Check for updates

Original Research

SAGE Open
January-March 2026: 1–18
© The Author(s) 2026
DOI: 10.1177/21582440261415716
journals.sagepub.com/home/sgo
Sage

A Bibliometric Analysis of Trends and Developments in Corpus Translation Studies (2015–2024)

Yiyang Hu[1] **and Kanglong Liu**[1]

Abstract

Corpus-based translation studies (CTS) has been shaped by theoretical developments within translation studies, while also incorporating insights from related fields such as linguistics and digital humanities. Although prior reviews have noted the field's increasing methodological diversity, this complexity has made it challenging for new researchers seeking to understand the field's trends and developments. Bibliometric analysis offers a systematic, data-driven approach to exploring scholarly activity and provides a valuable means to address this gap. However, its application in CTS has remained limited in recent years. This study addresses the gap by conducting a bibliometric analysis of 477 research articles published between 2015 and 2024, retrieved from the Web of Science (WoS) database. By examining publication trends, term frequencies, and citation patterns, it identifies shifts in key research themes and the underlying intellectual structures of CTS. The findings reveal that CTS has expanded and refined its research focus over the past decade, with a predominant emphasis on translation universals and a noticeable rise in interpreting studies. Network analyses further elucidate the intricate relationships among these themes, providing a deeper understanding of the field's evolving dynamics.

Keywords

corpus translation studies (CTS), bibliometric analysis, translation universals, interpreting studies, digital humanities

Introduction

Corpus translation studies (CTS) is widely recognized as one of the most prominent areas where digital humanities have been applied within translation studies (Gu, 2024). Regarding the relationship between corpus linguistics (CL) and digital humanities (DH), Maci and Sala (2022) argued that:

> Where DH is the territory, CL is the map to chart it and make it manageable; where DH is the field, CL is the trajectory along which to navigate it. (p. 4)

According to Yan et al. (2025), the origins of DH can be traced to corpus linguistics, with its applications subsequently extending into various other disciplines. In this regard, researchers in CTS have been at the forefront of integrating DH techniques into translation studies. Emerging from the convergence of multidisciplinary theories, CTS has defined key research areas over the past three decades. Previous top-down reviews have explored the state and trends of CTS, particularly in light of the growing influence of DH. While these reviews have offered valuable guidance for future research and have supported new scholars in the field, their scope may be constrained by the inherent subjectivity in the selection of research articles.

Bibliometrics provides a systematic and structured approach to conducting academic reviews, minimizing potential subjective biases while delivering quantitative, evidence-based insights for deeper analysis. Previous reviews (Laviosa & Liu, 2021a; Hu & Wang, 2022) have underscored the growing complexity of the CTS landscape. Recognizing that the relationships between research fields are fluid and interconnected, we contend

[1]The Hong Kong Polytechnic University, China

Corresponding Author:
Kanglong Liu, Department of Language Science and Technology, The Hong Kong Polytechnic University, HHB812, LST, Hong Kong SAR 999077, China.
Email: klliu@polyu.edu.hk

Data Availability Statement included at the end of the article

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

that a more nuanced understanding of CTS can be attained by exploring the interconnections and interactions among these fields. Network-based methods in bibliometrics are particularly effective for uncovering the intricate dynamics within CTS, as they enable the analysis of large datasets of research articles. Despite its potential, bibliometric analysis has been underutilized in CTS, with Pang (2019) being one of the few exceptions. As digital humanities have deepened and extended the research fields of CTS, it is crucial to detect the intricate relations among different research topics. Given the rapid developments in CTS over the past 5 years, this study aims to map the changes systematically by drawing on bibliometric techniques. By tracing how the academic publishing, research topics, and knowledge structures wax and wane over the last decade, the following study can guide new scholars to gain an overview of the landscape, grasp the key literature, and predict the research trend of CTS.

Evolution and Scope of CTS

Foundations and Developments of CTS

The emergence of CTS in the 1990s was driven by the convergence of theories from various disciplines. Baker (1993) highlighted that the recognition and elevation of the status of translations were influenced by Even-Zohar's (1979) polysystem theory, which treats translations as texts in their own right that are equal in status to original works. This elevated status of translations intersected with J. R. Firth's focus on linguistic usage in context, the rise of corpus linguistics (Sinclair, 1991), and Toury's (1995) concept of translation norms in descriptive translation studies. Translated texts are regarded as valuable and authentic linguistic data, comparable to original texts, for the construction of parallel, comparable, and multilingual corpora (Baker, 1995). Baker (1995) also emphasized the importance of building translation corpora for descriptive research, as well as for applications in education and terminology. CTS thus moved away from traditional paradigms that prioritized source texts as the basis for comparison, and instead positioned translated texts as independent research objects situated within specific communicative and sociocultural contexts.

Translation universals have been a central focus of research since the inception of CTS. Drawing on earlier theoretical frameworks, Baker (1993, 1995, 1996) identified several hypotheses about translation patterns, including simplification, explicitation, normalization, and leveling out. These hypotheses were later refined and categorized by Chesterman (2004), who distinguished between S-universals and T-universals. S-universals refer to linguistic features that emerge from comparing translations with their source texts, while T-universals are identified by comparing translations with comparable non-translated texts (Chesterman, 2004). While these classifications expanded the theoretical scope of translation universals, early studies were often limited in empirical scope due to constraints such as language pairs, text types and individual translator styles, making it difficult to draw broad, generalizable conclusions.

The concept of translation universals has also faced criticism from several scholars. Tymoczko (1998) warned that observed regularities in translated texts may reflect the methodological preferences or ideological biases of particular research communities rather than universal tendencies. Chesterman (2004), in his reflection on corpus-based studies, cautioned against overgeneralizing translation universals, noting the lack of consistent criteria for corpus design and selection. House (2008) went further by rejecting the very notion of translation universals, arguing that it overlooks critical variables such as translation direction, genre conventions, and historical change. Similarly, Becher (2010) challenged the classification of explicitation as a universal, proposing instead a pragmatic explanation grounded in context-specific communicative norms. In response to these critiques, scholars began to move toward more nuanced models of variation in translated language. One such development was the reconceptualization of translation universals within the broader framework of constrained language (Lanstyák & Heltai, 2012), which views translated texts as one among several types of language shaped by context-specific constraints. This framework allows for the modeling of diverse linguistic and situational factors that influence translated language, offering a more flexible and empirically grounded approach to variation across language varieties.

Research within corpus-based translation studies (CTS) has evolved significantly, moving beyond the initial focus on translation universals to encompass a broader range of interdisciplinary inquiries. The increasing integration of digital humanities into translation studies has played a key role in driving this expansion. As noted by Laviosa and Liu (2021a), CTS research now falls into two major domains: descriptive and applied studies. Descriptive studies continue to investigate translation universals, along with the analysis of register variation and translational norms. In contrast, applied studies focus on the practical use of corpora in areas such as machine translation, translation quality assessment, and translator training. For example, Wołk (2019) demonstrated how parallel corpora can enhance statistical machine translation for low-resource languages like Polish. This broader scope is further reflected in the review by Hu and Wang (2022), who examined the development of translation studies through the lens of digital

humanities. They identified CTS as the foundational stage, which has since expanded to incorporate methodologies from critical discourse analysis and multimodal analysis. They also outlined a second stage characterized by the adoption of text mining techniques to investigate topics such as translation strategies, methods, and communicative practices.

The integration of digital humanities into CTS has led to significant methodological advancements and the identification of new research gaps. Gu (2024) argued that corpus linguistic methods represent the most impactful application of digital humanities in translation and interpreting studies. Laviosa and Liu (2021b) outlined these developments, which range from research object selection and feature extraction to the use of statistical analyses. In areas such as subtitling, news, and audio description, building effective corpora often requires the application of NLP techniques. The extraction and quantification of specific features also depend on data mining and statistical methods.

These top-down reviews offer valuable guidance for new researchers by identifying emerging topics and methodological trends in CTS. However, as the volume of research continues to grow, the field has become increasingly complex, raising questions about whether these reviews adequately represent the breadth of CTS. In addition, the relationships among research topics over time (diachronic) and within a given period (synchronic) remain unclear, highlighting the need for further exploration into how research areas develop and interact. Methodological advances driven by digital humanities have broadened the scope of CTS, expanding its research objects and fostering links with adjacent disciplines, which in turn has led to the formation of new scholarly communities. To complement earlier top-down approaches, this study adopts a systematic bottom-up method based on bibliometric techniques to capture subtle shifts and evolving patterns within the field.

Bibliometrics

The term "bibliometrics" was first introduced by Pritchard (1969), who defined it as "the application of mathematics and statistical methods to books and other media of communication" (p. 349). Unlike traditional reviews that rely on selective close reading, bibliometric analysis reduces expert bias by systematically collecting all relevant articles from a database. It also enables the tracing of diachronic changes in research paradigms through sophisticated statistical methods (Chen, 2016).

Bibliometric analysis has been used to trace the development of various subfields within applied linguistics. Liao and Lei's (2017) study is one of the earliest bibliometric investigations in corpus linguistics. Similar approaches have been applied in related areas, including applied linguistics more broadly (Lei & Liu, 2019) and English for academic purposes (Hyland & Jiang, 2021). Early studies often relied on *n*-gram analysis to identify dominant topics and monitor their progression over time. The introduction of tools such as CiteSpace (Chen, 2016) has greatly improved the efficiency of statistical processing and the visualization of bibliometric data, enabling the construction of co-occurrence networks, co-citation maps, and structural variation models.

Early quantitative reviews in translation studies can be traced to bibliographic compilations from the 1980s, which concentrated primarily on interpreting studies. Pioneering work in this area includes Gile's (2000) review on conference interpreting and his later study on translator and interpreter training (Gile, 2005). These early analyses were typically limited in scale due to the manual effort required for citation analysis. However, recent advances in bibliometric and corpus-based tools have significantly reduced the need for manual coding, allowing for more comprehensive and systematic investigations. As a result, bibliometric research has expanded across multiple levels and subfields, including general translation studies (Pan & Wu, 2024), book reviews in translation journals (L. Li & Liang, 2024), journalistic translation (Ping, 2021), and interpreting studies (Martínez-Gómez, 2015). The growing adoption of bibliometric methods underscores their value in deepening our understanding of specific domains, revealing hidden connections, and supporting the training of emerging scholars (Gile, 2000).

To the best of our knowledge, Pang (2019) conducted the first bibliometric analysis of CTS, covering publications from 2008 to 2018. Using CiteSpace, the study mapped the CTS research landscape by examining basic bibliographic data, keyword trends, and co-citation patterns among citing and cited works. In the past 5 years, the growing integration of digital humanities into translation studies has contributed to a notable increase in research outputs across a wider range of subfields. This shift suggests the emergence of new connections between traditional core topics and digitally driven areas of inquiry. To investigate these recent developments, the present study builds on Pang's (2019) work by adopting a similar bibliometric approach, with a focus on the following research questions:

RQ 1: What are the main characteristics of CTS publications from the past decade (2015–2024), particularly in terms of publishing journals, contributing countries, and affiliated institutions?

RQ 2: What evolutionary trends can be observed in the most prominent research topics within CTS, and how are these topics thematically interconnected?

RQ 3: Which publications have exerted the greatest influence on CTS over the past decade, and what underlying knowledge structure can be uncovered through co-citation analysis?

Methodology

Data Collections

The dataset was retrieved from the WoS collection database, which includes the SSCI and A&HCI databases and provides a representative and comprehensive source for this study. Based on previous CTS reviews, the retrieval strategy was expanded to include "interpreting" and "translation practices" as search terms. To broaden the scope of subfields, categories such as "computer science" and "artificial intelligence" were also incorporated into the search criteria (see Figure 1). The dataset was then refined by excluding reviews, book chapters, conference papers, and journal articles in press. After these initial filters, 840 articles remained for manual screening.

The relevance of research articles to CTS was verified through close reading of abstracts and the application of specific exclusion criteria to eliminate unrelated studies. Articles were excluded if the term "corpora" was used without reference to linguistic content. In particular, studies that referred to corpora as isolated sentences or fragmented language samples, even when the term appeared in the title, were excluded. Since the focus of this bibliometric analysis was on CTS, the exclusion criteria used in the previous bibliometric study of corpus linguistics by Crosthwaite et al. (2023) were also applied. In addition, each article had to meet the three criteria defined by Hu (2015) for inclusion in CTS: the use of translation corpora, the application of statistical methods, and analyses grounded in translation theories, linguistics, literary theories, or cultural studies. After applying these criteria, 477 research articles remained for further manual screening.

Analytical Procedures

For **RQ1**, the overview of publication counts, journals, authors, and institutions was gathered by using Biblioshiny in the Bibliometrix package (Aria & Cuccurullo, 2017). Compared to other bibliometric tools, Bibliometrix offers greater flexibility, as it can be integrated with statistical and visualization software like R Studio (R Core Team, 2022). This flexibility facilitated a series of pre-analyses to reduce noise and address synonyms in the WoS collection database.

To address **RQ2**, the 2-grams and 3-grams from titles and abstracts were extracted by Bibliometrix in R Studio. All terms from the authors' keywords were also included in the extraction due to their high representativeness, regardless of whether they were 1-grams. The preliminary result of the term lists indicated that 2-grams and 3-grams from titles and abstracts represented the key concepts in the CTS subfield more effectively than 1-grams or 4-grams. 1-grams tended to be general terms like "translation," "interpreting," and "discourse," while 4-grams had a low frequency after preliminary analysis. Drawing on the stopword list function in Bibliometrix, a custom list was created to remove noise and standardize synonyms. This process involved manually reviewing key terms and compiling synonym lists (see Appendix). For instance, formulaic expressions in academic writing, such

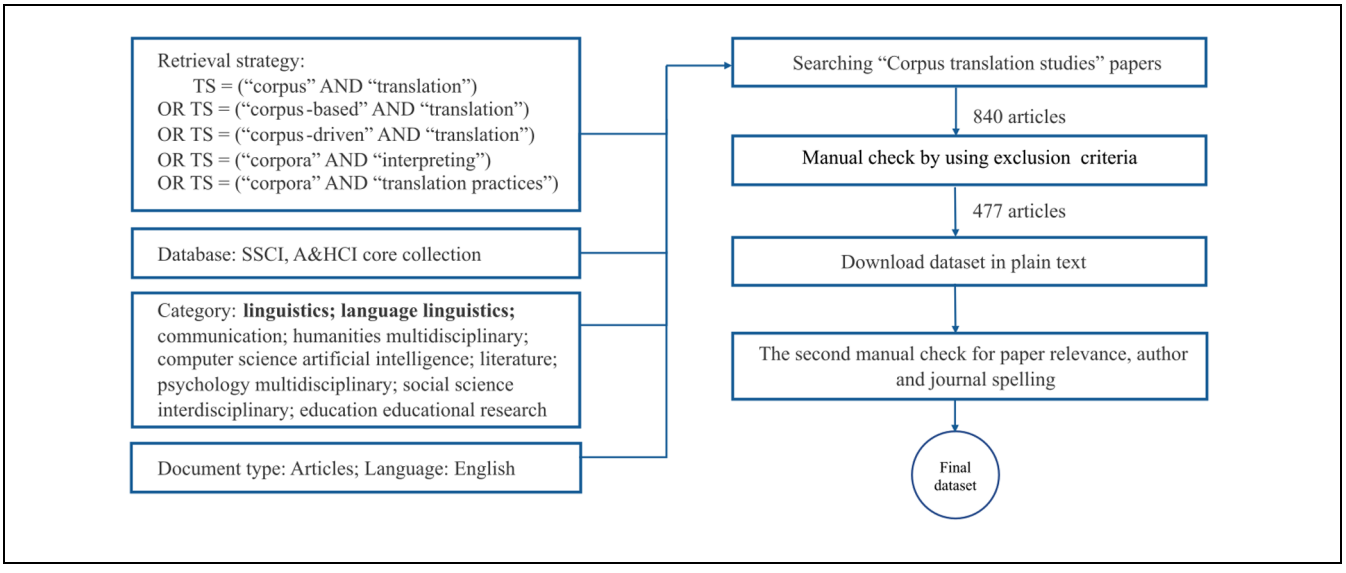

Figure 1. The procedures of data collection.

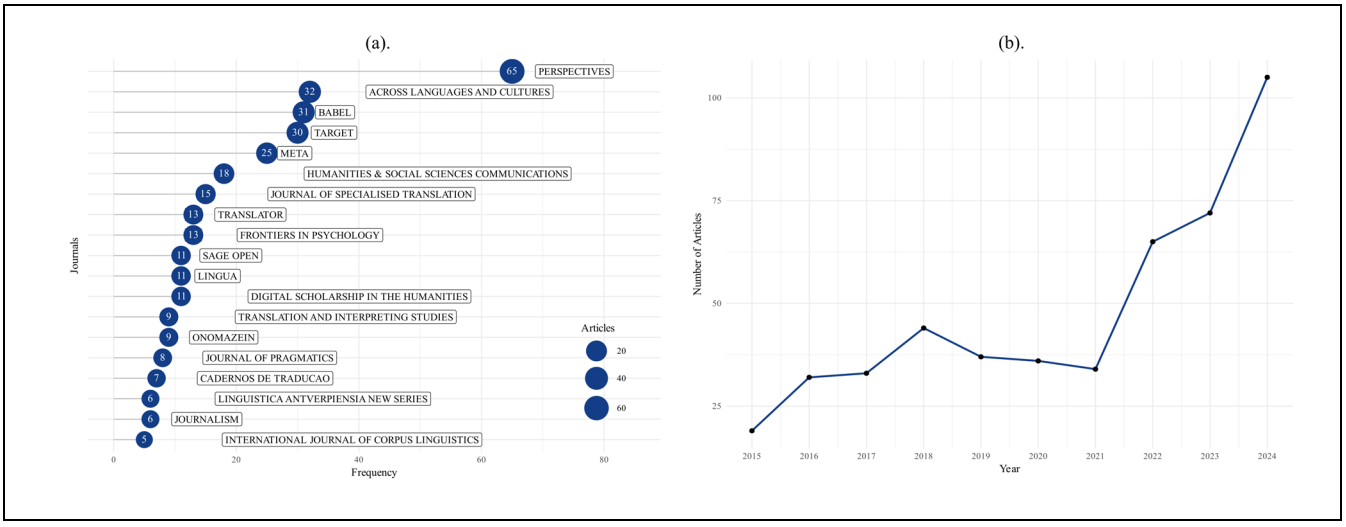

Figure 2. (a) Most relevant journals of CTS from 2015 to 2024 and (b) annual production of CTS articles from 2015 to 2024.

as *study finds* were removed; the term corpus-based translation studies and its variants were consolidated under the synonym CTS.

The overall frequency of all terms was first calculated to gain an overview of research themes over the last decade. Then Kleinberg's (2003) burst detection algorithm was applied to capture the sharp and sudden onset of core terms in adjacent years. Burst detection was performed using Sci2 (Sci2 Team, 2009) with a gamma value of 1.0 and a density scaling parameter of 2.0. To examine the interrelations of terms in CTS, the network analysis approach in Bibliometrix was used to visualize vertex sizes and clusters. Since Biblioshiny supported network analysis for only one type of *n*-gram (1/2/3-gram) in a single field (titles, abstracts, or keywords), only 2-grams from abstracts were chosen as vertices in the network.

The vertices in the pre-analysis were further filtered by referencing the noise reduction list and the synonyms list (see Appendix). One overall network for the past decade, along with two networks for each 5-year period, was constructed to examine the co-occurrence patterns of the principal research themes. After multiple trials, no more than 50 vertices were included in the network, based on method parameters including the Fruchterman-Reingold layout, the Louvain clustering algorithm, and association normalization. The threshold for vertex inclusion was a minimum of 2 edges.

For **RQ3**, regarding citation and co-citation, this study calculated the overall citation frequency for both the cited works and the articles in our dataset. Next, a co-citation network was constructed based on the cited works. It should be noted that some citation items in the WoS database were incomplete. The missing information for certain highly-cited works and key vertices was manually corrected through close reading. A record of these corrections can be found in the Appendix.

Results and Discussion

The following section is divided into three parts, each addressing one of the research questions. The first section provides an overview of the basic information on CTS articles. The second section focuses on the principal research themes by analyzing overall term frequency, burst strength of terms, and co-occurrence networks. The final section outlines the state of CTS based on citation and co-citation analyses.

General Publication Trends

Emerging Cross-Disciplinary Engagement in CTS Research. As shown in Figure 2a, the dataset reveals a predominance of CTS articles published in journals specializing in translation studies such as *Perspectives*, *Babel*, *Target*, *Meta*, and *Across Languages and Cultures*. Beyond these discipline-centric publications, the distribution of articles highlights CTS research's interdisciplinary resonance, with contributions from journals in psychology, linguistics, digital humanities, and media studies. The diverse academic engagement of scholars in CTS suggests that the research themes in CTS form interconnections across fields, such as cognition in translation behaviors, the social implications of translation products, and the influence of translation technology. Two fully Open Access academic journals, *Sage Open* and *Humanities and Social Sciences Communication*, contribute notable strength, but to a lesser degree than journals specializing in translation studies.

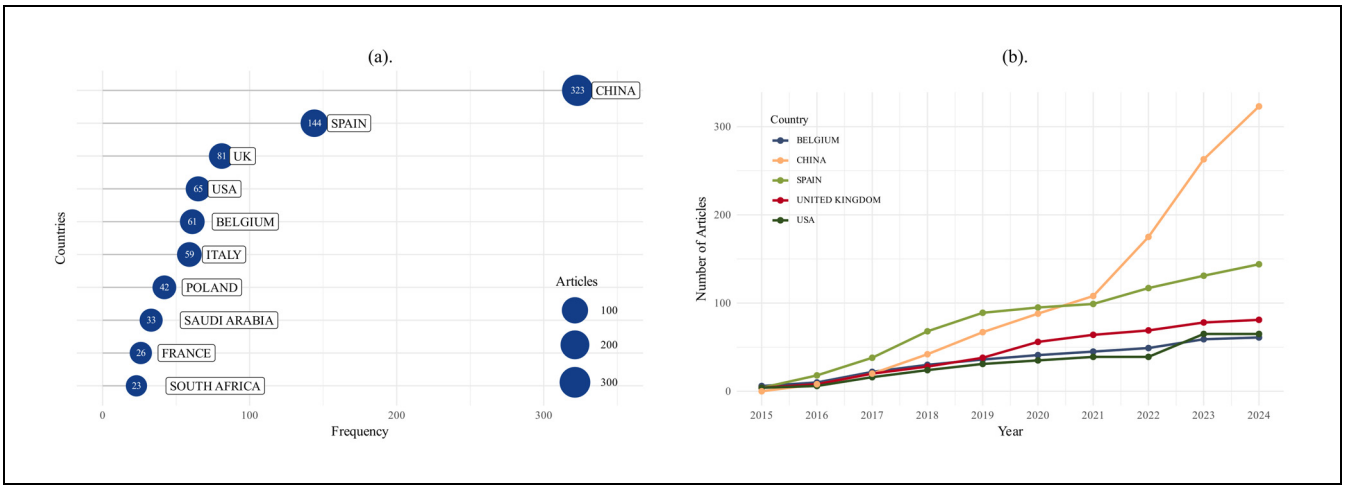

Figure 3. (a) Most productive countries in CTS from 2015 to 2024 and (b) cumulative annual production of CTS articles by country from 2015 to 2024.

The number of published articles depicted in Figure 2b initially shows a slight decline before stabilizing for nearly 8 years from 2015 to 2021. In 2022, a dramatic surge is observed, with the number of CTS articles doubling compared with the previous year. The frequency of annual publications continues to rise in 2024, approaching three times the total from 2021. This significant increase warrants further investigation in the current study to explore the factors driving the recent surge in the popularity of CTS.

Countries and Institutions. The countries of all authors involved in CTS are shown in Figure 3a. China has produced more than half of the CTS articles over the past decade, making it the top contributing country. Spain ranks second, followed by other countries including the UK, the USA, and Belgium. Figure 3b illustrates the formation of rankings based on cumulative annual production over the 10-year period. Notably, Spain leads in cumulative publication frequency from 2016 to 2020, with China surpassing Spain in 2021. Since 2022, CTS has experienced considerable momentum in China, while Spain, the UK, the USA, and Belgium have continued to show stable growth. The significant disparity between China and other countries shown in Figure 3b partly explains the overall sharp increase in CTS articles in 2022.

To gain further insight into the countries involved in CTS, the most contributing countries are identified based on the corresponding authors (see Figure 4). The ranking is similar to the overall distribution shown in Figure 3a. However, there are notable differences in academic collaboration, as indicated by the MCP (Multiple Country Publication) metric. A relatively larger proportion of MCP articles is observed in the UK, Belgium, Poland, Switzerland, and Australia, whereas high-production countries in CTS, such as China and Spain, show a tendency for more intra-country collaborations.

The collaboration map between countries, based on the interactions between corresponding authors and other authors, is shown in Figure 5. Despite the relatively lower MCP percentage, China still has the largest number of collaborative works in CTS. Notably, wide lines connecting South Africa, the UK, Spain, Australia, and China suggest strong collaboration. Multiple lines also cluster within Europe. However, minimal involvement is observed in many countries in the Global South, including those in South America, Africa, and Southeast Asia. One possible explanation is the limited academic resources and funding issues in the Global South. Compared with the Global South, the Council of Europe has supported projects to create corpora for the languages of the Union (Laviosa, 2004). South Africa, which held the first international conference of CTS in 2003, is more likely to collaborate with other countries in CTS, thus forming the exception in the Global South. The inter-country collaboration patterns suggest that languages from countries with less involvement tend to receive less attention, reducing linguistic diversity in CTS. This finding supports the limitation highlighted in a previous review (Laviosa & Liu, 2021b), which notes that languages with limited diffusion are rarely explored in CTS.

Figure 6 illustrates the most productive institutions and their cumulative production frequency. Ghent University and The Hong Kong Polytechnic University

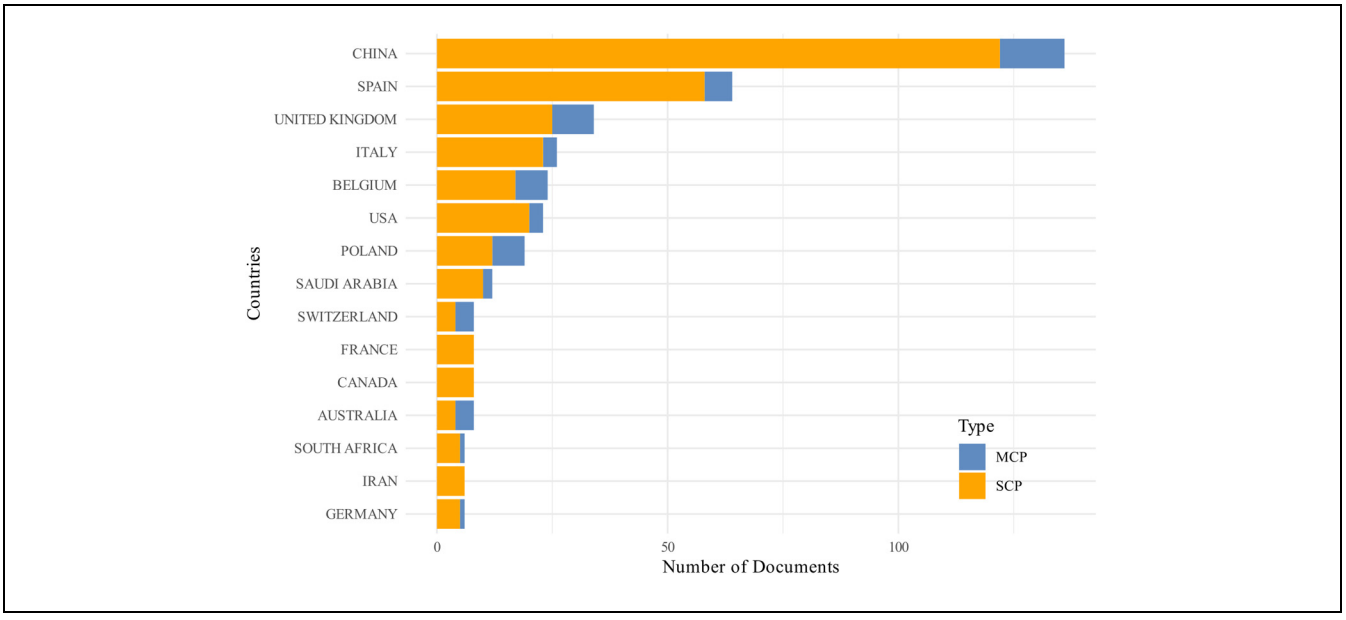

Figure 4. Contribution of countries by corresponding authors.
Note. MCP = multi-country partnership indicates international collaboration; SCP = single-country partnership indicates domestic collaboration.

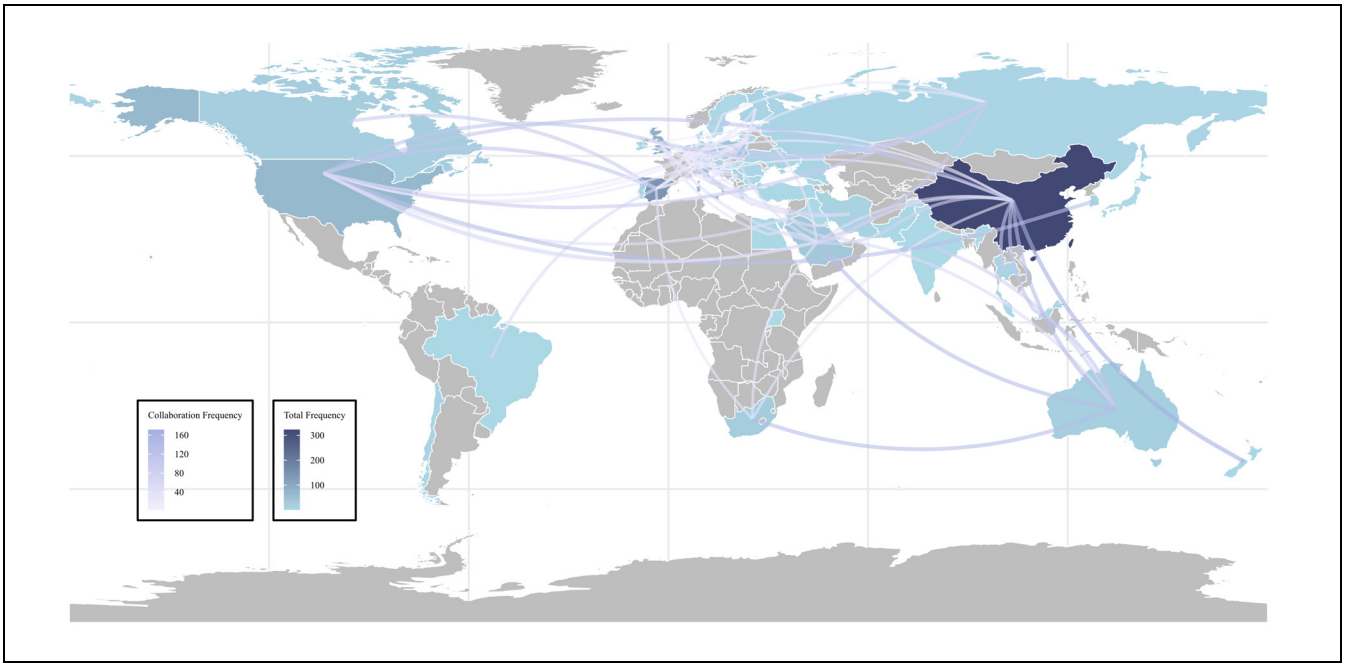

Figure 5. World map of academic collaborations in CTS from 2015 to 2024.

have been key hubs in CTS over the past decade. Among the top 10 productive institutions, three are based in China. Ghent University has played a leading role in CTS since 2016. The CTS output at Macquarie University has remained steady since 2021. In 2022, a notable surge in production occurred at The Hong Kong Polytechnic University, which has continued to maintain its momentum into 2024.

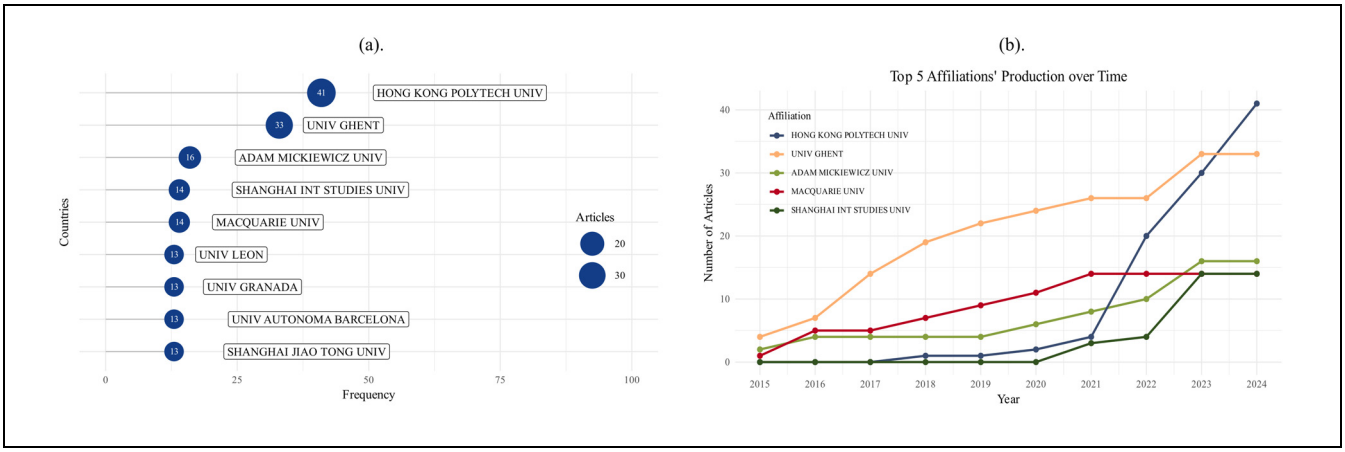

Figure 6. (a) Most productive institutions in CTS from 2015 to 2024 and (b) annual productions of institutions in CTS from 2015 to 2024.

Table 1. Overall Key Terms from Titles, Authors' keywords and Abstracts in CTS Articles.

Rank	Term	Frequency	Range	Rank	Term	Frequency	Range
1	CTS	317	203	8	Translated texts	82	57
2	Source texts	134	94	9	News translation	81	43
3	Parallel corpus	117	94	10	Political discourse	80	35
4	Discourse analysis	116	62	11	Translation strategies	80	49
5	English translations	113	78	12	Translation universals	60	38
6	Machine translation	101	46	13	Comparable corpus	58	43
7	Target texts	100	73	14	Literary translation	51	35

Principal Research Themes

Analysis of Term Frequency. As outlined in Section 3.2, key terms were extracted from 2-grams to 3-grams in article titles, author-supplied keywords, and abstracts, and were supplemented by 1-grams from author keywords. To identify the most salient themes, terms were filtered based on their dispersion across at least 7% of the corpus (minimum threshold: 35 articles). The filtering process yielded the results shown in Table 1. Notably, 2-grams dominate the list of prominent terms, with the exception of the collective term *CTS* (corpus translation studies and its variants).

The high-frequency terms in Table 1 align with foundational frameworks in translation studies, particularly Baker's (1995) research paradigm. These include core concepts such as *source texts*, *target texts*, and corpus typologies (*parallel* corpus, *comparable* corpus). The enduring prominence of these terms, first articulated nearly three decades ago, underscores their sustained methodological relevance in contemporary CTS. Subsequent high-frequency terms reflect evolving research foci, including *discourse analysis*, *machine translation*, *literary translation*, and *news translation*, indicating a diversification of scholarly inquiry within the field.

Burst Detection of Terms. The overall high-frequency terms provide an overview of the research paradigms and fields, but they do not capture sudden shifts in relatively low-frequency terms over the past decade. Kleinberg's (2003) burst detection algorithm was applied to more effectively identify subtle changes among the vast array of terms. The results of the burst detection analysis are shown in Figure 7.

The horizontal line graph illustrates the sudden increase in terms related to CTS from 2015 to 2024. The colored lines for each term indicate the onset, duration, and end of the bursts. For most terms, the burst lasts no more than 2 years, with exceptions including *literary translation*, *target languages*, *text mining*, and *machine translation*. Compared with the first three periods, more burst terms emerged between 2018 and 2021. In the early years of CTS (2015–2019), terms related to discourse analysis (e.g., *English news*, *institutional translation*) and

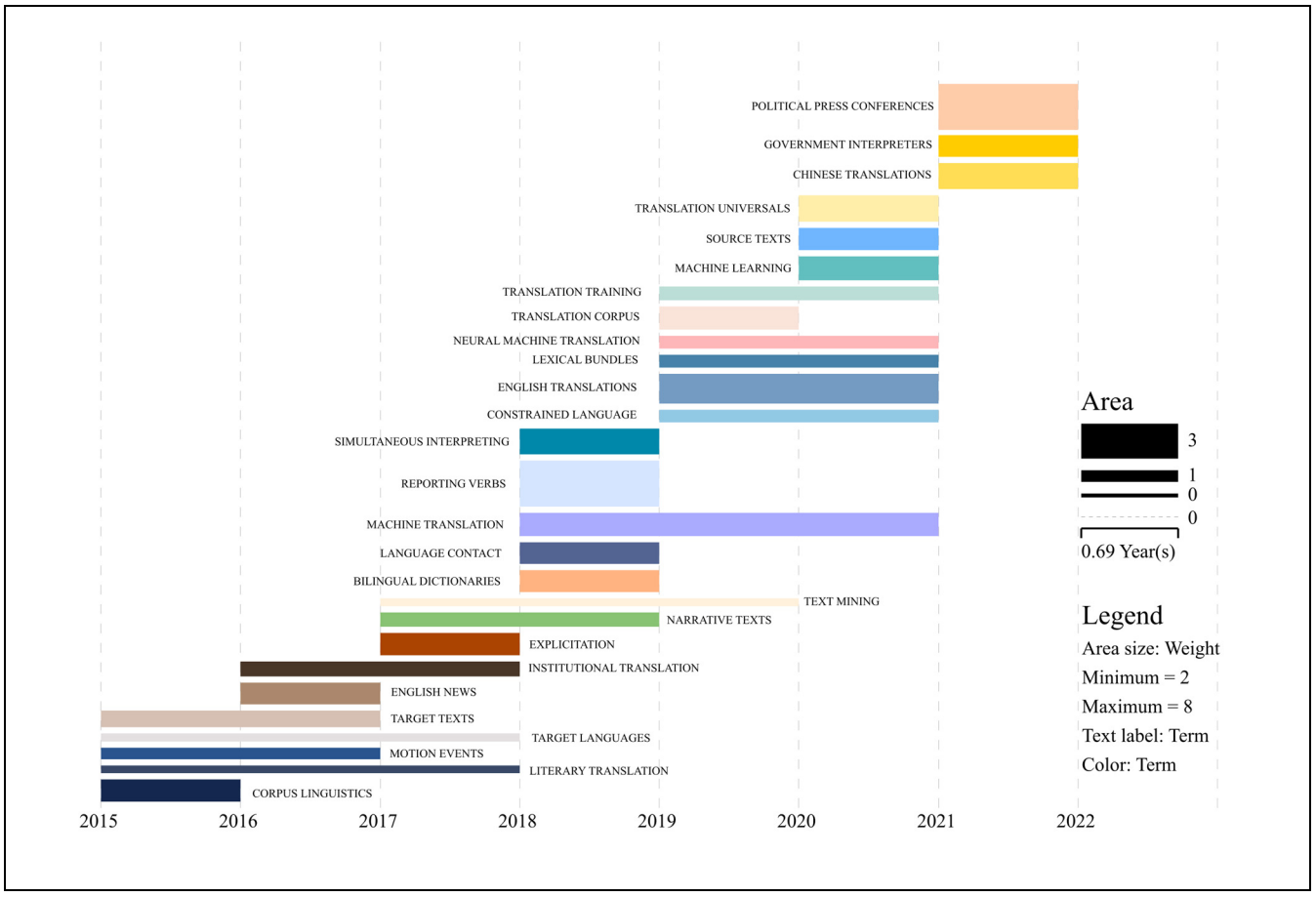

Figure 7. Horizontal line graph of burst detection.

literary stylistics (e.g., *literary translation*, *narrative texts*) were particularly prominent. In terms of text mode, more interpreting products were included in CTS, as evidenced by the prominence of *simultaneous interpreting* (2018–2019), *government interpreters* (2021–2022), and *political press conferences* (2021–2022). Notably, computational techniques have influenced CTS throughout the last decade, including *corpus linguistics*, *text mining*, *machine translation*, *neural machine translation*, and *machine learning*. Another notable thread has emerged around bilingual and multilingual topics, such as *language contact* (2018–2019), *bilingual dictionaries* (2018–2019), and *constrained language* (2019–2021).

The burst detection analysis suggests that CTS research has broadened its scope by incorporating influences from adjacent disciplines, such as sociolinguistics, communication, literary studies, and computational linguistics. Notably, the framework of constrained communication appears to support the integration of these adjacent disciplines by offering a means of modeling the factors that influence language use (Lanstyák & Heltai, 2012). According to Kotze and van Rooy (2024), these specific factors range from language activation, modality and register, and text production, to proficiency and task expertise, corresponding to many terms illustrated in Figure 7. For instance, the prominence of terms related to interpreting products suggests cognitive factors in language activation, situational variables in text production, and the interpreters' language background. The threads around bilingual and multilingual topics directly reflect the complex language activation involved in translation activities.

Term Co-Occurrence. The overall term frequency and burst detection do not reveal how the terms are interconnected. Although isolated terms can be linked in the frequency list and burst detection in an intuitive manner, it remains challenging to draw conclusions about how one group of terms is connected with another. To illustrate the interconnections between term groups, a co-occurrence network of 2-grams in abstracts was generated (see Figure 8).

The co-occurrence network, based on the Louvain cluster algorithm, divides the terms into five main

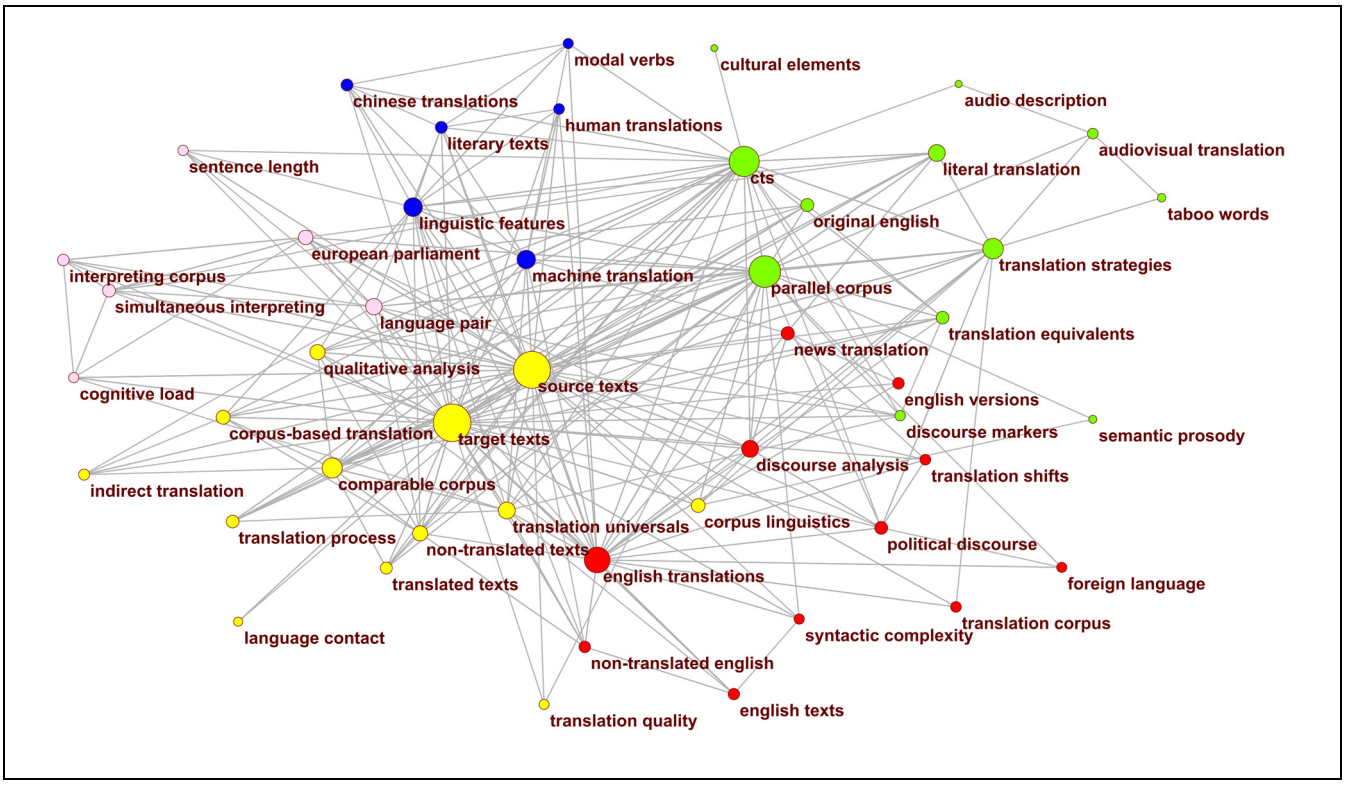

Figure 8. Co-occurrence network of 2-grams in CTS abstracts from 2015 to 2024.

clusters. The core research paradigm of CTS is represented at the center of the yellow and green clusters, encompassing terms such as *source texts*, *target texts*, *comparable corpus*, and *parallel corpus*. Interestingly, the vertex parallel corpus is positioned closer to the green and red clusters, where the key research objects in discourse analysis include *news translation*, *discourse markers*, *semantic prosody*, and *audiovisual translation*. In comparison with parallel corpus, the term comparable corpus is surrounded by terms such as *translation universals*, *translated texts*, and *non-translated texts*. According to Baker (1995), comparable corpora facilitate the exploration of the nature of translated texts, while parallel corpora are typically employed in studying translating norms in a specific context. The two contrasting vertex organizations appear to indicate different research paradigms in corpus design.

The same Louvain algorithm for the co-occurrence network was applied to the dataset for the first period (2015–2019) and the second period (2020–2024) to illustrate diachronic changes in term relationships. Compared with Figure 8, Figure 9a shows a slight decrease in the number of vertices meeting the minimum condition of two edges, attributable to the smaller number of articles in the first period. Consequently, the less dense connections between the clusters were reflected in the absence of bridging vertices between the blue and pink clusters. The blue cluster of *language contact* in Figure 9a is absent, while the vertex *audiovisual translation* became connected to *translation strategies* within the green cluster in Figure 9b. The introduction of new terms into CTS is also evident in the peripheral vertices, which are connected to only one or two central vertices, such as *syntactic complexity*, *taboo words*, and *machine learning*.

CTS has increasingly integrated concepts from quantitative linguistics to address longstanding challenges in translation studies. As indicated by Figure 9b, some terms like *syntactic complexity*, *cognitive load*, and *machine learning*, have gained prominence in recent CTS research. A notable example was the application of dependency grammar to measure syntactic complexity in CTS, which exemplifies successful interdisciplinary collaboration (H. Xu & Liu, 2023; Fan & Jiang, 2019). The flexibility of dependency grammar has enabled the holistic syntactic description of translated languages, thereby revealing the role of working memory in shaping dependency structures (C. Xu & Liu, 2022). This trend suggests that advanced measurement techniques from quantitative linguistics will likely play an increasingly influential role in shaping the future of CTS.

To provide a more accurate description of the two networks, the betweenness and closeness values for the main

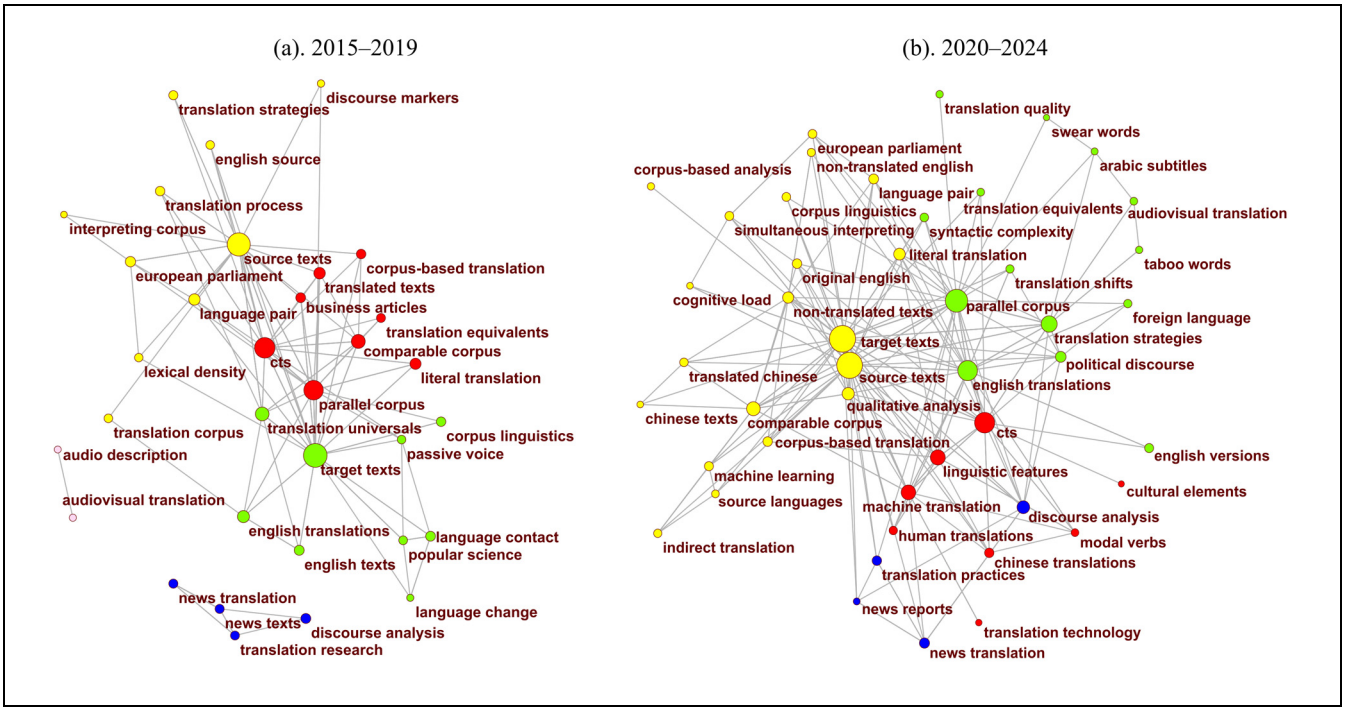

Figure 9. Co-occurrence network of 2-grams in CTS abstracts across two periods.

Table 2. Properties of Prominent Vertices in the Co-Occurrence Network across Two Periods.

2015–2019			2020–2024		
Vertex	Betweenness	Closeness	Vertex	Betweenness	Closeness
Target texts	135.564	0.031	Target texts	277.684	0.017
Source texts	93.45	0.029	Source texts	265.831	0.017
CTS	33.605	0.027	Parallel corpus	175.908	0.015
Parallel corpus	29.278	0.026	Translation strategies	90.532	0.013
Translation universals	5.582	0.023	English translations	88.3	0.014
Comparable corpus	3.529	0.023	CTS	86.577	0.014
English translations	3.403	0.02	Machine translation	54.711	0.013
European parliament	1.779	0.02	Comparable corpus	13.208	0.012
Passive voice	1.293	0.02	Linguistic features	9.804	0.013
News texts	0.625	0.333	Discourse analysis	6.868	0.012

vertices were calculated and extracted (see Table 2). The top vertices in terms of betweenness, which represent core concepts in the research paradigm, show little change in the vertex list. However, the main vertices in the second period generally exhibit higher betweenness values, indicating a tendency for these central vertices in CTS to serve as hubs along the shortest paths between other vertices. Interestingly, an opposite trend is observed in the closeness values, which decline across the two periods. The lower closeness values indicate fewer direct links between the main vertices and the peripheral vertices in the network. The peripheral vertices tend to represent more concrete topics, indirectly linked to the core concepts through mediating vertices. For instance, specific topics in Figure 9b, such as *indirect translation* and *audiovisual translation*, are not directly connected to the core concepts. The changes in betweenness and closeness corroborate our earlier observation that research fields in CTS are becoming both broader and more specific over the past decade.

Table 3. The Most Cited Global Works.

Cited references	LCS	Cited references	LCS
Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds), *Text and technology: In honor of John Sinclair* (pp. 233–250). John Benjamins Publishing Company.	80	Laviosa, S. (2002). Corpus-based translation studies: Theory, findings, applications. Rodopi.	31
Toury, G. (1995). *Descriptive translation studies and beyond*. John Benjamins Publishing Company.	37	Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager* (pp. 175–186). John Benjamins Publishing Company.	30
Newmark, P. (1988) . *A textbook of translation*. Prentice Hall.	37	Blum-Kulka, S. (1986). Shift of cohesion and coherence in translation. In L. Venuti (Ed.), *The translation studies reader* (pp. 298–313). Routledge.	27
Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, *43*(4), 557–570.	31	Olohan, M. (2004). *Introducing corpora in translation studies*. Routledge.	26
Biber, D. Johansson, S. Leech, G. Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.	31	Venuti, L. (2008). *The translator's invisibility: A history of translation* (2nd ed.). Routledge.	26

Note. See Appendix for the complete list of works.

Intellectual Structures in CTS

The investigation of the principal research themes in Section 4.2 revealed both the core concepts of CTS and their intricate relationships with other topics. Since the birth of CTS, scholars have clarified, refined, and critiqued specific concepts, thereby adding complexity to the field's conceptual landscape. A more comprehensive, document-based account of the discipline's development can be achieved through citation analysis. As Gile (2005) notes, citation analysis can document and quantify impressions of specific areas over time. In the following sections, both citation and co-citation analyses are employed to examine the intellectual structures underpinning CTS.

The Most Cited Global Works. A total of 17,631 references were identified in 477 research articles in CTS. Among these, the top 10 cited works were extracted and presented in Table 3. These highly cited works might not be directly related to the topics in the recent studies, but they could serve as conventional citations that support the ideas presented by the authority cited (Gile, 2005). A close review of these works was conducted to summarize and analyze the influence of early intellectual structures in CTS.

Recent studies in CTS continue to be shaped by the research paradigm established approximately three decades ago. The foundational research paradigms in CTS emerged in the early 1990s, when Baker (1993, 1995, 1996) and Laviosa (1998, 2002) sought to lay the groundwork for the field by conducting preliminary studies and exploring potential topics, methods, and applications. These early studies also stimulated efforts toward the systematic development and dissemination of CTS methodology (Olohan, 2004). Notably, translation universals and their associated debates have continued to influence many CTS studies over the past decade. The key hypotheses related to translation universals have remained central to CTS research, particularly simplification (Baker, 1993, 1996; Laviosa, 1998) and explicitation (Baker, 1993, 1996; Blum-Kulka, 1986; Olohan & Baker, 2000). The most cited global works likewise reflect the enduring influence of certain translation theories and frameworks from adjacent disciplines, such as descriptive translation studies (Toury, 1995) and corpus linguistics (Biber et al., 1999).

The Most Cited Local Works. Among the 477 articles in corpus-based translation studies (CTS), Kruger and van Rooy (2016a) achieve the highest LCS (21) and GCS (51). The significant influence of Kruger and van Rooy's (2016a) study has spurred a growing interest in investigating the linguistic features of constrained or mediated languages, as evidenced by numerous subsequent studies in Table 4. These studies predominantly adopted corpus designs that examine the effects of various factors and their interactions on the linguistic characteristics of constrained languages. For instance, Kajzer-Wietrzny (2015)

Table 4. The Most Cited Local Works.

Paper	LCS	GCS	Paper	LCS	GCS
Kruger, H., & van Rooy, B. (2016a). Constrained language: A multidimensional analysis of translated English and a non-native indigenized variety of English. *English World-Wide*, *37*(1), 26–57.	21	51	Kruger H., & van Rooy, B. (2016b). Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. *Language Sciences*, *56*, 118–131.	8	20
Kruger, H. (2019). That again: A multivariate analysis of the factors of conditioning syntactic explicitness in translated English. *Across Languages and Cultures*, *20*(1), 1–33.	11	17	Ferraresi, A., Bernardini, S., Petrović, M. M. & Lefer, M. A. (2018). Simplified or not simplified? The different guises of mediated English at the European Parliament. *Meta*, *63*(3), 717–738.	8	15
Kajzer-Wietrzny, M. (2015). Simplification in interpreting and translation. *Across Languages and Cultures*, *16*(2), 233–255.	10	15	Oster, U., & Molés-Cases, T. (2016). Eating and drinking seen through translation: A study of food-related translation difficulties and techniques in a parallel corpus of literary texts. *Across Languages and Cultures*, *17*(1), 53–75.	8	12
Liu, K., Liu, Z., & Lei, L. (2022). Simplification in translated Chinese: An entropy-based approach. *Lingua*, *275*, 103364.	9	13	Ruano, P. (2017). Corpus methodologies in literary translation studies: An analysis of speech verbs in four Spanish translations of Hard Times. *Meta*, *62*(1), 94–113.	6	9
Plevoets, K., & Defrancq, B. (2018). The cognitive load of interpreters in the European Parliament: A corpus-based study of predictors for the disfluency uh(m). *Interpreting*, *20*(1), 1–29.	8	47	Cvrček, V., & Chlumská, L. (2015). Simplification in translated Czech: A new approach to type-token ratio. *Russian Linguistics*, *39*, 309–325.	4	19

Note. Local Citation Score (LCS) refers to the number of times a work in the local dataset is cited by other works within that same dataset. Global Citation Score (GCS) refers to the number of times a work is cited in the Web of Science (WoS) database. See Appendix for the complete list of works.

analyzed the influence of production modes and language pairs on lexical features in interpreted English, translated English, and native English. Similarly, Ferraresi et al. (2018) investigated lexical features in translated and interpreted texts across different language pairs, concluding that source languages exert a stronger influence on lexical features than production modes. Compared to earlier CTS research, recent studies demonstrate a shift toward examining more specific factors in translated texts, such as language varieties in contact situations, source languages, and production modes. This evolution reflects a broader trend toward methodological refinement and a deeper engagement with the complexities of translation processes.

The hypotheses related to translation universals have been prominent topics in recent CTS articles, many of which have received high citation counts. Nearly half of the articles listed in Table 4 focus on the two translation universals: simplification and explicitation. Regarding simplification, recent studies have demonstrated the influence of multiple variables on constrained languages (Liu et al., 2022; Ferraresi et al., 2018; Kajzer-Wietrzny, 2015). It is worth noting that critical reflection on previous indicators can also contribute to a deeper conceptualization of simplification. For example, Cvrček and Chlumská (2015) examined the statistical limitations of the traditional type-token ratio (TTR) in measuring lexical simplification and proposed the new indicator. Their critique aligns with Sampson's (2024) call of "getting serious with statistics" when addressing the challenges of applying TTR in linguistic research (p. 47). Liu et al. (2022) expanded the scope of simplification from the lexical level to the syntactic level by employing the informetric indicator of entropy of parts of speech. In the case of explicitation, research has primarily focused on reported-speech constructions, exploring both intralinguistic and cross-linguistic factors (Kruger, 2019; Kruger & van Rooy, 2016b). Given the sustained scholarly interest in translation universals, we anticipate that this topic will remain a central focus in future studies of CTS.

The rise in interpreting studies, as shown in Table 4, aligns with the pink cluster identified in the term co-occurrence network in Figure 8. In addition to interpreting studies linked to translation universals, recent research has also focused on specific phenomena, such as disfluency (Plevoets & Defrancq, 2018) and the relationship between hedges and gender (Magnifico & Defrancq, 2017). These phenomena, including disfluency and hedges, often have a more direct connection to real-world practices, highlighting the potential of CTS to contribute to translation quality assessment (Chesterman, 2004). Traditional corpus linguistics methods, such as keyword analysis and concordances, continue to be employed to support qualitative analyses and contextual interpretation in stylistics (Johnson, 2016; Ruano, 2017).

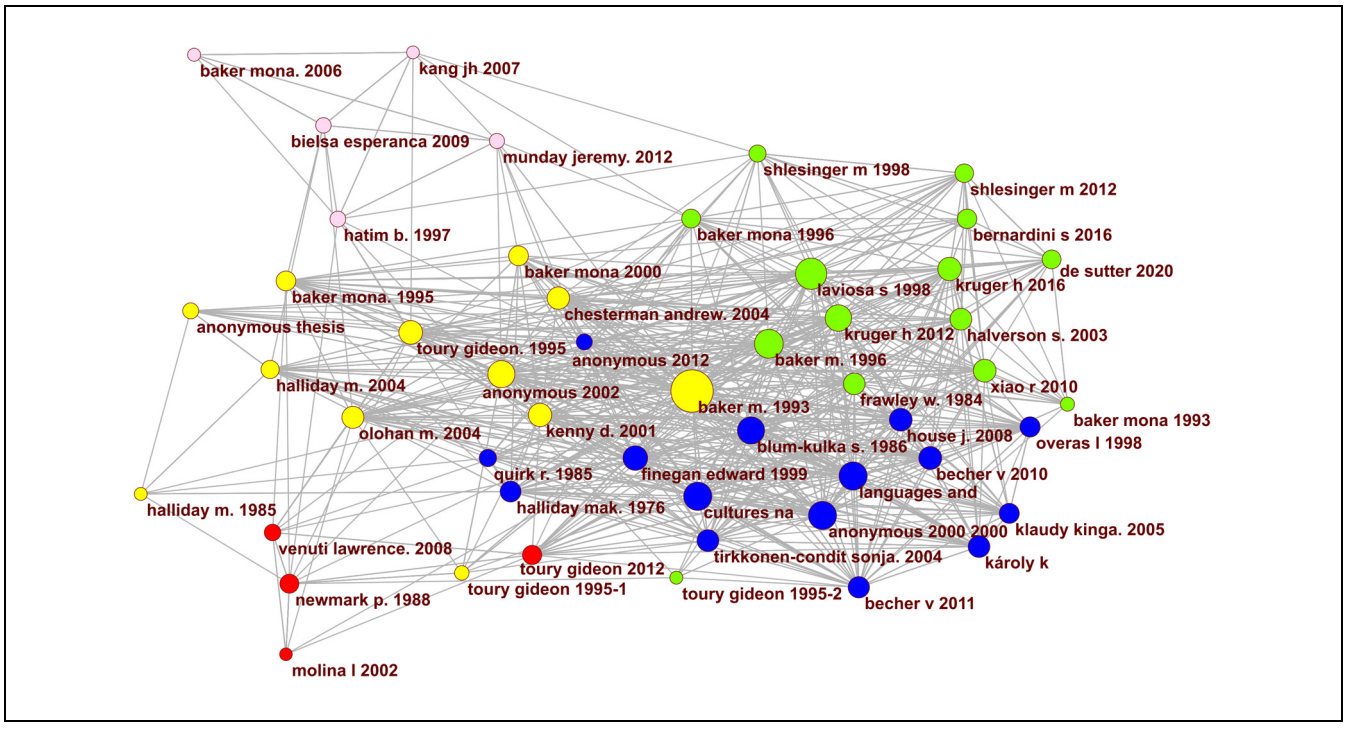

Figure 10. Co-citation network for the cited works in CTS from 2015 to 2024.

With the ongoing advancements in digital humanities, it can be predicted that more sophisticated tools for identifying linguistic and multimodal features will be integrated into CTS research.

Recent highly cited articles in CTS have demonstrated significant methodological advancements in corpus compilation, measurement indicators, and statistical analyses. As highlighted in the preceding discussion, contemporary research increasingly incorporates the control or manipulation of multiple social and individual variables in corpus design to address specific research questions. In terms of measurement, CTS has expanded its scope to investigate both local and global linguistic features, including lexical bundles (Y. Li & Halverson, 2020), the aggregation of diverse linguistic features (Kruger & van Rooy, 2016a), and holistic indicators such as entropy (Liu et al., 2022). Furthermore, recent CTS research has adopted more sophisticated statistical methods to analyze complex relationships among multiple variables, including multidimensional analysis (Kruger & van Rooy, 2016a), generalized additive mixed-effects models (Plevoets & Defrancq, 2018), conditional inference trees (Kruger, 2019), and principal component analysis (de Sutter et al., 2018).

Co-Citation Analysis. A close reading of globally and locally influential works with high citation counts enables us to identify the early intellectual structures and recent developments in research topics within CTS. The boundaries between various topics in CTS are not rigid, as there often exists a continuum connecting different research areas. Since exploring this continuum can illuminate the complex relationships between core areas, a co-citation network was constructed for the most frequently cited works, as illustrated in Figure 10. A few works remain marked as anonymous in the network, due to incomplete information in the WoS database. The incomplete references were retained, as they might function as important bridges connecting other vertices within the network.

Several co-citation clusters can be identified based on the colors and distances of the vertices in the network. At the center of the network, the yellow, blue, and green vertices represent early foundational works in CTS, which tend to be larger due to their higher co-citation scores (e.g., Baker, 1993; Blum-Kulka, 1986; Kenny, 2001; Laviosa, 1998). The blue vertices adjacent to the green cluster are linked to critical reflections on translation universals (e.g., Becher, 2010; House, 2008). The top-right section of the network pertains to interpreting studies, featuring works by Shlesinger (1998), Shlesinger and Ordan (2012), and Bernardini et al. (2016). The vertices near the right side of the green cluster mainly correspond to studies conducting multidimensional analyses of linguistic features in translational languages (de Sutter & Lefer, 2020; Kruger & van Rooy, 2016a; Xiao, 2010).

Another notable pattern in Figure 10 is that both the cited prescriptive and descriptive grammar books tend to cluster at the bottom left (Biber et al., 1999; Quick et al., 1985). Although these comprehensive works are not core issues in CTS, they assist researchers in explaining certain grammatical features of translational languages by linking structural properties to functional explanations.

The pink vertices located in the top-left section of the network, together with some red vertices, primarily focus on discourse analysis and sociological perspectives within translation studies. News discourse appears to be the typical research object in this field (e.g., Bielsa & Bassnett, 2009; Kang, 2007). The vertices related to systemic functional linguistics (SFL), marked on the left side of the yellow clusters and one blue vertex (Halliday, 1985, 2004; Halliday & Hasan, 1976), also form a subcluster on the right side of the network. The affinity between SFL and discourse analysis is likely due to the SFL framework's capacity to link linguistic choices to social contexts. Notably, Toury's (1995) work serves as the bridging vertex between the network center and the three clusters (pink, yellow, and red). The three clusters on the left in Figure 10 further support the earlier observation of distinct research paradigms. As shown in the term co-occurrence network in Figure 8, the field of discourse analysis, compared with the field of translation universals surrounding the comparable corpus, forms another main cluster which is closely associated with the parallel corpus.

Conclusion

The advancement of digital humanities has introduced new complexities to the field of CTS. To examine the historical development and emerging trends in CTS, this study conducts a bibliometric analysis of 477 research articles published over the past decade. Based on the results from basic profiles, principal research themes, and intellectual structures, the main findings are summarized as follows:

(1) The general frequency of the journals in the dataset reveals the multidisciplinary features of CTS. China and European countries have been two centers in CTS where several universities are productive in this field. The Global South is less involved in CTS, thus reducing linguistic diversity.
(2) The popularity of computational techniques, interpreting studies, and constrained communication is identified from the term threads in the burst detection analysis. Term co-occurrence networks identify two research fields, namely translation universals and discourse analysis, which surround the vertices of comparable corpora and parallel corpora, respectively.
(3) The topic of translation universals is reaffirmed as the main theme in CTS through the close reading of the most cited global and local works. Notably, the co-citation network corroborates the two research fields identified in the term co-occurrence analysis. The co-citation clusters further reveal the affinity between SFL and discourse analysis.

Our bibliometric analysis of CTS supports Kuhn's (1996) view of the relationship between research paradigms and normal science. Core concepts such as source texts, target texts, and various corpus types remain central to the cluster, reflecting well-established patterns in CTS research design. Recent empirical studies in CTS, as part of the normal science in the field, have either reaffirmed or updated translation hypotheses by incorporating new measurements of linguistic features along with advanced statistical methods. In the interpretation of linguistic features, recent studies, especially corpus-based discourse analyses of translations, also attempt to associate textual entities with sociological factors. Compared with the results of Pang's (2019) bibliometric analysis, corpus-based interpreting studies have gained significant attention, forming a distinct cluster in the network. These trends suggest that recent developments in CTS not only reinforce the phenomena and theories established through scientific practices within the research paradigm but also expand their application under the influence of other disciplines.

This study is not without limitations. The dataset is sourced from research articles in the SSCI and A&HCI core collections in WoS, excluding other academic genres such as book chapters. In addition, as Biblioshiny does not support network construction based on multiple *n*-grams, this study is limited to using 2-grams as proxies for most terms. Future research may expand the dataset to include articles from additional academic databases and employ more advanced bibliometric methods to uncover more nuanced connections and trends. Bibliometric methods may also be integrated into the program of translation studies, as they assist novice researchers in navigating the increasingly complex academic landscape.

ORCID iDs

Yiyang Hu https://orcid.org/0009-0001-4101-8667
Kanglong Liu https://orcid.org/0000-0003-3962-2563

Ethical Considerations

This study does not involve human participants or animal subjects; therefore, ethical review and approval are not required.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is supported by The Hong Kong Polytechnic University (Grant Nos. P0051009 and P0046370).

Declaration of Conflicting Interests

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

Data Availability Statement

Data concerning the study are available on Open Science Framework (https://osf.io/t6qpd/overview?view_only = bbc907cf14744caf8b9c8f71b848e0a4).

References

Aria, M., & Cuccurullo, C. (2017). Bibliometrix : An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*, 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–250). John Benjamins Publishing Company.

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, *7*(2), 223–243. https://doi.org/10.1075/target.7.2.03bak

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager* (pp. 175–186). John Benjamins Publishing Company.

Becher, V. (2010). Abandoning the notion of "translation-inherent" explicitation: Against a dogma of translation studies. *Across Languages and Cultures*, *11*(1), 1–28. https://doi.org/10.1556/acr.11.2010.1.1

Bernardini, S., Ferraresi, A., & Miličević, M. (2016). From EPIC to EPTIC : Exploring simplification in interpreting and translation from an intermodal perspective. *Target*, *28*(1), 61–86. https://doi.org/10.1075/target.28.1.03ber

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.

Bielsa, E., & Bassnett, S. (2009). *Translation in global news*. Routledge.

Blum-Kulka, S. (1986/2000). Shift of cohesion and coherence in translation. In L. Venuti (Ed.), *The translation studies reader* (pp. 298–313). Routledge.

Chen, C. (2016). *CiteSpace: A practical guide for mapping scientific literature*. Nova Science Publishers.

Chesterman, A. (2004). Beyond the particular. In A. Mauranen & P. Kujamäki (Eds.), *Translation universals: Do they exist* (pp. 33–50). John Benjamins Publishing Company.

Crosthwaite, P., Ningrum, S., & Schweinberger, M. (2023). Research trends in corpus linguistics: A bibliometric analysis of two decades of Scopus-indexed corpus linguistics research in arts and humanities. *International Journal of Corpus Linguistics*, *28*(3), 344–377. https://doi.org/10.1075/ijcl.21072.cro

Cvrček, V., & Chlumská, L. (2015). Simplification in translated Czech: A new approach to type-token ratio. *Russian Linguistics*, *39*, 309–325. https://doi.org/10.1007/s11185-015-9151-8

de Sutter, G., Cappelle, B., De Clercq, O., Loock, R., & Plevoets, K. (2018). Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translations. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, *16*, 25–39. https://doi.org/10.52034/lanstts.v16i0.440

de Sutter, G., & Lefer, M. A. (2020). On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives*, *28*(1), 1–23. https://doi.org/10.1080/0907676x.2019.1611891

Even-Zohar, I. (1979). Polysystem theory. *Poetics Today*, *1*(1/2), 287–310. https://doi.org/10.2307/1772051

Fan, L., & Jiang, Y. (2019). Can dependency distance and direction be used to differentiate translational language from native language? *Lingua*, *224*, 51–59. https://doi.org/10.1016/j.lingua.2019.03.004

Ferraresi, A., Bernardini, S., Petrović, M. M., & Lefer, M. A. (2018). Simplified or not simplified? The different guises of mediated English at the European Parliament. *Meta*, *63*(3), 717–738. https://doi.org/10.7202/1060170ar

Gile, D. (2000). The history of research into conference interpreting: A scientometric approach. *Target*, *12*(2), 297–321. https://doi.org/10.1075/target.12.2.07gil

Gile, D. (2005). Citation patterns in the T&I didactics literature. *Forum*, *3*(2), 85–103. https://doi.org/10.1075/forum.3.2.05gil

Gu, C. (2024). One-third of a century on: The state of the art, pitfalls, and the way ahead relating to digital humanities approaches to translation and interpreting studies. *Digital Scholarship in the Humanities*, *39*(1), 154–161. https://doi.org/10.1093/llc/fqad076

Halliday, M. A. K. (1985). *An introduction to functional grammar*. Edward Arnold.

Halliday, M. A. K. (Ed.). (2004). *An introduction to functional grammar* (3rd ed.). Routledge.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

House, J. (2008). Beyond intervention: Universals in translation. *TRANS*, *1*(1), 6–19.

Hu, K. (2015). *Introducing corpus-based translation studies*. Springer.

Hu, K., & Wang, X. (2022). Translation studies in digital humanities: State of the art, issues and future directions. *Foreign Languages and Their Teaching*, *327*, 111–121.

Hyland, K., & Jiang, F. (2021). A bibliometric study of EAP research: Who is doing what, where and when? *Journal of English for Academic Purposes*, *49*, 1–12. https://doi.org/10.1016/j.jeap.2020.100929

Johnson, J. H. (2016). A comparable comparison? A corpus stylistic analysis of the Italian translation of Julian Barnes' Il

Senso di una fine and the original text the sense of an ending. *Language and Literature*, *25*(1), 38–53. https://doi.org/10.1177/0963947015623360

Kajzer-Wietrzny, M. (2015). Simplification in interpreting and translation. *Across Languages and Cultures*, *16*(2), 233–255. https://doi.org/10.1556/084.2015.16.2.5

Kang, J. (2007). Recontextualization of news discourse: A case study of translation of news discourse on North Korea. *Translator*, *13*(2), 219–242. https://doi.org/10.1080/13556509.2007.10799239

Kenny, D. (2001). *Lexis and creativity in translation: A corpus-based approach*. Routledge.

Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, *7*, 373–397. https://doi.org/10.1023/a:1024940629314

Kotze, H., & van Rooy, B. (2024). Introduction: The constrained communication framework for studying contact-influenced varieties. In B. van Rooy & H. Kotze (Eds.), *Constraints on language variation and change in complex multilingual contact settings* (pp. 1–28). John Benjamins Publishing Company.

Kruger, H. (2019). That again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures*, *20*(1), 1–33. https://doi.org/10.1556/084.001

Kruger, H., & van Rooy, B. (2016a). Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide*, *37*(1), 26–57. https://doi.org/10.1075/eww.37.1.02kru

Kruger, H., & van Rooy, B. (2016b). Syntactic and pragmatic transfer effects in reported-speech constructions in three contact varieties of English influenced by Afrikaans. *Language Sciences*, *56*, 118–131. https://doi.org/10.1016/j.langsci.2016.04.003

Kuhn, T. S. (Ed.). (1996). *The structure of scientific revolutions* (3rd ed.). The University of Chicago.

Lanstyák, I., & Heltai, P. (2012). Universals in language contact and translation. *Across Languages and Cultures*, *13*(1), 99–121. https://doi.org/10.1556/acr.13.2012.1.6

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, *43*(4), 557–570. https://doi.org/10.7202/003425ar

Laviosa, S. (2002). *Corpus-based translation studies: Theory, findings, applications*. Rodopi.

Laviosa, S. (2004). Corpus-based translation studies: Where does it come from? Where is it going? *Language Matters*, *35*(1), 6–27. https://doi.org/10.1080/10228190408566201

Laviosa, S., & Liu, K. (2021a). Corpora: A lens into translation phenomena. *Translation Quarterly*, *101*, 1–4.

Laviosa, S., & Liu, K. (2021b). The persuasiveness of corpora in translation studies. *Translation Quarterly*, *101*, 5–20.

Lei, L., & Liu, D. (2019). Research trends in applied linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics*, *40*(3), 540–561. https://doi.org/10.1093/applin/amy003

Liao, S., & Lei, L. (2017). What we talk about when we talk about corpus: A bibliometric analysis of corpus-related research in linguistics (2000–2015). *Glottometrics*, *38*, 1–20.

Li, L., & Liang, L. (2024). A bibliometric analysis of book reviews published in translation journals between 2010 and 2021. *Journal of Scholarly Publishing*, *55*(1), 139–163. https://doi.org/10.3138/jsp-2023-0028

Li, Y., & Halverson, S. L. (2020). A corpus-based exploration into lexical bundles in interpreting. *Across Languages and Cultures*, *21*(1), 1–22. https://doi.org/10.1556/084.2020.00001

Liu, K., Liu, Z., & Lei, L. (2022). Simplification in translated Chinese: An entropy-based approach. *Lingua*, *275*, 103364.

Maci, S. M. & Sala, M. (Eds.) (2022). *Corpus linguistics and translation tools for digital humanities: Research methods and applications*. Bloomsbury.

Magnifico, C., & Defrancq, B. (2017). Hedges in conferencing interpreting: The role of gender. *Interpreting*, *19*(1), 21–46. https://doi.org/10.1075/intp.19.1.02mag

Martínez-Gómez, A. (2015). Bibliometrics as a tool to map uncharted territory: A study on non-professional interpreting. *Perspectives*, *23*(2), 205–222. https://doi.org/10.1080/0907676x.2015.1010550

Newmark, P. (1988). *A textbook of translation*. Prentice Hall.

Olohan, M. (2004). *Introducing corpora in translation studies*. Routledge.

Olohan, M., & Baker, M. (2000). Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures*, *1*(2), 141–158. https://doi.org/10.1556/acr.1.2000.2.1

Oster, U., & Molés-Cases, T. (2016). Eating and drinking seen through translation: A study of food-related translation difficulties and techniques in a parallel corpus of literary texts. *Across Languages and Cultures*, *17*(1), 53–75. https://doi.org/10.1556/084.2016.17.1.3

Pan, F., & Wu, X. (2024). Mapping intellectual structures and research trends of translation studies: A bibliometric analysis from 2007 to 2021. *Perspectives*, *32*(4), 736–757. https://doi.org/10.1080/0907676x.2023.2275750

Pang, S. (2019). New developments of corpus-based translation studies in the past decade (2008–2018): A bibliometric analysis. *Foreign Languages and Their Teaching*, *309*, 129–140.

Ping, Y. (2021). Towards two decades of journalistic translation research (2000–2019): A corpus-based bibliometric study of the translation studies bibliography. *Meta*, *66*(2), 406–426. https://doi.org/10.7202/1083185ar

Plevoets, K., & Defrancq, B. (2018). The cognitive load of interpreters in the European parliament: A corpus-based study of predictors for the disfluency uh(m). *Interpreting*, *20*(1), 1–29. https://doi.org/10.1075/intp.00001.ple

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*, *25*, 348–349.

Quick, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ruano, P. (2017). Corpus methodologies in literary translation studies: An analysis of speech verbs in four Spanish translations of hard times. *Meta*, *62*(1), 94–113. https://doi.org/10.7202/1040468ar

Sampson, G. (2024). *Structural linguistics in the 21st century*. Cambridge Scholars Publishing.

Sci2 Team. (2009). *Science of science (Sci2) tool*. Indiana University and SciTech Strategies. https://sci2.cbs.iu.edu

Shlesinger, M. (1998). Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta*, *43*(4), 486–493. https://doi.org/10.7202/004136ar

Shlesinger, M., & Ordan, N. (2012). More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target*, *24*(1), 43–60. https://doi.org/10.1075/target.24.1.04shl

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Toury, G. (1995). *Descriptive translation studies – And beyond*. John Benjamins Publishing Company.

Tymoczko, M. (1998). Computerized corpora and the future of translation studies. *Meta*, *43*(4), 652–660. https://doi.org/10.7202/004515ar

Venuti, L. (2008). *The translator's invisibility: A history of translation* (2nd ed.). Routledge.

Wołk, K. (2019). *Machine learning in translation corpora processing*. CRC Press.

Xiao, R. (2010). How different is translated Chinese from native Chinese? *International Journal of Corpus Linguistics*, *15*(1), 5–35. https://doi.org/10.1075/ijcl.15.1.01xia

Xu, C., & Liu, H. (2022). The role of working memory in shaping syntactic dependency structures. In J. W. Schwieter & Z. Wen (Eds.), *The Cambridge Handbook of Working Memory and Language* (pp. 343–367). Cambridge University Press.

Xu, H., & Liu, K. (2023). Syntactic simplification in interpreted English: Dependency distance and direction measure. *Lingua*, *294*, 103607. https://doi.org/10.1016/j.lingua.2023.103607

Yan, J., Li, Q., & Liu, H. (2025). Defining digital humanities and examining its relationship with linguistics through the lens of digital scholarship in the humanities. *Digital Scholarship in the Humanities*, *40*(1), 354–380. https://doi.org/10.1093/llc/fqae075

Appendix

The dataset of research articles, the information of terms, and the word lists to remove noises and combine synonyms are available on Open Science Framework (https://osf.io/t6qpd/?view_only = bbc907cf14744caf8b9c8f71b848e0a4).