transcriptions). While some research is addressing low-resource languages, more focus is placed on high-resource languages. Strengths, limitations, and quality of tools are difficult to generalize because they vary depending on text type, language combination, and subject matter. Therefore, reactions to the use of these tools also vary. Users do see potential for the tools, but they note the need to develop strategies to optimize quality.

Key messages

Translation technologies are being used for scholarly communication, but they have not reached their full potential for helping to create a more linguistically diverse scholarly communication ecosystem. At present, the main type of tool used is neural machine translation, which is employed primarily for translating written texts from other languages into English for publication as journal articles or book chapters, rather than to discover or read works that have been written in languages other than English. Little attention has been paid to how translation tools can be used to translate other types of texts (e.g., posters, slides, popularized texts) or to support oral forms of scholarly communication (e.g., conference presentations, interviews). More research focuses on high-resource (i.e., widely used) languages and less on low-resource languages, which risks widening the gap since data-driven tools/resources that already work relatively well are being perfected, while those that work less well are receiving less attention.

In the ways that they are currently being developed and applied, translation tools are serving to reinforce a monolingual English scholarly communication ecosystem, where non-Anglophones bear the responsibility for translating unidirectionally out of their languages and into English. Translation technologies can do more to facilitate linguistic diversity, but tools alone are not sufficient to shift the dynamics of linguistic privilege and marginalization in scholarly communication; technology must be supported by policies that value and promote multilingualism.

Methodology

The project team undertook a systematic review of the literature to investigate the current state of translation technologies within the scholarly communication ecosystem. The PRISMA protocol was adopted to guide the identification and selection of scholarly works for review. Following a multilingual search (English, French, Spanish, and Polish) in nine key research databases, 875 works were retrieved and screened according to a set of inclusion/exclusion criteria, and 40 were retained for closer investigation. These works were coded for themes that emerged from the reading. The themes were analyzed and discussed in relation to the topic of understanding the role of translation technology for shifting the dynamics of linguistic privilege and marginalization in scholarly communication.

1. Background

1.1 Scholarly communication

For decades, English has occupied a privileged position in scholarly communication. However, the seemingly solid foundation of using one main language for research has now begun to give way as scholars who speak other languages are increasingly advocating for multilingual scholarly communication. In principle, there is a logic to using one language for research (i.e., everyone who knows the language can participate in the conversation), but in practice, the single-language model has inequities. For instance, it requires more time and effort for non-Anglophones to read, publish, or present in English (Amano et al., 2023), and English may be used as a gatekeeper to exclude contributions from speakers of other languages (Habibie & Hultgren, 2022). As a result, Western viewpoints have higher visibility, and this may influence what gets studied, which methods are used, where findings are shared, and which communities benefit from the research (Angulo et al., 2021). The movement to create a multilingual scholarly communication ecosystem is gaining traction, with support coming from groups such as the Helsinki Initiative on Multilingualism in Scholarly Communication (Helsinki Initiative, 2019), UNESCO (2021), OPERAS SIG on Multilingualism (Balula, 2021), Acfas (St.-Onge et al., 2021), and others. Yet, to create a viable multilingual scholarly communication ecosystem, scholars will need practical support to find, read or write scholarly works in other languages. Can translation technologies go some way towards meeting this need?

1.2 Translation technologies

A wide variety of computer-assisted tools and resources exist to support the work of professional translators, such as electronic dictionaries or translation memories (Bowker, 2002). However, automatic translators or **machine translation tools** are computer tools that undertake the actual task of converting a text from one language to another. Machine translation tools have existed for decades, but early systems based on linguistics had very limited success (Hutchins & Somers, 1991). In recent years, the introduction of AI-based data-driven tools (e.g., Google Translate, DeepL Translator, or ChatGPT) has brought dramatic improvements to the quality of machine-translated text, making it a viable starting point for many purposes. But how viable is it in the context of scholarly communication?

Data-driven tools that use machine learning require massive amounts of training data to learn how to accomplish a task. To learn how to translate, an AI tool needs be trained on billions of pages of text that have been translated previously by professional translators (Forcada, 2017). This training material takes the form of a parallel corpus, which 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, 2022). However, for languages that are less widely used, such as many of Canada's Indigenous languages, or languages such as Welsh or Danish, it is more challenging to find a large collection of previously translated texts—especially if the translation is between two languages that are not widely used. This situation is referred to as a **low-resource situation** (Pérez-Ortiz, 2022). The concept of high- and low-resource applies not only to languages but also to domains and text types. For instance, some domains, such as administration, are commonly discussed. In contrast, other domains are highly specialized, such as nuclear physics. Likewise, some text types are easy to access (e.g., scientific abstracts), while access to others may be limited because they are hidden behind paywalls (e.g., full-text articles).

The translation quality of a data-driven machine translation tool is hugely influenced by the type of data included in the training corpus. The tools tend to work better for widely used languages, domains and text types because the training corpora contain a higher volume of relevant material from which the computer can learn. In contrast, for languages, domains and text types that are less common, the training corpus will be less robust, leading to lower quality translations. This concept of high- vs low-resource situations is important to keep in mind when assessing the potential of machine translation tools for scholarly communication. In addition, it is equally important 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 with a view to publication. Using a given translation tool may be more or less successful depending on the purpose of the translation.

2. Objectives

The overarching research question for this project is "How are translation technologies being used for multilingual scholarly communication in Canada and beyond?". Within this broad frame, we investigate the issue primarily from the perspective of tool *users* rather than developers. While the use and potential of these tools within Canada's scholarly community is of principal interest, we also examine the literature from other regions to uncover best practices, promising strategies, or known pitfalls from which Canadians can learn. Our review considers three main aspects:

1) To what extent are translation technologies being used for scholarly communication?
2) For what purposes within scholarly communication are these tools being used?
3) Which specific tool types, languages and text types are in play, and which are not?

To the extent possible based on the information found in the articles, we also make observations on the following aspects:

a) strengths and limitations of the tools;
b) reception of the tools (e.g., user satisfaction); and
c) promising strategies for integrating tools (e.g., ways of working with tools to optimize output).

3. Methods

We employ the systematic review approach to synthesize and provide a thorough and critical overview of previously published material on our topic of interest. This enables the identification of patterns and trends, as well as gaps. A systematic review of the literature can help to produce a reliable knowledge base by accumulating findings from a range of studies in a systematic and reproducible way (Briner & Denyer, 2012). A systematic review differs from other types of reviews (e.g., traditional narrative review) because it provides clear, easy-to-follow, and replicable process for synthesizing previous studies, using pre-determined criteria (Moher et al., 2009).

The review was conducted in English, French, Spanish, and Polish in accordance with the languages known to team members. However, some databases include abstracts in English for articles written in other languages, meaning that it is possible to retrieve articles in other languages as part of an English-language search. In the present project, articles written in Bulgarian, Portuguese, Russian, and Turkish were retrieved via English-language abstracts, and they were translated into English using neural machine translation to be evaluated for relevance and inclusion in the synthesis. For instance, the article by Dobrynina (2020) was translated from Russian to English, while the article by Kostadinova (2019) was translated from Bulgarian to English using DeepL Translator. This helped to broaden the study to include research conducted in languages other than English. More discussion about the inclusion and exclusion process and strategies is provided in the upcoming section on selection process (see section 3.2).

3.1 Search Process

Searches were conducted in nine academic databases to ensure wide coverage and reduce the risk that relevant articles would be missed. The databases are: 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. The search strategy included the identification of search terms that have been used in the literature to represent the two main subjects of the systematic research: translation technologies and scholarly communication.

3.1.1. Selection of search terms

This review focused on the *intersection* of two areas. These are 1) fully automatic machine translation tools, but *not* tools that support professional translators, such as electronic dictionaries, translation memories, or concordancers, and 2) scholarly communication, but *not* teaching academic writing or using tools in the university classroom. The list of terms and relevant alternative terms that have been used in the literature for these concepts was developed by one of the authors (LB) who has over 25 years of experience working in the domains of translation technologies and scholarly communication. It was verified by co-authors (PA and EK) who respectively have over 5 and over 15 years' experience working on scholarly communication and systematic reviews. Table 1 presents the search terms used for conducting the search in the selected databases.

Table 1: Search strategy

Terms relating to machine translation	**Terms relating to scholarly communication**
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 online translator* OR post-edit* OR translation engine* OR translation system* OR translation technolog* OR translation tool*	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 medical article* OR medical literature 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

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 communication. Figure 1 shows an example of a search query that has been formulated using the syntax required to query the 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 "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("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 "medical article*" OR "medical literature" 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

Figure 1: Search query sample from Scopus database

To ensure that we searched for articles that have been published in other languages, such as French, Spanish, and Polish, one of the authors (LB) developed a query prompt for ChatGPT-3.5. See Appendix A for the ChatGPT prompts, process, and outcomes of query translation.

3.2 Selection Process

In this review, we selected empirical studies that are quantitative studies (e.g., survey), qualitative studies (e.g., interview based), experimental research (e.g., comparing the use of translation tools to other translation options), or mixed methods. The authors developed inclusion and exclusion criteria to guide the selection of studies that are relevant in meeting our research focus.

3.2.1 Inclusion criteria

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

1) Focus on automatic translation tools (also known as machine translation (MT) tools), which are tools that tackle the main task of translating from one natural language to another. Examples of such tools are Google Translate, DeepL, and Microsoft Translator.
2) Focus on the scholarly communication context. This could be 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 an additional language (i.e., not the author's dominant language);
   d) *Accessing* a conference presentation through another language (e.g., automatic subtitles or an automatic transcription in another language).

3) Focus on the use or application of translation tools by and for scholars rather than on the development of a tool. An exception might be if the tool in question is being specifically developed or customized for scholarly communication (rather than a more general-purpose translation tool).

3.2.2 Exclusion criteria

Studies were excluded from analysis if they focused on the following:

1) Tools that do not perform the main task of translation but only offer support to professional translators (e.g., electronic dictionaries, translation memories, or concordancers).
2) Machine translation tools being used by professional translators or translator trainees.
3) Machine translation tools being used for language teaching or in language classrooms.
4) Machine translation tools applied in any context that is not directly related to scholarly communication (e.g., clinicians using translation tools to communicate with patients).
5) Development of machine translation tools for any application other than scholarly communication.
6) Any studies where the abbreviation "MT" refers to a concept other than machine translation.
7) Any studies where the terms "abstract" or "scientific abstract" appeared only as section headings.

3.3 Search and screening outcomes

To ensure broad coverage across multiple disciplines, we searched nine academic databases as described previously in the section on search process (see section 3.1). The Scopus database was searched on June 25, 2023, by one of the authors (EK) using the pre-determined search queries. The search returned 215 records which were exported to Excel spreadsheet for further analysis. After screening the records by title and abstracts (LB), 57 records were saved to the reference management tool Zotero for full text screening. The Web of Science (WoS) core collections database was also searched on June 28, 2023, by one of the authors (EK). The search in WoS returned 121 records which were exported to Excel spreadsheet for further analysis. After initial screening of the title and abstract (LB), 20 articles were found to be potentially relevant to our research focus, hence they 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. Next, the PubMed, MLA and ERIC databases were searched on July 31, 2023, by one of the authors (PA). After searching the databases using the search query, PubMed returned 13 articles, MLA returned five articles, while ERIC returned two articles, and these were all exported to Excel spreadsheet for further screening. After initial title and abstract screening of the records found, six records from PubMed), one record from MLA, and four 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). The search returned 500 records which were exported to Excel spreadsheet for further screening. After the title and abstract screening of the records, 61 records found to be potentially relevant to our study were saved to Zotero for full-text screening. To ensure that we accounted for articles published in languages other than English, we also searched Redalyc, a Spanish-oriented database, and Érudit, a French-oriented database on August 15, 2023. Using the translation query that one of the authors (LB) developed by prompting ChatGPT-3.5, the Redalyc database 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 database using the French search query returned no results, and the search in Scopus and WoS using the Polish search query (developed by EK) also returned no results. To further ensure that we accounted for contents published in open sources that may not be captured in proprietary databases, we searched Google Scholar on September 8, 2023. After screening the title and abstracts of the first 10 pages of the search results, 15 articles were found to be potentially relevant to our study focus. Hence, they were saved to Zotero for full-text screening.

3.4 Full-text screening outcome

Full-text screening was carried out independently by two authors (LB and PA) and the reasons for inclusion or exclusion were provided, which then allowed the authors to compare, discuss and resolve any differences. At end of the full-text assessment of 160 potentially relevant articles, 120 articles were excluded. The reason for excluding 91 of the articles is that they focused on the use of machine translation for purposes beyond scholarly communication (e.g., journalism, automatic text summarization, developing a bilingual dictionary, language learning, citizen science, translator training, coursework by student in classroom settings). Moreover, five articles focused on different settings (e.g., machine translation use by trauma surgeons or clinicians), while six articles focused on barriers faced by non-Anglophones in scholarly communication without a mention of translation tools. In addition, six articles were review articles and editorials with no empirical data, ten articles focused on different translation-related topics but not specifically on translation tools, and two articles were retracted by the publishers. Consequently, 40 articles that met our inclusion criteria were included for qualitative analysis.

The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) protocol was used to present the search process and outcomes (Moher et al., 2009). The PRISMA protocol was created by an international network of health-based collaborators to provide the framework for systematic literature review to ensure methodological rigour and quality (Pati & Lorusso, 2018). As such, we found it appropriate for reporting the process of identifying records, screening the records found, reporting eligible articles, and finally, reporting included articles. Figure 2 presents the search process and inclusion/exclusion procedures, including the reasons for excluding some full-text articles.

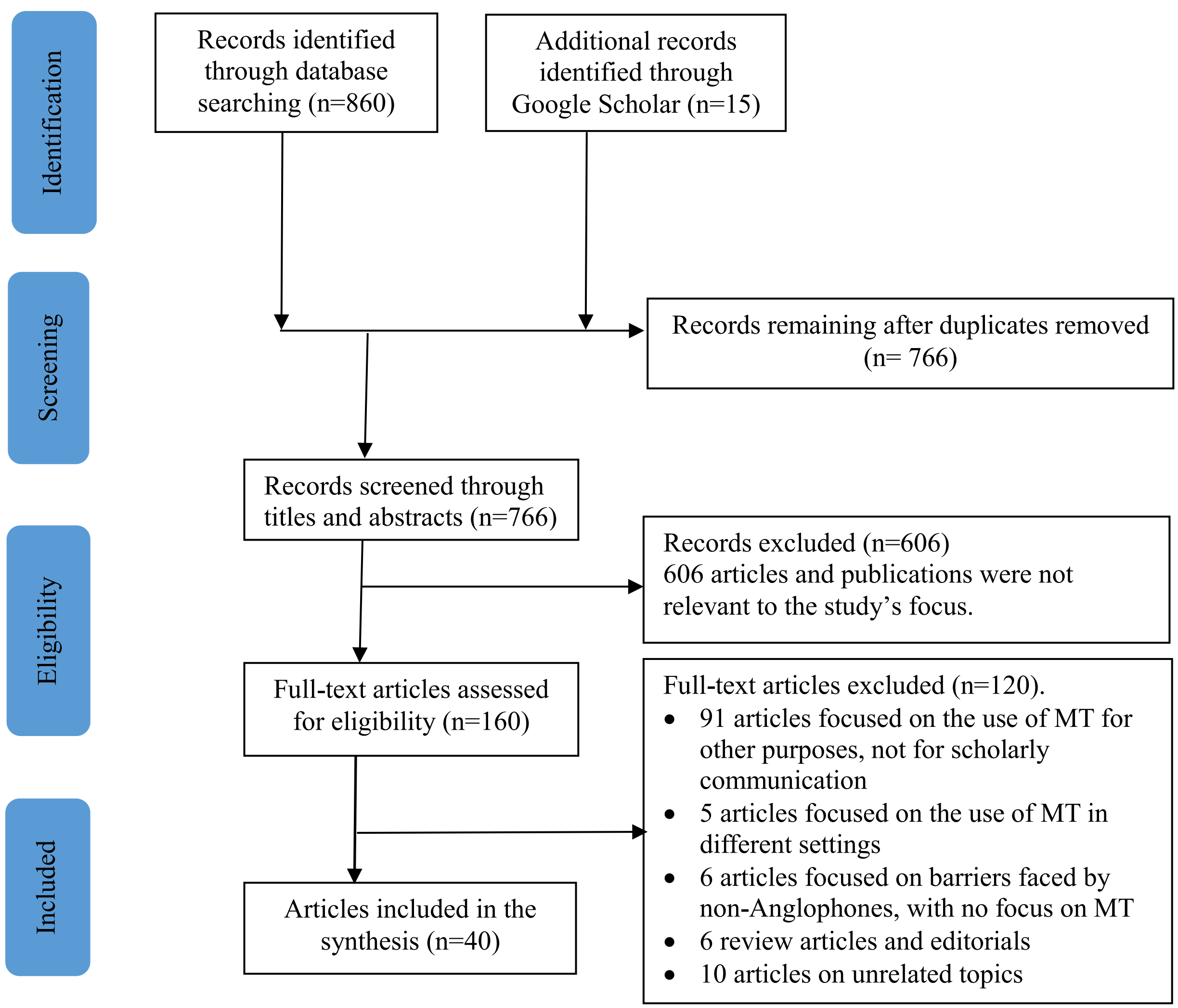

Figure 2: PRISMA protocol of search process and outcomes

3.5 Data extraction and coding approach

The coding scheme was pre-developed by one of the authors (LB) who is an expert in the field of translation technologies and scholarly communication. It was reviewed by other members of the team (EK and PA) who have relevant experience in systematic reviews and scholarly communication. The coding scheme was developed by LB based on their knowledge of the field and their experience using translation technologies in scholarly communication. Once the coding scheme had been developed, one of the authors (PA) then applied the coding scheme in NVivo and coded the 40 included 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 coding 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 of the pre-developed codes to include two additional themes: 5) evaluation methods and 6) evaluation outcomes. Hence, the coding scheme was revised to accurately represent the findings in the included studies, consisting of six main themes and associated sub-themes. For a detailed description of the themes, please see Appendix B for the codebook imported from NVivo (v.14). To check the accuracy of the initial coding done by one of the authors (PA), a second author (LB) randomly selected five articles from the included studies and coded the articles using the revised coding scheme, without seeing the coding outcomes done by PA. Afterwards, their codes were compared, and the authors discussed and resolved the few minor discrepancies in their codes.

3.6 Qualitative analysis of included studies

The analysis of the coded data was done qualitatively, synthesizing the findings from the included studies to identify existing knowledge on how translation tools have been used for scholarly communication. The qualitative analysis also helped to identify gaps in existing studies, leading to recommendations for future studies. To get an initial insight into the content of the coded data, a query of the 100 most frequent words that appeared in the 40 studies was generated in NVivo, resulting in the word cloud presented in Figure 3. As shown in the word cloud, “translators” is the most frequent word, while “English” comes in second place, followed by “language,” “sentences,” and “text”. This word cloud highlights the prominence of English in scholarly communication. Indeed, only six other languages feature in the top 100 words, but these fall into positions much further down the list:

- English: 2nd position with 212 occurrences;

- Chinese: 16th position with 50 occurrences;
- Japanese: 23rd position with 44 occurrences;
- French: 34th position with 30 occurrences;
- German: 63rd position with 20 occurrences;
- Spanish: 64th position with 20 occurrences;
- Italian: 84th position with 15 occurrences.

Moreover, the word “English” appears 212 times in the 40 included studies — over four times more often than “Chinese”, which is mentioned just 50 times in total. Indeed, with 212 occurrences, English is mentioned more times than all of these other languages combined (50 + 44 + 30 + 20 + 20 + 15 = 179). This word cloud therefore highlights not only the central position held by English, but also the fact that high-resource languages are receiving more attention than low-resource languages.

Figure 3: Word cloud of the 100 most frequent words in the included studies

Table 2: Characteristics of included studies

First Author	Year	Country	Publication	Methods	Language	Domain	Text type	Translation tools
Bawden, R.	2020	UK	Conference paper	Mixed-methods	English (source), target languages: Basque, Chinese, French, German, Italian, Portuguese, Russian, Spanish	Science & Technology	Scientific abstracts from Medline database, and Basque medical journal	Custom-built prototypes
Bowker, L.	2020	Canada	Journal article	Workshop	Chinese, English	Multidisciplinary	Not reported	Free online MT systems
Bowker, L.	2019	Canada	Conference paper	Workshop	Any (not specified)	Multidisciplinary	Scientific articles	Any free online MT systems
Bowker, L.	2018	Canada	Conference paper	Experiment	French, Spanish, English	Humanities & Social Sciences	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	Multidisciplinary	Scientific abstracts and theses/dissertatio ns	Google Translate
Daniele, F.	2019	Italy	Journal article	Experiment	English, Italian	Science & Technology	111 scientific abstracts from PubMed	Google Translate
Dobrynina, O. L.	2021	Russia	Journal article	Mixed-methods	Russian, English	Science & Technology	Scientific texts	Custom-built prototype - AiGobex

Esmailpour, R.	2020	Iran	Journal article	Experiment	English, Farsi	Multidisciplinary	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)	Multidisciplinary	Scientific thesis/research paper	Not specified (just "machine translation")
Kostadinova, D.	2019	Bulgaria	Journal article	Experiment	Bulgarian, English	Multidisciplinary	Scientific texts	Microsoft Translator, Google Translate
Lin, L. H. F.	2021	China	Journal article	Survey of 110 participants	Chinese, English	Science & Technology	Scientific papers in Engineering	Google Translate
Matsumura, Y.	2018	Japan	Conference paper	Experiment	English, Japanese	Science & Technology	Scientific texts	Custom-built NMT prototype tool
Mino, H.	2021	Japan	Workshop report	Experiment	English, Japanese	Science & Technology	Scientific abstract	Custom-built NMT prototype tool
Morishita, M.	2019	Japan	Workshop report	Experiment	English, Japanese	Science & Technology	Scientific papers	Custom-built prototype
Nayak, P.	2019	Ireland	Conference paper	Experiment	English to Basque	Science & Technology	Biomedical texts	Custom-built prototype
Neves, M.	2018	Germany	Conference paper	Experiment	English (source), target languages: Chinese, French,	Science & Technology	Scientific abstracts from Medline and EDP database	Moses SMT, OpenNMT, and various custom-built prototypes

					German, Portuguese, Romanian, Spanish			
O’Brien, S.	2018	Ireland	Book chapter	Experiment	Source languages: Arabic, Chinese, French, German, Romanian, Spanish; target language: English	Multidisciplinary	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	Multidisciplinary	Scientific abstracts from theses and dissertations in 86 European repositories	Data-driven MT
Sel, İ.	2022	Turkey	Journal article	Experiment	Turkish, English	Multidisciplinary	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	Multidisciplinary	Scientific articles from SciELO	Moses SMT system
Soares, F.	2021	UK	Journal article	Experiment	English, Japanese	Science & Technology	Scientific articles on COVID-19	Google Translate, Microsoft Bing Translator
Soares, F.	2018	Brazil	Conference paper	Experiment	English and Portuguese	Multidisciplinary	Scientific abstracts from theses and	Moses SMT system, OpenNMT

							dissertations in CAPES TDC (Thesis and Dissertation Catalogue)	system, Google Translate
Sun, Y.	2022	China	Journal article	Experiment	English and Chinese	Science & Technology	Scientific abstracts and academic articles	Custom-built prototype MT tool
Sun, Y.-C.	2022	Taiwan	Journal article	Experiment	Chinese, English	Humanities & Social Sciences	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	Multidisciplinary	Academic abstracts	Google Translate
Takakusagi, Y	2021	Japan	Journal article	Experiment	English and Japanese	Science & Technology	Scientific text	DeepL Translator
Takeshita, S.	2022	Germany	Conference paper	Experiment	English (source); target: German, Italian, Chinese, Japanese	Science & Technology	Scientific articles	DeepL Translator, LLM (BART)
Tehseen, I.	2018	Pakistan	Book chapter	Experiment	English, Urdu	Science & Technology	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	Humanities & Social Sciences	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	Science & Technology	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	Humanities & Social Sciences	MultiUN Corpus (United Nations documents)	Google Translate SMT version
Winiharti, M.	2021	Indonesia	Journal article	Experiment	Indonesian to English	Multidisciplinary	3 Scientific articles in Japanese, management, math	Google Translate
Xie, Q.	2020	South Korea	Journal article	Experiment	English, Chinese	Humanities & Social Sciences	Scientific articles	Custom-built prototype MT tool
Xu, J.	2021	France	Conference paper	Experiment	English and French	Science & Technology	Scientific abstracts in the domain of biomedicine.	Custom-built prototype MT tool
Yamamoto, S.	2021	Japan	Journal article	Experiment	English to Japanese	Science & Technology	Summaries of Scientific articles	Google Translate
Zhang, B.	2023	Unknown	Conference paper	Experiment	English, Spanish, Portuguese, French, Korean, Malayam, German, Japanese, Dutch, Turkish, Kannada	Multidisciplinary	Historical search queries in the field of e-Commerce	Custom built prototype
Zhivotova, A. A.	2020	Russia	Conference paper	Experiment	Russian, English	Science & Technology	Scientific abstracts on articles about unmanned aerial vehicles	Google Translate, DeepL Translator, Amazon Translate

Zomer, G.	2021	UK	Conference paper	Experiment	English, Spanish, Portuguese	Humanities & Social Sciences	Scholarly publications	Custom-built language checker
Zou, C.	2023	China	Journal article	Case study	Chinese to English	Science & Technology	Academic articles on engineering.	Google Translate, also mentions Baidu Translate
Zulfiqar, S.	2018	Egypt	Journal article	Experiment	German, English	Science & Technology	Scientific texts on various subfields of chemistry from German-language academic databases.	Google Translate (NMT version) and DeepL Translator

4. Results

This section presents the findings from the analysis of the 40 retained studies. It first presents some general characteristics of these studies as summarized in Table 2, with a focus on country of research (based on the first author's affiliation), domain, publication venue, and methods of data collection. Next, it discusses the following 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.

4.1 Study characteristics

As presented in Figure 4, five studies (12.5%) were conducted by first authors based in Japan, three studies (7.5%) each in Canada, China, Taiwan, UK, and USA, followed by two studies (5%) each in Brazil, Germany, Ireland, and Russia. One study was conducted in each of the other countries, including France, and Italy, among others. 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. It is of interest to note that while English-speaking countries such as the UK and USA are represented, there is a greater volume of research on this topic taking place in countries where English is not the (only) national language. In other words, it appears as though researchers from nations that are not (exclusively) Anglophone are motivated to find ways of facilitating multilingualism in scholarly communication.

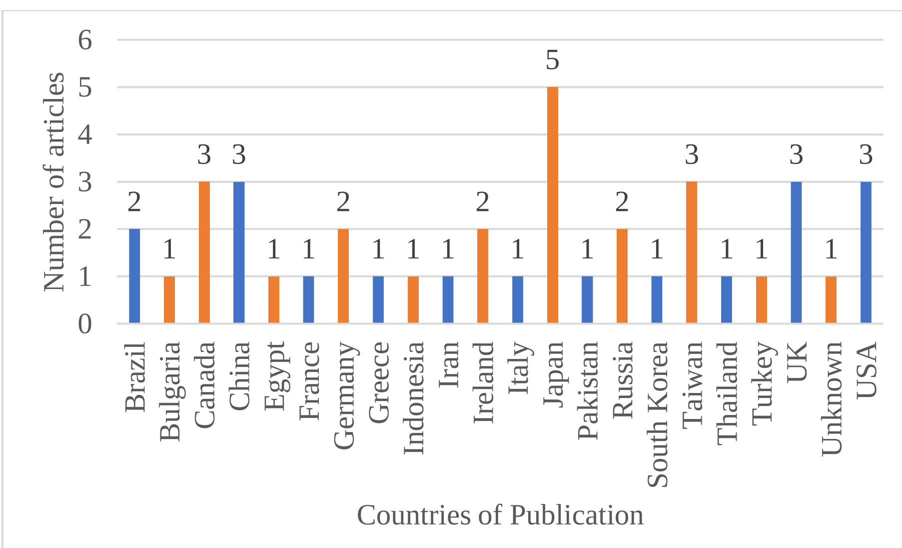

Figure 4: Countries of publication of the included studies

Half of the studies focused on translating texts in Science and Technology, while only six (15%) investigated the translation of texts in the Humanities and Social Sciences. The remaining 14 (35%) covered texts from both areas. This suggests that there may be a greater interest (or need) on the part of researchers in scientific or technical fields for employing translation tools for scholarly communication.

The majority of the studies were published as journal articles (n=21, 52.5%), followed by conference papers (n=15, 37.5 %). Other studies were published as book chapters and workshop reports. In terms of methods used in the studies, 32 studies (68%) used experiments for testing and evaluating the quality of machine translation output for scholarly communication. Four studies (10%) used mixed-methods approach by combining surveys, interviews, and analysis of written texts produced with translation tools. A few studies used pilot studies and workshops to produce or evaluate scholarly publications using translation tools.

4.2 Purpose of the study

Of the 40 included studies, 16 (40%) focused on evaluating the translation quality of the output produced by machine translation tools such as Google Translate and DeepL, by checking the performance of the tools and quality of translated texts (e.g., Bawden et al., 2020; Bowker, 2019; Daniele, 2019; Kostadinova, 2019). Thirteen studies (32.5%) focused on developing translation tools to aid scholarly communication practices of researchers (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 a tool based on SciPar, a multilingual parallel corpus of thesis and dissertation abstracts with 9.17 million sentence pairs in 31 language pairs. The tool was developed to facilitate the translation of scientific text to and from other languages to improve equitable access to scientific knowledge and accelerate research. Ten (25%) of the studies focused on using machine translation tools to produce scholarly content, exploring how scholars and researchers apply these tools for scholarly writing (e.g., Bowker, 2019, 2020; O'Brien et al., 2018; Sun & Yang, 2023; Sun et al., 2022). In contrast, just three (7.5%) studies explore how translation tools can be used for discovering scholarly content, such as through database searching (Bowker, 2018; Dobrynina, 2021; Esmailpour et al., 2020).

4.3 Tools

Findings showed that, overwhelmingly, the main type of translation tool in use in the scholarly communication community is machine translation, and more specifically neural machine translation (NMT). Only three studies involved statistical machine translation (e.g., Neves et al., 2018; Soares et al., 2018), which is an architecture that has been superseded by NMT since the introduction of the latter in late 2016. At the other end of the spectrum, only one study (Takeshita et al., 2022) incorporated research involving Large Language Models (LLMs), which began to gain traction in late 2022. Since the period covered by this review is January 2017 to September 2023, it is not surprising that NMT, which was the dominant machine translation architecture during this period, is the focus of the majority of the studies. Given the lag in publication, some studies conducted before the introduction of NMT may have been published later. Meanwhile, studies on LLMs conducted in late 2022 or early 2023 had not yet been published by the time that data collection for this systematic review ended in early September 2023.

The most frequently used NMT tool is Google Translate, with 18 (45%) of the included studies using it for translating scholarly texts or evaluating its translation quality (e.g., Chang et al., 2020; Daniele, 2019; Esmailpour et al., 2020; Kim & LaBianca, 2018; Lin & Morrison, 2021; O'Brien et al, 2018; Soares et al., 2018; Sun, 2022; Sun & Yang, 2023; Sun et al., 2022). The high use of Google Translate in the included studies could be because this tool is easily accessible and free of charge and because it can translate in more than 130 languages (Winiharti, Syihabuddin, & Sudana, 2021). The second most frequently used machine translation tool in the studies is DeepL Translator (e.g., Takakusagi et al., 2021; Takeshita et al., 2022; Wahab et al., 2020), which operates in 31 languages and accounts for four (10%) of the included studies. Only a few studies investigated other tools, such as Microsoft Translator, Baidu or Amazon translate (e.g., Kostadinova 2019). Meanwhile nine (22.5%) studies focused on comparing two or more translation tools in order to identify the one best suited to their scholarly communication needs (e.g., Bowker, 2018; Neves et al., 2018; Soares et al., 2021). While the majority of studies focus on using existing tools (e.g., Google Translate or DeepL), 13 (32.5%) research teams set out to develop custom-built prototypes, which suggests that tools made for general purposes may not sufficiently meet the specialized needs of multilingual scholarly communication (e.g., Sun, 2022; Xie et al., 2020; Xu et al., 2021).

While machine translation is the technology that is currently of the most interest to the scholarly community, a few other types of tools are investigated. For instance, Zhang & Misra (2023) examine 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 that can help to improve source-text quality, which in turn can improve the quality of machine translation output. However, tools beyond machine translation are largely overlooked by the scholarly community at present. For instance, terminology extraction tools, which can identify and extract terms from texts in specialized domains, are not featured in any of the studies included in this review. In addition, a particularly striking absence is the lack of investigation into tools for speech processing or subtitling. We did not specifically build these terms (e.g., “speech processing”, “subtitling”) into our search query because these tools can be used in monolingual environments and the focus of this review is on tools for supporting multilingual scholarly communication. However, these tools can be combined with machine translation to produce multilingual subtitles or multilingual transcriptions of spoken texts, such as conference presentations or interviews. Therefore, if these tools are being combined with machine translation tools, then studies on this topic should have been retrieved by our search query. Since no studies were retrieved on topics combining machine translation with subtitling or speech processing tools, this appears to be an area where there is significant room for research and development in the future since scholarly communication encompasses more than written or published texts (e.g., conferences).

4.4 Languages

It was noted in the background section (see section 1.1) that English has emerged as the key language for scholarly communication in recent decades. In addition, it was noted that the underlying approach employed by many translation tools is a data-driven approach (see section 1.2). For languages that are widely used and between which there is lots of translation activity (i.e., high-resource languages and language pairs), it is relatively easy to collect training data and the resulting tools produce relatively good quality translations. However, for languages that are less widely used or between which there is relatively little translation activity (i.e., low-resource languages), it is challenging to gather sufficient training data and the resulting translation quality is likely to be lower.

One noteworthy observation is that all 40 of the studies included in this review feature English as one of the languages (see Appendix C for details). As illustrated in Figure 5, a closer look reveals that in the majority of cases (25/40 or 62.5%), English is the target language, while in only nine (22.5%) studies, English is the source language. Three additional studies are bidirectional, meaning that translation takes place both into and out of English. The remaining three studies use multiple language pairs, including some that involve English. The high proportion of studies that use English as the *target* language is striking because it suggests that translation tools are not necessarily helping to displace English as the key language of scholarly communication. Instead, it appears that non-Anglophone scholars are using machine translation tools to reduce the burden of preparing English-language publications, but this is not necessarily creating a genuinely multilingual scholarly communication ecosystem.

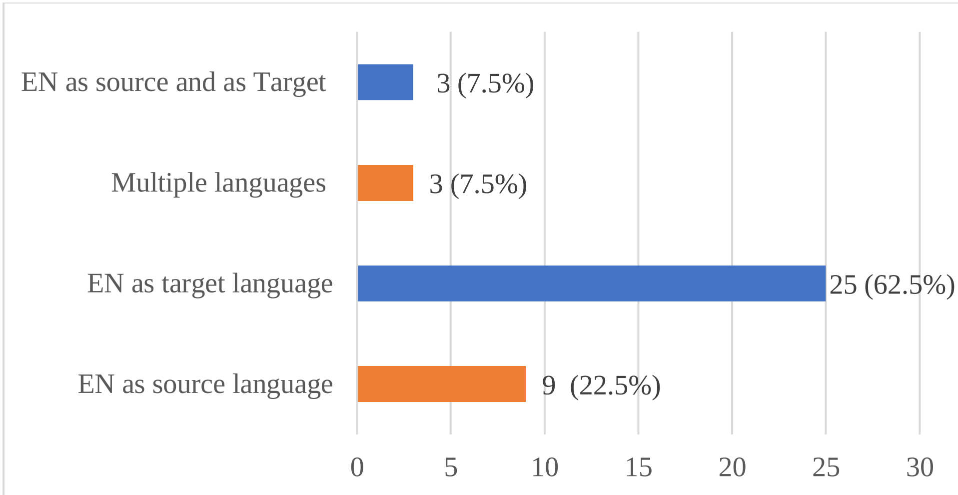

Figure 5: Distribution of studies involving English by translation direction

With regard to other languages, 25 (62.5%) of the included studies focused on high-resource language pairs, with the most used languages being English-Japanese or Japanese-English (e.g., Matsumura et al., 2018; Mino et al., 2021; Morishota et al., 2019; Soares et al., 2021; Takakusagi et al., 2021; Takeshita et al., 2022), and English-Chinese or Chinese-English (Bawden, et al., 2020; Chang et al., 2020; Neves et al., 2018; 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 ten (25%) 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 machine translation 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 translated into English more successfully with the help of machine translation tools (e.g., Esmailpour et al, 2020; Kostadinova 2019).

4.5 Text types

Scholarly communication covers a variety of text types. As explained in the background section (see section 1.2), data-driven technologies (e.g., Google Translate, ChatGPT) require a large volume of data, but this also has be to be the *right kind of data*. For instance, the linguistic and textual 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 that you want to translate must be well represented in the training data that is used to train the translation tool.

In the studies included in this review, 13 (32.5%) explored how machine translation could be used to help produce written texts such as articles or papers (e.g., Esmailpour et al., 2020; Kostadinova, 2019; Lin & Morrison, 2021: O'Brien et al., 2018; Sun, 2022; Sun et al., 2022). However, a number of the studies involved using tools that had been trained using only abstracts (9 or 22.5%) (e.g., Bawden et al., 2020; Daniele, 2019; Nayak et al., 2020: Soares et al., 2018; Sun & Yang, 2023; Xu et al., 2021) or theses/dissertations (4 or 10%) (e.g., Chang et al., 2020: Sel & Hanbay, 2022; Soares et al., 2018). Meanwhile, no systems were trained using books. Accessibility of texts for inclusion in a training corpus is undoubtedly a factor: books tend to have copyright restrictions and full-text articles are often behind paywalls, while abstracts and theses/dissertations are more likely to be accessible in databases or institutional repositories. 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.

All of the included studies focused on written texts, and none considered spoken or multimodal formats (e.g., subtitles). In addition, some types of written texts, such as posters, slides, plain language summaries, and popularized science communication texts, were overlooked. There is a need to investigate how translation tools process these other text types that are commonly found in scholarly communication contexts. For instance, the abbreviated style used on slides may be difficult for machine translation tools to disambiguate, whereas the plain language used in lay summaries or popularized texts could make these good candidates for machine translation. And as mentioned previously, conferences and interviews are other key contexts in scholarly communication that are not currently being well served by translation tools, but which could potentially be areas for future research.

4.6 Evaluation methods

As noted previously, one key area of interest for the scholarly community is 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 (human) evaluation, which can be more accurate, but which is also more costly and time-consuming to implement. Of the 27 included studies that dealt with an aspect of translation quality assessment, 18 (66.6%) used automatic metrics, while 15 (55.5%) employed manual evaluation. Of these, a few studies incorporated both types of evaluation method (e.g., Matsumura et al., 2018; Xie et al., 2020).

Among those studies that use automatic metrics, the most popular is the BLEU score, which is used in 13/18 (72.2%) cases (Bawden et al., 2020; Matsumura et al., 2018; Mino et al., 2021; Morishota et al., 2019; Nayak et al., 2020; Neves et al., 2018; O'Brien et al., 2018; Soares et al., 2018; Soares et al., 2019; Sun, 2022; Xu et al., 2021; Zhang & Misra, 2023; Zhivotova et al., 2020). Other studies employed some less popular 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 neural machine translation tools. There are other metrics that are specifically designed for neural machine translation (e.g., BERTscore, BLEURT), but the fact that these are not employed in the studies could point to low awareness and a need for improved machine translation literacy among some researchers, who may not necessarily have a background in translation technologies.

A manual evaluation approach is used in 15/27 studies (55.5%), where subject experts and/or human translators are engaged to check the accuracy of machine-translated texts. Some studies employed native speakers of the target language who had good knowledge of the source language (e.g., Bawden et al., 2020; Dobrynina, 2021; Esmailpour et al., 2020), evaluating the translated texts vis-à-vis international academic and publication conventions. For instance, Dobrynina (2021) conducted manual evaluation of machine translation by checking the accuracy of translation, consistency of translation, and compliance with the rules of English grammar.

4.7 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 communication system.

4.7.1 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 domains (see section 1.2). 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 background section (see section 1.2), translation can be undertaken for different purposes (e.g., discovering or reading a text in another language, or preparing a text for publication in another language), and machine translation 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; Windsor et al., 2019; Soares et al., 2021; Yamamoto et al., 2021; Zhovotova 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; Soares 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), Chinese (Sun, 2022), and Japanese (Yamamoto et al., 2021), among others.

The volatility and unpredictability of data-driven tools has 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. 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-resource and low-resource languages will widen since the volume of text produced in high-resource languages is growing more quickly than the volume available for low-resource languages.

For studies that involve comparing a custom-built prototype against an existing general-purpose translation tool (e.g., Google Translate), the results seem to suggest that tools customized for scholarly communication perform better for these tasks (e.g., Sel & Hanbay 2022; Matsumura et al., 2018).

4.7.2 Reception of the tools

There is a noticeable interest in and some positive attitudes towards translation tools in scholarly communication. For instance, Kim and LaBianca (2018) found that students and faculty are positive in their attitudes towards the ethical use of machine translation tools for academic writing, particularly at words 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 machine translation as an aid for writing 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).

4.7.3 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) 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" (p. 619). Similarly, Zhonotova et al., (2019) found that pre-editing a source text helps to improve machine translation quality. Sun & Yang (2023) report that back-translation is the most prevalent approach used by participants to maximize the benefits of the translation tools. They suggest that is important to develop students' abilities to write in a machine translation-friendly style. Meanwhile, in cases where a scholar intends to publish a text, it is important not to rely solely on the tools but to verify and, where necessary, post-edit the translated texts (e.g., Dobrynina, 2021).

To improve the accuracy of the translation by DeepL Translator, Takakusagi et al. (2021) suggest that users should "clarify the subject and predicate, avoid using compound sentences as much as possible, and describe the English technical term" (p. 3). Similarly, Winiharti and Sudana (2021) submit that Google Translate needs improvement with regard to forms and context-based meaning because it applies literal translation, and it is "not yet capable of identifying the context or topics being discussed in the text" (p. 717). Dobrynina (2021) also advises authors to become acquainted with the grammatical, lexical, and stylistic features of the original language, drawing their attention to the advantages and limitations of machine translation systems. Wahab et al., (2020) advise that there is need for preprocessing texts before translation to avoid falling prey to the garbage-in-garbage-out phenomenon. This confirms the findings of Zhivotova et al., (2020) that pre-editing source texts helps to improve machine translation quality.

Overall, the authors of the studies included in this review emphasize that there is a need for awareness, training, and machine translation 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 machine translation literacy through the provision of training or guidelines on the effective use of machine translation tools.

5. Implications

This section presents policy and practical implications emanating from the analysis of the included studies.

5.1 Policy implications

Overall, it is worth emphasizing that technology alone cannot create or sustain a multilingual scholarly communication ecosystem. It is essential for tool use to be supported by meaningful policies that address areas such as those outlined below.

1) There is a clear **appetite** for publishing in languages other than English, so funding agencies, journals and institutions must develop policies, offer support, and implement value structures to facilitate and appropriately recognize scholarly communication in multiple languages or in languages other than English.
2) People ***are*** using language technologies for scholarly communication, so journals and funding agencies should not ignore this. The tools are not perfect, but simply banning them is not a helpful approach; rather, there is a need for guidelines and policies about **responsible** and **transparent** use of tools in scholarly communication.
3) At present, many non-Anglophone scholars are using translation tools to **reduce the burden** of publishing in English; however, this does not actively contribute to the creation of a multilingual scholarly communication ecosystem. Policies are needed to **shift the responsibility** away from expecting speakers of other languages to use translation tools to produce texts (in English) and towards encouraging the use of these tools to access and engage with research that has been written in other languages.

4) Currently, language technologies are being used **mainly for written texts**, but it is clear the field is evolving rapidly (as demonstrated by interest in the recently emerged LLM-based tools), so policy makers should not ignore **speech and subtitling technologies** and their potential for scholarly communication (e.g., conferences, interviews).
5) To produce quality output, data-driven tools require not only a large quantity of data but also **the right kind of data**. At present, many data-driven tools developed to translate scholarly texts rely on abstracts or theses/dissertations for training corpora since book and journal content is often behind a paywall. Policies that meaningfully support **open access** could help to ensure that a greater volume and more diverse range of academic text types are available to develop high-quality training corpora.
6) Language technologies can play an important role in ensuring more equity and diversity in scholarly communication, but there could be implications for the human translation and interpretation industry. Policy makers should pay attention to the need for **sustainable** solutions.

5.2 Practical implications

1) English and French are high-resource languages, so there is stronger potential for existing data-driven translation tools to be used successfully for Canada's official languages rather than for Indigenous languages or low-resource languages from other regions.
2) While it is encouraging to see that some researchers are investigating low-resource languages, most attention is still focused on high-resource languages. As a result, there is a risk that the gap between central or major languages and those on the (semi)peripheries will grow even wider if more attention is paid to perfecting translation tools for languages that are already well served, rather than developing tools for languages that are underserved.
3) Translation tools created for general purposes may not be sufficient for translating specialized scholarly content well.
4) Automatic evaluation metrics may not be sufficient for reliably determining the quality of machine-translated texts intended for scholarly publication.
5) Translation tools are not perfect or foolproof, so it is important for tool users to improve their machine translation literacy, such as by understanding the potential for data bias, recognizing the limitations of automatic evaluation metrics, having reasonable expectations with regard to 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.
6) Emphasis is currently being placed on using machine translation to produce scholarly texts (in English), rather than on using translation tools to discover and access content in other languages. As a result, the responsibility for multilingualism in scholarly communication continues to rest on the

shoulders of non-Anglophone scholars while English and English-speaking scholars remain in a privileged position.

7) Scholars who need support delivering or understanding oral presentations or participating in interviews or discussions in another language are not currently being helped by translation tools, even though these tools may have the potential to assist.

6. Conclusions

The theme for this iteration of the SSHRC Knowledge Synthesis Grants is *shifting dynamics of privilege and marginalization*. Accordingly, the overarching goal of this systematic review was to gain a better understanding of how translation tools are being used in the context of scholarly communication, and more specifically to learn whether these tools are helping to shift this largely monolingual English space towards a more linguistically diverse ecosystem.

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, the overwhelming majority of items retrieved were published in English, which confirms the central role that this language plays in scholarly communication. 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 translation technologies and scholarly communication. Forty 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.

One key observation is that the field is evolving rapidly, as evidenced by the fact that only a few studies focused on statistical machine translation – a paradigm that dominated the field for nearly 20 years until neural machine translation was introduced in 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 policy makers, who must respond quickly to support researchers.

Our review reveals that translation technologies are indeed being used in the scholarly community; however, the focus is firmly on prototypical scholarly texts (i.e., scientific abstracts and journal articles) with no attention paid to how these tools might support oral forms of scholarly communication (e.g., conferences) or less prototypical texts (e.g., slides, posters, popularized texts). Moreover, the main tools of interest are general-purpose data-driven machine translation tools (e.g. Google Translate), which are used most often and most successfully to translate between high-resource languages (i.e., languages that are widely spoken), although tools that have been customized for scholarly communication appear to perform even better. Less research focuses on low-resources languages, which are already less widely spoken. There is therefore a risk that the gap between central and (semi)peripheral languages will be exacerbated since the tools to support the former are already good and are getting better, while tools to support the latter are developing at a slower pace and are producing lower quality results. One concerning situation that has been revealed by our review is that English – which has long occupied an outsized place in the scholarly communication sphere – continues to take centre stage even in the presence of translation technologies. All 40 of the studies included in our review had English as one of the languages under investigation, and the majority focused on how translation tools can be used to help non-Anglophone scholars write in English. Translation technologies have the potential to help level the playing field by allowing users to discover and access works written in other languages; however, at present, the primary use of these tools is to facilitate the publication of scholarship in English. In other words, translation tools are not yet helping to displace English as the key language for scholarly communication, and the responsibility for translating remains firmly on the shoulders of non-Anglophone scholars.

A greater effort is therefore needed to shift the dynamics of linguistic privilege and marginalization in the scholarly communication ecosystem. Our study demonstrates that the solution required is not solely a technological one. Current translation technologies, 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. Why then are these tools mainly being used to translate out of other languages and 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 communication primarily through the medium of English. Our review illustrates that while translation tools can help, technology alone is not enough to achieve or sustain a multilingual scholarly communication ecosystem.

6.1 Future areas of research

More research is needed in the following areas:

- developing translation tools for low-resource languages;

- identifying optimal customizations for translation tools to be used in scholarly communication contexts;
- identifying the best use cases for translation tools in the context of scholarly communication;
- understanding the potential and limitations of translation tools for handling additional text types that are relevant to scholarly communication, such as slides, posters, plain language summaries, popularized texts, oral speech and multimodal texts (e.g., subtitles);
- investigating the potential and limitations of large language models (LLMs) for translation in scholarly communication;
- identifying and developing appropriate automatic and other types of evaluation metrics for neural machine translation tools and translation tools based on LLMs;
- investigating the potential of LLMs for other language-related tasks, such as text simplification or text summarization, which can in turn support multilingual scholarly communication (e.g., by making source texts more translation-friendly, or by facilitating the publication of multilingual abstracts);
- determining how policies related to scholarly communication can work in conjunction with translation technologies to support multilingualism and linguistic diversity.

7. Knowledge Mobilization Activities

The primary target audience for this knowledge synthesis consists of members of the research community who want to work towards creating a more multilingual scholarly communication ecosystem. This includes researchers themselves, but also journal editors and academic publishers, as well as policy makers and funding agencies. Finally, members of the wider public may also be interested in certain aspects of this knowledge synthesis, such as how multilingual research could benefit society more broadly. We have planned a variety of knowledge mobilization activities to reach these different groups.

7.1 Researchers

For researchers, we will publish the final project in open access as a review paper for a peer-reviewed scholarly journal such as *Annual Review of Information Science & Technology* or a similar journal. In addition, we will present our findings at two different online seminars that address international research audiences:

- Presentation for the UNESCO Chair on Open Science research group, coordinated by the Chair holder Prof. Vincent Larivière, Université de Montréal, Canada (scheduled for 8 February 2024).

- Presentation for the Scholarly Communication Research Group seminar series on Science and (Semi) Peripheries, coordinated by Dr. Krystian Szadkowski, Adam Mickiewicz University (scheduled for February 2024)

7.2 Editors, publishers and policy makers

To reach an audience that includes not only researchers but also academic publishers, journal editors and policy makers, we will present our findings at three events, including one in French:

- SSHRC Virtual Forum on Shifting Dynamics of Privilege and Marginalization (25 January 2024).
- Public seminar on Multilingualism in Research organized by the Helsinki Initiative on Multilingualism in Scholarly Communication (scheduled for March or April 2024). This seminar will be coordinated by Janne Pölönen, Secretary General of the Publication Forum, Federation of Finnish Learned Societies. Our project will be presented alongside projects from the Linguistic Diversity group of the DIAMAS (Developing Institutional open Access publishing Models to Advance Scholarly communication) Project and the Multilingual SIG of OPERAS (Open Access in the European Research Area through Scholarly Communication), among others.
- *Pour la science en français : la traduction automatique, une avenue prometteuse* – une journée de réflexion organisée par le Commissaire à la langue française au Québec (26 mars 2024).

7.3 Wider society

Finally, although our research focuses more on the academic community, we plan to engage members of Canadian society more broadly in the final months of the project by submitting popularized pieces to venues such as the following:

- Our languages/Nos langues: this bilingual (EN/FR) blog, which is maintained by the Government of Canada's Translation Bureau, invites contributions from the public on any language-related topic and encourages readers to add comments and engage with the authors. One of the authors (LB) has already contributed four posts on various aspects of translation technologies to the blog over the past several years.
- *The Conversation Canada*: This is an independent source of news and views from the academic community that is delivered directly to the public in the form of journalistic or popularized research.

Bibliography

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

Balula, A., Caliman, L., Costa, C., Fiorini, S., Jarmelo, S., Leão, D., Mounier, P., Nomine, J.F., Padez, M.J., Plag, C., Pölönen, J., del Rio Riande, G., Silva, M., Stojanovski, J., Szulińska, A., Volt, I., Toledo, E.G., and Withanage, D. (2021). Innovative models of bibliodiversity in scholarly publications. *SIG Multilingualism White Paper*. https://doi.org/10.5281/zenodo.4817508

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. (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. (2002). *Computer-aided translation technology*. Ottawa: University of Ottawa Press.

Briner, R. B., & Denyer, D. (2012). *Systematic review and evidence synthesis as a practice and scholarship tool*. In D. Rousseau (Ed.), *The Oxford Handbook of Evidence-Based Management: Companies, Classrooms, and Research*, 112-129. Oxford University Press.

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/

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

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

Dobrynina, O. L. (2021). Academic Writing for Publication Purposes and Machine Translation: Is the Symbiosis Possible? *Www. Vovr. Elpub. Ru; Www. Vovr. Ru Журнал Издаётся с 1992 Года*, *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

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

Habibie, P., & Hultgren, A. K. (eds). (2022). *The Inner World of Gatekeeping in Scholarly Communication*. Cham: Palgrave Macmillan.

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

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

Kostadinova, D. (2019). On some particular cases of translation and self-translation. *Езиков свят - 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

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* 6(7), e1000100. https://doi.org/10.1136/bmj.b2535

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

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_11

Pati, D., & Lorusso, L. N. (2018). How to Write a Systematic Review of the Literature. *HERD: Health Environments Research & Design Journal* 11(1):15-30. http://doi.org/10.1177/1937586717747384

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

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 Conference*, 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

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. (2022). Analysis of Chinese Machine Translation Training Based on Deep Learning Technology. *Computational Intelligence and Neuroscience*, *2022*, e6502831. https://doi.org/10.1155/2022/6502831

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

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.

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: ChatGPT prompts, process, and outcomes of query translation

Prompts for ChatGPT

1. You are an academic database query tool.
2. Reconstruct the following query for the Web of Science database but substitute the English keywords with their French equivalents.

3. (TS=("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 "online translator*" OR "post-edit*" OR "translation engine*" OR "translation system*" OR "translation technolog*" OR "translation tool*")) AND (TS=("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 "medical article*" OR "medical literature" 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.

4. Reformulate the following Web of Science query for Scopus.
5. (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 "article* médica*" OR "littérature médicale" 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 allow a user to enter a search term in one language and retrieve texts written in another language.	1	2
g. Speech processing tool (e.g., voice recognition tools,	Tool that takes spoken language as input or output and that can be combined with a machine	0	0

speech generation/synthesis tools, automatic subtitling tools)	translation tool to produce subtitles or speech in another language.		
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: Distribution of articles by language and translation direction

Language	Articles	Frequency
English as source language	Bawden et al., 2020; Daniele, 2019; Nayak et al., 2020; Roussis et al, 2022; Soares et al., 2021; Tehseen et al., 2018; Takeshita et al., 2022; Xu et al., 2021; Yamamoto et al., 2021;	9
English as target language	Bowker 2018; Bowker, 2020; Chang at al, 2020; Dobrynina, 2021; Esmailpour et al, 2020; Kostadinova, 2019; Lin & Morrison, 2021; Mino et al., 2021: Neves et al., 2018; O'Brien et al., 2018; Sel & Hanbay, 2022; Soares et al., 2018; Soares et al., 2019; Sun & Yang, 2023; Sun et al., 2022; Takakusagi et al., 2021; Tongpoon-Patanasorn & Griffith, 2020; Wahab et al., 2020; Windsor et al., 2019; Winiharti & Sudana., 2021; Xie et al., 2020; Zhivotova et al., 2020; Zomer & Frankenberg-Garcia., 2021; Zou et al., 2023; Zulfiqar et al. - 2018	25
Multiple languages (including English)	Bowker 2019; Kim & LaBianca, 2018; Zhang & Misra – 2023;	3
English as source and target language	Matsumura et al., 2018; Morishita et a., 2019; Sun, 2022;	3