Article

Bibliometric analysis and visualization of translation assessment: Research theme, evolution and hotspots

Dai Tao[1,2], Lin Siew Eng[1,*]

[1] Education Department, Faculty of Social Sciences & Liberal Arts, UCSI University, Kuala Lumpur 56000, Malaysia

[2] School of Foreign Languages and Cultures, Southwest University of Science and Techonology, Sichuan 621000, China

* Corresponding author: Lin Siew Eng, linse@ucsiuniversity.edu.my

Abstract: Translation assessment, referring to evaluating various aspects related to translation, is crucial to the improvement of translation competence and quality. This study, taking valid papers on translation assessment published in WoS core collection from 2000 to 2022 as research samples, visualizes and reviews the research theme, research evolutions and emerging hot topics of translation assessment through bibliometric analysis. The number of annual publications shows that despite the general growth, the number of articles on translation assessment fluctuates from year to year. By Cite Space-based analysis of keywords co-occurrence, keyword clustering, time zone and burst detection, it finds that translation assessment mainly covers five themes, namely translation competence, translation quality, machine translation, translation teaching and training, and others. The overall evolution has been a gradual shift from concepts, metrics and evaluations of topics related to translation assessment into a topic of in-depth and interdisciplinary study. The hotspots in recent years include translation competence acquisition, neural machine translation and translation quality estimation. The findings reveal research evolutions and hot spots, propose implications for further translation assessment research and provide references for scholars who are interested in this topic.

Keywords: translation assessment; translation competence; translation quality; bibliometrics analysis

1. Introduction

Translation is the delivery of meaning in one language into another by the translator, which is an indispensable carrier in cross-culture communication (Nida, 2003). With the development of globalization, translation has become increasingly prominent in its role of communication, and the demand for translators and interpreters is becoming more and more urgent (Bai, 2018). Translation assessment (TA) usually refers to the use of various materials and information related to translation to assess translation ability and level. It is like figuratively sitting together with learners and collecting information about their learning process (Kiraly, 2000). In addition to the collection of qualitative and quantitative data, assessment also includes the interpretation of data, the synthesis of various data and finally the comparison and judgment according to the teaching objectives (Xiao, 2012). In short, assessment is the process that aids decision-making by gathering, synthesizing, and interpreting information (Airasian, 1997).

Translation assessment is a fruitful research field in Translation Studies (Han et al., 2022). It is very important because assessment with fine structure and reasonable contents is crucial to the enhancement of translation competence and quality

(Nikolaeva and Korol, 2021; Akulina and Tikhona, 2021). Besides, assessment can be viewed as a pedagogical activity. It is instructive for teaching because it gathers information about learners to detect the individual strengths and weaknesses, then guide them to move in the expected direction (Dickson et al., 2020). Brunette (2000) also notices that different translation scholars interpret the term translation assessment inconsistently, depending on their theoretical positions. Therefore, there are a number of studies about translation assessment, for instance, mixed researches (Nikolaeva and Korol, 2021; Korol, 2020; Korol, 2021), quantitative researches (Akulina and Tikhonova, 2021; Su, 2022; Han et al., 2022; Angelone, 2020) and qualitative researches (Han, 2020; Huang and Xin, 2020). However, there is a lack of systematic reviews on translation assessment despite many studies, let alone any using bibliometric analysis and visualization, such as research hotspots and trends. Systematic review makes it possible for us to know the structure of a field and to regard the field in its entirety (Tymoczko, 2005).

Bibliometric analysis (BA) is defined as a method to evaluate and monitor the progress of given area with mathematical and statistical techniques (Yilmaz, 2019). It can offer quantitative analysis for publications in written material (Ellegaard and Wallin, 2015). The review of studies over periods can not only discover the distributed architecture features and patterns, but also monitor trends in relevant areas and assess these trends as well as future research. Bibliometrics has proved to be a significant tool for reviewing the development in TA (Gile, 2000). By using the bibliometric approaches, we can vividly and visually describe the latest progress, frontier topics and existing gaps in a certain research subject field (Guo et al., 2019). CiteSpace is an automation software for BA developed by Professor Chen Chaomei, which is extensively used to visualize trends and patterns in scientific literature. As a all-in-one software, it integrates functions like drawing visual co-citation map, locating inflection point, searching key node and analyzing regional evolution (Chen, 2006).

This paper analyses publications in the WoS core collection from 2000 to 2022 related to TA using bibliometric techniques and Citespace, aiming to examine four questions:
1) What are the overall trend of TA from 2000–2022?
2) What research themes have emerged on TA from 2000 to 2022?
3) How did research on TA evolve?
4) What are the hotspots of TA research?

By answering the four questions, topics, citations and keywords of the published literature about translation assessment can be visually analyzed, research theme and evolution can be scientifically explored, current emerging hot spots can be revealed, which help propose implications for further research and provide references for scholars who are interested in this topic.

2. Data source and methodology

When analyzing information for a research, the top priority lies in its richness and neutrality. The dataset for this article comes from the Web of Science (WoS), which is considered the most important and commonly utilized scientific database in

various areas (Pranckute, 2021; Janmaijaya et al., 2018). With a strict screening mechanism, WoS is a valuable database for acquiring academic information from around the world. Meanwhile, it merely contains influential academic papers from different subjects, based on Garfield's law of concentration in bibliometrics (Pan et al., 2019). Though there are many other databases, such as Google scholar, the whole attention will be centered on WoS in this study.

Translation studies is not a discipline featured by the compactness of terms (Marco, 2007), and the related publications are not always topically clear-cut (Huang and Xin, 2020). The interpretation of the term "translation assessment" by different scholars is considerably inconsistent, depending on their theoretical standpoints (Brunette, 2014). Therefore, "translation assessment" was a fuzzy concept and used in ways that were not always consistent. The very act of translation assessment sometimes is referred to other synonymous terms or related phrases, such as "translation evaluation" (Chapelle and Brindley, 2010; Xiao, 2012; Maier, 2014; Yang, 2019), "translation criticism" (Palumbo, 2009; Munday, 2016), "translation analysis" (McAlester, 1999) and "translation revision" (Arango-Keeth and Koby, 2003; Karoubi, 2016). However, some scholars point that there existing some differences among these terms, for example, the term evaluation refers to translation competence assessment, namely the process of translation evaluation for teaching purposes, while the term assessment applies to translation quality assessment, validating the suitability of the translation as a product to be presented to the customer (Arango-Keeth and Koby, 2003).

After consulting correlated literature and several translation experts, the retrieval strategy is set as TS = ("translation assessment" OR "translation evaluation" OR "translation criticism" OR "translation near/1 assessment" OR "translation near/1 evaluation*"), which takes the concentration and coverage of the literature into consideration. They are meaningful words that can represent the topic and cover the studies related to translation assessment to guarantee the retrieval accuracy. The literature sources were not streamlined in order to embrace the related interdisciplinary literature in this research. The datum was retrieved from the core collection of Web of Science (WoS) on March 28, 2023. The period is from 2000 to 2022, because the related studies before 2000 were limited and rare. The total number of publications is 1460.

1460 recorders were reviewed to select the representative and accurate ones as well as to remove the irrelevant and duplicate ones. Ultimately, 990 valid literature recorders were retained as data samples and saved in plain text format for bibliometric analysis. **Figure 1** shows the research design of this paper.

In stage 1, 990 literatures related to translation assessment were identified. Among which, the document types include: article, review and proceedings papers. "article" and "proceeding papers" are the main document types.

There are rich information involving in the exported records from WoS, for instance, authors, title, abstract, publication year, and references etc. Based on the adequate information from Stage 1, the bibliometric analysis and information visualization can be accomplished effectively in Stage 2.
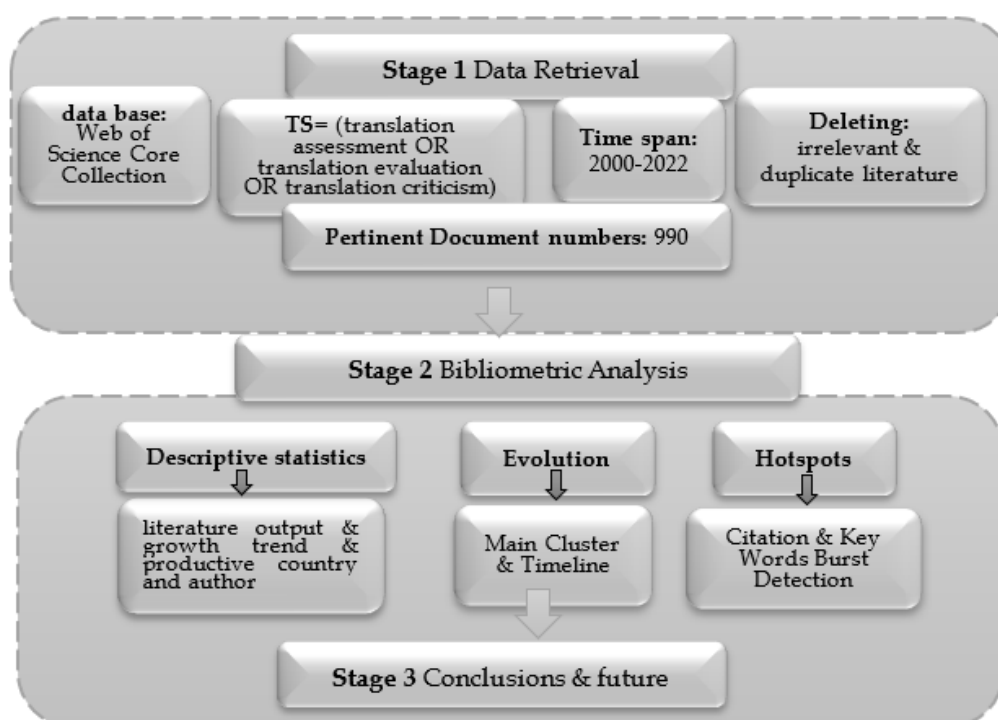
Figure 1. Research design.

The internationally widely-use bibliometric analysis software, Cite-Space, developed by Chaomei Chen was chosen in this research. It is a comprehensive research tool with which the maximum use can be made of the information contained in the literature for a systematic and time-line analysis of previous research, as well as use titles, abstracts and keywords to predict its future developments (Che et al., 2022). In this study, keywords co-occurrence, keywords and cited literature bursts were accomplished with Cite-Space.

Stage 3 comes with some conclusions and future work.

Therefore, this paper will explore translation assessment by BA to analyze papers taken from the Web of Science Core Collection. Research findings contain publication trend, cluster analysis, keyword co-occurrence (time zone) and burst detection to demonstrate the evolution and hotspots.

3. Findings and discussions

3.1. Descriptive analysis

Descriptive statistics can be utilized to indicate the overall trends of variables and carry out data mining (Pan et al., 2019). Counting the amount of academic literature in a discipline is one way to measure its degree of research development. By drawing and analysing the time distribution of the number of references, not only the research status of this discipline can be appraised efficiently, but also the development vitalities and trends would be expected (Diao et al., 2022). In brief, the publication number is an important indicator reflecting the development trends of scientific research (Guo et al., 2019).

The annual publication volume of journals from 2000 to 2022 was calculated (see **Figure 2**). On the whole, 990 papers published during the past 22 years, with an average annual publication of 45 papers. Before 2000, there are very few articles on translation assessment with intermittent publication year. Since 2000, related articles show an increasing trend. The Translator started a special issue on "Evaluation and Translation" devoting to the evaluation and quality of translation in 2000 (Maier, 2000), might have been partially responsible for the articles' appearance on translation assessment at this time. In its introduction, it says that conversations about the quality and value of translation have always been particularly bothering. Nowadays, however, the rising demand for translation in areas such as business and technology, requiring specialized translations as well as a more scientific approach to value determination. This is the reason why the special issue was set up.

According to the line graph, translation assessment research can be allotted into the following three stages: (1) From 2000 to 2007, as the initial period, a total of 88 papers were published, with an average annual number of 11 papers. The number of papers was a few and the growth was slow. During this period, there have been 2 pullbacks, almost all of them accompanied by rallies, the fastest one is in 2005. In 2005, papers related to machine translation appear. (2) During the stable period from 2008 to 2013, a total of 171 papers were published, with an average annual number of 19 papers. Although accompanied by a certain degree of fluctuation, but the overall trend is upward. In 2014, the number of papers reached a peak of 36 papers, which increased significantly. (3) From 2014 to 2022, for the fast developed period, a total of 726 papers were published, with an annual average of 81 papers, and translation assessment research tended to be stable. Although there was a decline in 2015, it grew rapidly in the following year and maintained a stable number, which indicates to some extent that translation assessment research has received continuous attention from a certain number of scholars and formed its own relatively independent research field.

Figure 2. Annual publication volume of TA (2000–2022).

3.2. Research themes

Cluster analysis refers to an exploratory data mining technique for identifying and analyzing the classification of significant terms and context within a certain research field (Si et al., 2019). The gathered data is transformed into several structured clusters using a series of algorithms, then gain a deeper understanding of

the research topics, such as the distribution and structures (Olawumi and Chan, 2018).

The foundation of cluster analysis relies on one of the most basic, simplest, and most often overlooked methods of understanding and learning, namely, grouping "objects" into "similar" groups, therefore, each subject in its group is more similar to other subjects than to subjects outside the group. This process involves many different algorithms and methods to make similar types of clusters.

In this paper, Cite Space 6.2R2 was used to cluster the keywords of 990 literatures, and LLR algorithm was used to identify 12 research clusters and the corresponding keywords of each cluster. The Log-Likelihood ratio (LLR), possessing the function of producing clusters with low inter-class similarity and high intra-class similarity, was the clustering technique. Thus, cluster analysis helps to classify large amounts of information into controllable units, and then deduces each cluster's information.

The past 20 years' data were obtained to analyze research themes. The time slice length was set at 2 years and the selection criteria g-index was $k = 15$. On the basis of LLR algorithm, cluster analysis shows 277 nodes, 266 connections, and 12 clusters. The modularity was Q = 0.8585, Weighted Mean Silhouette was S = 0.9863. The 12 research clusters and keywords are showed in **Table 1**, which are in order as follows: #0 translation competence, #1 translation quality, #2 statistical machine translation, #3 machine translation, #4 machine translation evaluation, #5 legal translation, #6 translation quality assessment, #7 translation evaluation, #8 neural works, #9 neural machine translation, #10 translator training, #11 translation strategies. Based on the clustering results, this paper integrates the above 12 research clusters with their related keywords, then summarizes 5 domains of translation assessment.

Table 1. Main clusters and keywords of TA (2000–2022).

No	Frequency	Centrality	Year	Label	Keywords
1	40	1	2014	Translation Competence	Translation Competence Acquisition; Competence Levels; Machine Translation; Trainee Translators
2	20	1	2015	translation quality	human evaluation; translation competence; mt evaluation; bologna process
3	20	0.992	2010	statistical machine translation	bidirectional language model; low-resource languages; distributional semantics; speech translation
4	19	0.99	2014	machine translation	translation competence; evaluation metrics; translation didactics; qualitative research
5	16	0.961	2016	machine translation evaluation	error analysis; natural language processing; college English; translation teaching
6	12	0.993	2017	legal translation	translator education; translation assessment; l2 translation; language and speech interfaces
7	11	0.989	2011	translation quality assessment	comparable corpora; corpus-based translation studies; literary translation; machine translation
8	11	0.971	2007	translation evaluation	automatic evaluation; translation errors; measuring instruments; translation test
9	10	0.959	2019	neural networks	attention mechanism; accuracy; correlation; neural mt

Table 1. (*Continued*).

No	Frequency	Centrality	Year	Label	Keywords
10	8	1	2017	neural machine translation	quality estimation; machine learning; transformer; phrase-based statistical machine translation
11	6	0.968	2011	translator training	English for translation and interpreting; framework of reference; academic performance; bilingual sub-competence
12	6	0.942	2016	translation strategies	tourism text; culture-specific elements; translation techniques; Skopos theory

There existing some duplicate and similar contents among the twelve clusters, as seen in **Table 1**, for example "statistical machine translation" "neural machine translation" and "machine translation evaluation", which are related to the cluster "machine translation", therefore, the four clusters are combined into one. Similarly, cluster "translation quality" and cluster "translation quality assessment" both focus on quality, thus they are synthesized into one cluster "translation quality". The "translation evaluation" cluster, as the research topic in this study, is eliminated. Thus, this paper, on the basis of cluster analysis results and through further study and analysis of literature, merges the above 12 clusters and keywords to derive five research themes, i.e., translation competence, translation quality, machine translation, translation teaching and training, and others, as shown in **Figure 3**.
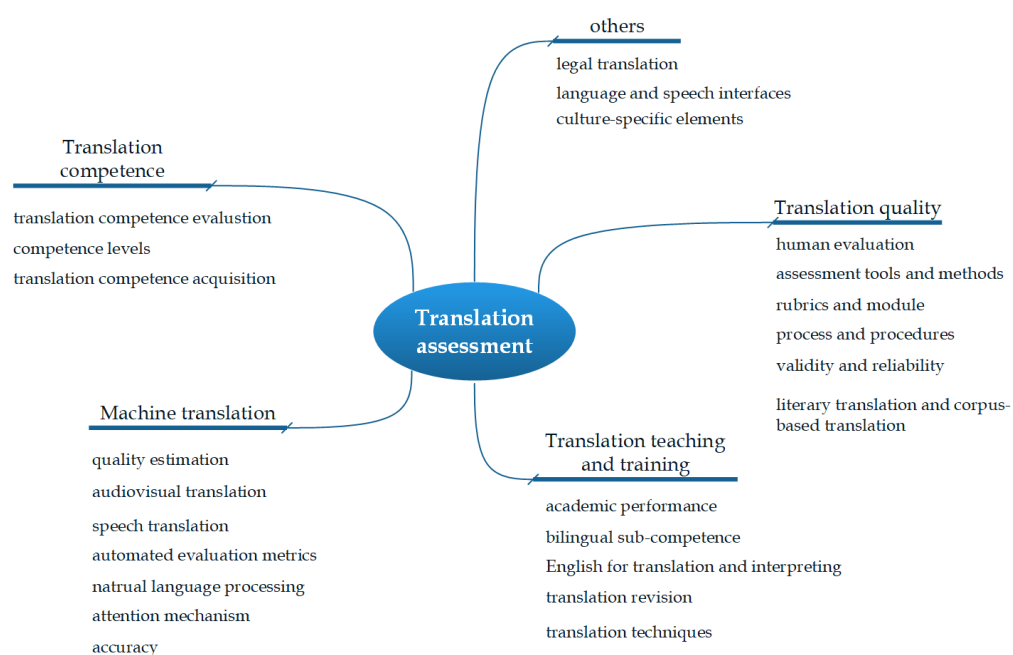

Figure 3. Research theme of translation assessment (2000–2022).

The first main cluster is translation competence, in which translator's competence evaluation, translation competence acquisition and competence levels are included. In this cluster, scholars mainly focused on the construct of translation competence, and translation competence assessment of the translators.

The second main cluster is translation quality. It includes rich contents serving for assessing translation quality, such as assessment tools and methods, rubrics and module, process and procedures, validity and reliability, literary translation and

corpus-based translation, comparison between human translation and machine translation.

The third main cluster is machine translation, including key words such as quality estimation, audiovisual translation, speech translation, automated evaluation metrics, natural language processing, attention mechanism.

The fourth main cluster is translation teaching and training. This cluster focuses on assessment in education, such as academic performance, bilingual sub-competence, English for translation and interpreting, translation revision, translation techniques etc. In addition, how to assess and improve students' translation skills in teaching process is frequently discussed in this cluster.

The fifth main cluster is others, which includes legal translation, language and speech interfaces, culture-specific elements etc.

3.3. Research evolutions

There is a growing interest in tracking research topics. One of the methods is to visualize various time patterns and detect the trend on this basis (Erten et al., 2003). The typical words to depict the central content of a paper are key words, and the network analysis of keyword co-occurrence is helpful to detect the research concentration in a certain filed (Diao et al., 2022). This paper visualizes the keyword co-occurrence network with respect to time zone, as shown in **Figure 4**. The time span is from 2000 till now, which shows the whole process of the evolution. The search term of translation assessment come from the title, abstract and key words. The size of a node implies the frequency of keyword occurrences. The larger the node, the higher the frequency. In addition, keywords near the center of each nod represent the core topics for that time period.

Since 2000, key words occurred intensively. From 2000 to 2005, The main topics studied by scholars are translation quality and its assessment, translation competence, translation assessment, revision and translation error. In other words, the studies of translation assessment begin with translation quality assessment and translation competence assessment. Some scholars develop models for translation quality assessment (al-Qinai, 2000; House, 2001; Williams, 2001; etc.) and revisions for translation quality control (Chakhachiro, 2005), others come up with different approaches to translation evaluation (Waddington, 2001; Bowker, 2001; Li, 2001; etc.). The construct, acquirement, assessment of translation competence is also explored (Martínez and Hurtado, 2001; Orozco and Hurtado Albir, 2002; Pym, 2003; etc.).

From 2005 to 2012, scholar's research themes gradually evolved into machine translation evaluation. Scholars concentrate on the evaluation metrics of machine translation and how to improve it (Yao et al., 2006; Zhu and Wang, 2006; Yang et al., 2008; etc.). Within machine translation theme, the statistical machine translation system was attached attention to Sadat and Habash (2006); Hassan et al. (2006); Schwenk and Estève (2008). During this period, although the research theme still covered translation assessment and translation evaluation, it was less compared with machine translation (Li, 2006; Albir, 2007; Colina, 2008; etc.).
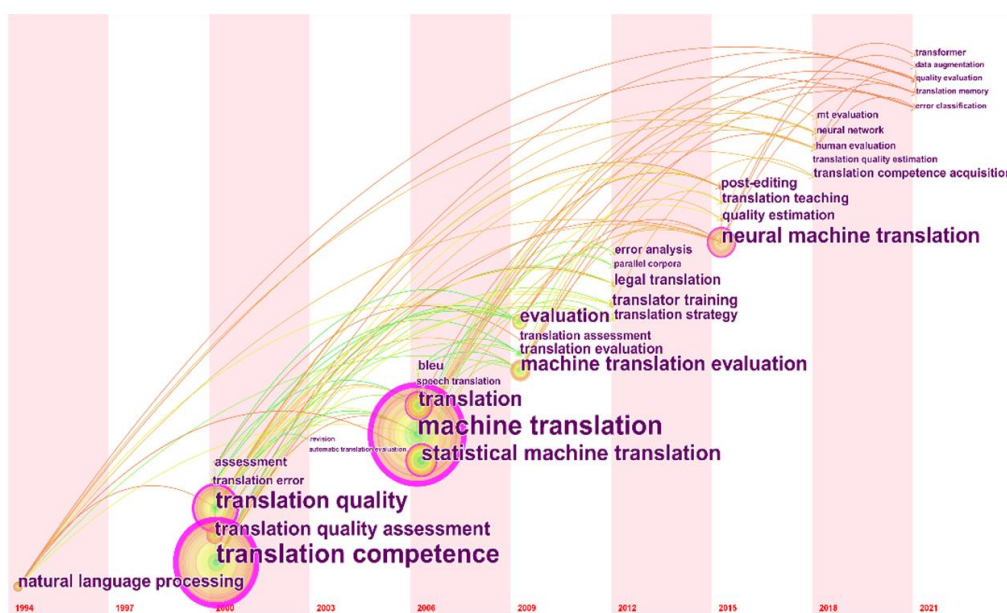
Figure 4. Translation assessment research evolution (2000–2022).

From 2012 to 2015, the research topic began to center around translation assessment in teaching and training. The translation competence acquisition and evaluation of EFL learners or translators were studied frequently (Göpferich, 2013; Pym, 2013; Károly, 2014; etc.). In addition, error patterns were identified and corpus were constructed to help assess the translations (Popescu, 2013; Lee and Ronowicz, 2014).

New research theme revolving around neural machine translation evaluation emerged from 2015 to 2018 (Guzmán et al., 2017; Moorkens, 2018). With the rise of this topic, scholars also began to pay attention to the quality of machine translation (Burchardt et al., 2016; Castilho and O'Brien, 2017). Translation quality improvement through edition, revision and other methods are also studied during this period (Daems, et al., 2017; De Sutter et al., 2017; Liu et al., 2017; etc.).

Since 2018, not centering on the certain topic, the research themes began to be all-sided, which included translation competence acquisition (Cortez-Godínez, 2019; PACTE, 2019; Şerban and Muresan, 2019; etc.), translation quality estimation (Jani et al., 2019; Korol, 2020; Han, 2020; etc.), machine translation evaluation (Comelles and Atserias, 2019; Liu and Xiao, 2019; Chatzikoumi, 2019; etc.), human evaluation (Fomicheva and Specia, 2019), error classification (Carl and Báez, 2019; Wang and Wang, 2020; Du and Saeheaw, 2020).

To sum up, the research evolution of translation assessment from 2000 to 2022 can be summarized as below: (1) the rise in the number of participants. At first, scholars studying translation assessment mainly come from the filed of translation and linguistics. Later on, scholars possessing different research backgrounds, such as psychology, testing, pedagogy, computer science join to study on this topic. (2) The interdisciplinary nature of the research content is becoming increasingly apparent. Initially, translation assessment principally evaluating competence and quality, the former is closely related to pedagogy field, for instance, error analysis, post-editing, translator training; while the latter is closely linked to literature and linguistic filed.

With the emergence of machine translation, its assessment is closely connected to the fields of science and technology, such as computer, software and various algorithms, presenting a stronger interdisciplinary feature. (3) the development of research angle: in the early stage of machine translation, scholars concentrate on the evaluation metrics and its reliability. With the deepening of the research on machine translation, scholars shifted to the quality of machine translation and how to improve it. (4) For research methods, usually scholars utilize controlled experiments to compare the translation quality or translation competence. Since 2012, case study and survey are more popular. (5) Half research about machine translation during before 2018 came from collected papers of international conference. This is probably because research on machine translation lies in the early stage and is the research hotspot at that time. Since 2018, research in this filed came from journals, indicating that translation assessment has received sustained attention from a certain number of scholars.

3.4. New hot spots

In order to explore the research hot spots, the keywords with high frequency as well as their centrality are firstly made statistics. According to the **Table 2** below, translation competence, machine translation and translation quality are the top three keywords with high frequency and high centrality.

Table 2. High frequency keywords and centrality.

No.	Frequency	Centrality	Keywords	No.	Frequency	Centrality	Keywords
1	50	0.42	Translation competence	7	11	0.08	Translator training
2	50	0.34	Machine translation	8	11	0.08	Machine translation evaluation
3	31	0.21	Translation quality	9	10	0.08	Legal translation
4	23	0.17	Neural machine translation	10	8	0.03	Statistical machine translation
5	13	0.14	Translation quality assessment	11	6	0.03	Competence

In order to further understand the research hotspots of translation assessment, the burst detection of keywords and citations is carried out with the help of Citespace.

Keywords burst and cited literature burst provide evidence that specific keywords and citations are associated with spikes in occurrence frequency. Keywords burst appears when a topic has attracted or is attracting attention at a certain time from researchers, and cited literature burst indicates that the research circle has paid or is paying particular attention to the possibility of the paper's potential contribution (Diao et al., 2022). Therefore, burst detection, regarded as a sign of a highly active field of research, able to discover emerging trends as well as fleeting ones (Pollack and Adler, 2015). To explore the emerging hot spots in translation assessment, this paper conducted sudden detection of keywords and citations in the literature published in the past eight years, and extracted the detection results in the past five years based on the burst intensity, as shown in **Table 3**.

Table 3. Burst detection of translation assessment (2018–2022).

Keywords	Strength	Year	Citations	Strength	Year
Neural mt	**1.19**	2018–2018	Edinburgh neural machine translation systems for WMT 16	**1.78**	2017–2017
Translation quality	**1.09**	2018–2018	Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics	**2**	2017–2019
Translation competence	**1.91**	2019–2019	First results of PACTE group's experimental research on translation competence acquisition: the acquisition of declarative knowledge of translation	**2.09**	2019–2019
Student translator	**1.57**	2019–2019	English language teaching for translator and interpreter trainees: Syllabus analysis and design	**2.09**	2019–2019
Statistical machine translation	**1.41**	2019–2019	Findings of the 2016 Conference on Machine Translation (WMT16)	**2.18**	2018–2019
Translation competence acquisition	**2.28**	2019–2020	Results of PACTE's Experimental Research on the Acquisition of Translation Competence: the Acquisition of Declarative and Procedural Knowledge in Translation. The Dynamic Translation Index	**3.04**	2018–2020
Translation quality estimation	**1.1**	2020–2020	Predictor-estimator using multilevel task learning with stack propagation for neural quality estimation	**1.55**	2019–2020
Neural machine translation	**2.22**	2020–2022	Open NMT: Open-Source Toolkit for Neural Machine Translation	**1.86**	2020–2022
Translator competence	**1.06**	2021–2022	Occasioning translator competence: Moving beyond social constructivism toward a postmodern alternative to instructionism	**2.33**	2019–2020
Neural network	**1.06**	2021–2022	Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates	**2.03**	2021–2022

According to the keyword burst detection results, the burst strength of translation competence acquisition and neural machine translation is the highest, with 2.28 and 2.22 respectively. The former shows that assessing the acquisition of one's translation competence is a focus in research field in recent years, while the latter shows that in machine translation evaluation system, neural machine translation has surpassed statistical machine translation and attracted increasing attention by researchers. other keywords like neural mt and translation quality burst since 2018, demonstrating researchers concerned about the evaluation of translation quality; translation competence, student translator and statistical machine translation began to burst in 2019, showing students' translation competence is a hotspot at that time and statistical machine translation still being explored by scholars. However, the burst duration time of the above key words last only one year. Since 2020, the burst duration time of keywords last more than one year, like neural machine translation, translator competence and neural network, indicating these areas received more sustained attention.

In terms of citations burst detection, the identified citations center on the machine translation field, signifying that TA is not only explored by linguistic and language field, but also by computer science filed. Classic papers about translation competence started a sustained burst since 2018, with the highest strength. Next is machine translation, started to burst in 2017. Among the 10 high-cited articles, 7 belong to empirical research in which experiment or case study is conducted to obtain related data, demonstrating empirical study is the mainstream in TS field in

recent years. In addition, one article about machine translation comes from conference proceedings, which indicates the hotspot in this filed is discussed in conferences. By translation competence acquisition experiment, Beeby et al. (2015) found that after training, the professional and expert translators have a less obvious concept of and approach to translation compared with students due to the former prefers a static concept of translation while the latter has obtained more dynamic educational input. Kudo (2018) showed by experiment that using segmentation ambiguity as noise to enhance the accuracy and robustness of neural machine translation is feasible. Due to the highly-fragmented view on translation quality and the fundamentally-different quality evaluation methods between human and machine translation, there are some problems existing in requesters and users. In order to solve it, Lommel et al. (2014) developed the Multidimensional Quality Metrics (MQM) framework to declare and describe translation quality metrics through a shared vocabulary of "issue types". PACTE (2016) presents the results of the 2016 Machine Translation Conference, in which five machine translation tasks, three evaluation tasks, an automatic post-editing task and bilingual document alignment task are included.

The results of comparative analysis of keywords and citation burst show that "machine translation, translation competence, translation quality" are the common high-frequency topics in the two detections. Machine translation refers to the study of how to translate using computers, and it has long been considered one of the most challenging assignments in the natural language processing field (NLP) (Wang et al., 2022). The rise and popularity of machine translation triggered a wide-ranging exploration of translation quality and translation competence evaluation by scholars, since whether machine translation is as good as human translation is an interesting topic. It can be seen that the emerging hot topics recently in translation assessment field are mainly related to translation quality estimation and translation competence acquisition.

4. Conclusions

This study, taking valid papers on translation assessment published in WoS core collection from 2000 to 2022 as research samples, visualizes and reviews the research theme, research evolutions and emerging hot topics of translation assessment through bibliometric analysis, proposes implications for further translation assessment research and provides references for scholars in this field.

This paper shows the outcome of the BA of research papers on TA in the WoS core collection. By combing 990 literatures related to TA, the research themes, evolutions and hotspots are summarized. The research findings include:(1) Research on TA come across three stages of development: preliminary stage (2000–2007), growth stage (2008–2013), and rapid stage (2014–2022). The growth trend of this research field showed that translation assessment has received continuous attention from a certain number of scholars and formed its own relatively independent research field. (2) Research themes from 2000 to 2022 can be categorized into five themes, including translation competence, translation quality, machine translation, translation teaching and training, and others. (3) Research evolution of TA from

2000 to 2021 can be presented in several aspects: the rise of participants number, the interdisciplinary nature of the research content, the development of research angle, as well as research methods and publish journals. (4) The burst detections investigate the newly developing part in this field, which include: translation competence acquisition. translation quality estimation and machine translation evaluation.

It is undeniable that there are still some limitations in the paper. On the one hand, WoS is a representative and authoritative database, however, the same keywords are retrieved in other database, the number of articles are quite different. Therefore, optimizing translation assessment research based on database is the direction for future research. On the other hand, translation assessment is a topic involving many disciplines, like linguistics, education, computer science etc. Limited by the words number and interdisciplinary integration ability, the paper does not carry out a detailed and in-depth analysis from the aspect of interdisciplinary fusion. This will be the direction of further improvement in the follow-up research.

Author contributions: Writing original draft, DT; revising and making improvemnts, LSE. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

Adom, D., Adu-Mensah, J., & Dake, D. A. (2020). Test, measurement, and evaluation: Understanding and use of the concepts in education. International Journal of Evaluation and Research in Education (IJERE), 9(1), 109. https://doi.org/10.11591/ijere.v9i1.20457

Airasian, P.W. (1997). Classroom assessment, 3rd ed. McGraw-Hill.

Akulina, K. V., & Tikhonova, E. V. (2021). Application of the translation quality assessment system in the training of linguist-translators. Yazyk i Kul'tura, 56, 129–145. https://doi.org/10.17223/19996195/56/8

Albir, A. H. (2007). Competence-based Curriculum Design for Training Translators. The Interpreter and Translator Trainer, 1(2), 163–195. https://doi.org/10.1080/1750399x.2007.10798757

Albir, P. Group. A. H., Galán-Mañas, A., Kuznik, A., Olalla-Soler, C., Rodríguez-Inés, P., & Romero, L. (2020). Translation competence acquisition. Design and results of the PACTE group's experimental research. The Interpreter and Translator Trainer, 14(2), 95–233. https://doi.org/10.1080/1750399x.2020.1732601

Alos, J. (2015). Explicating the implicit: an empirical investigation into pragmatic competence in translator training. The Interpreter and Translator Trainer, 9(3), 287–305. https://doi.org/10.1080/1750399x.2015.1100398

al-Qinai, J. (2002). Translation Quality Assessment. Strategies, Parametres and Procedures. Meta, 45(3), 497–519. https://doi.org/10.7202/001878ar

Angelone, E. (2020). The impact of screen recording as a diagnostic process protocol on inter-rater consistency in translation assessment Erik Angelone, Kent State University. Journal of Specialised Translation, 34, 32-50.

Angelone, E. (2021). Broadening the scope of error categories in translation assessment through screen recording. Across Languages and Cultures, 22(2), 143–157. https://doi.org/10.1556/084.2021.00021

Arango-Keeth, F., & Koby, G. S. (2003). Assessing assessment. Beyond the Ivory Tower, 117–134. https://doi.org/10.1075/ata.xii.09ara

Bai, L., Feng, L., & Yan, M. (2018). Constructs and principles of the Chinese English Translation Proficiency Rating Scale. Modem Foreign Language, 41(1), 101-110.

Baniata, L. H., Park, S., & Park, S.-B. (2018). A Multitask-Based Neural Machine Translation Model with Part-of-Speech Tags Integration for Arabic Dialects. Applied Sciences, 8(12), 2502. https://doi.org/10.3390/app8122502

Beeby, A., Castillo, L., Fox, O., Galán Mañas, A., Hurtado Albir, A., Kuznik, A., Massana, G., Neunzig, W., Olalla, Ch., Rodríguez Inés, P., & Romero, L. (2015). Results of PACTE's Experimental Research on the Acquisition of Translation Competence. Translation Spaces, 4(1), 29–53. Portico. https://doi.org/10.1075/ts.4.1.02bee

Bowker, L. (2002). Towards a Methodology for a Corpus-Based Approach to Translation Evaluation. Meta, 46(2), 345–364. https://doi.org/10.7202/002135ar

Brunette, L. (2000). Towards a Terminology for Translation Quality Assessment. The Translator, 6(2), 169–182. https://doi.org/10.1080/13556509.2000.10799064

Brunette, L. (2000). Towards a Terminology for Translation Quality Assessment. The Translator, 6(2), 169–182. https://doi.org/10.1080/13556509.2000.10799064

Burchardt, A., Lommel, A., Bywood, L., Harris, K., & Popović, M. (2016). Machine translation quality in an audiovisual context. Target, 28(2), 206–221. Portico. https://doi.org/10.1075/target.28.2.03bur

Cao, D. (2014). Teaching and learning legal translation. Semiotica, 2014(201). https://doi.org/10.1515/sem-2014-0022

Carl, M., & Báez, M. C. T. (2019). Machine translation errors and the translation process: A study across different languages. Journal of Specialised Translation, 31, 107-132.

Castilho, S., & O'Brien, S. (2018). Acceptability of machine-translated content: A multi-language evaluation by translators and end-users. Linguistica Antverpiensia, New Series – Themes in Translation Studies, 16. https://doi.org/10.52034/lanstts.v16i0.430

Chakhachiro, R. (2005). Revision for quality. Perspectives, 13(3), 255–238. https://doi.org/10.1080/09076760508668994

Chapelle, C. A., & Brindley, G. (2010). What is language assessment. 247-267.

Chatzikoumi, E. (2019). How to evaluate machine translation: A review of automated and human metrics. Natural Language Engineering, 26(2), 137–161. https://doi.org/10.1017/s1351324919000469

Che, S., Kamphuis, P., Zhang, S., Zhao, X., & Kim, J. H. (2022). A Visualization Analysis of Crisis and Risk Communication Research Using CiteSpace. International Journal of Environmental Research and Public Health, 19(5), 2923. https://doi.org/10.3390/ijerph19052923

Colina, S. (2008). Translation Quality Evaluation. The Translator, 14(1), 97–134. https://doi.org/10.1080/13556509.2008.10799251

Comelles, E., & Atserias, J. (2018). VERTa: a linguistic approach to automatic machine translation evaluation. Language Resources and Evaluation, 53(1), 57–86. https://doi.org/10.1007/s10579-018-9430-2

Cortez-Godínez, J. (2019). Influencia del nivel socioeconómico en la adquisición de la Competencia Traductora. Sendebar, 30(0). https://doi.org/10.30827/sendebar.v30i0.6594

Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017). Identifying the Machine Translation Error Types with the Greatest Impact on Post-editing Effort. Frontiers in Psychology, 8. https://doi.org/10.3389/fpsyg.2017.01282

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. https://doi.org/10.52034/lanstts.v16i0.440

Diao, B.C., Adnan, N., Harji, M. B., Ravindran, L. (2022). Evolution and hotspots of peer instruction: a visualized analysis using CiteSpace. Education and Information Technologies. 1-18.

Ding, Z.Y., Zang, H.F., Zhang, Q., Miao, G.M. & Chi, Y.H. Automatic machine translation evaluation based on sentence structure information. Paper presented at the 2009 International Conference on Asian Language Processing.

Du, W., & Saeheaw, T. (2020). Errors in translation: A case study of Chinese into English translation. Language Learning in Higher Education, 10(1), 129–142. https://doi.org/10.1515/cercles-2020-2009

Dugonik, J., Bošković, B., Brest, J., & Sepesy Maučec, M. (2019). Improving Statistical Machine Translation Quality Using Differential Evolution. Informatica, 30(4), 629–645. https://doi.org/10.15388/informatica.2019.222

Dussart, A. (2005). Faux sens, contresens, non-sens… un faux débat ? Meta, 50(1), 107–119. https://doi.org/10.7202/010661ar

Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? Scientometrics, 105(3), 1809–1831. https://doi.org/10.1007/s11192-015-1645-z

Erten, C., Harding, P. J., Kobourov, S. G., Wampler, K., & Yee, G. (2004). <title>Exploring the computing literature using temporal graph visualization</title> SPIE Proceedings. https://doi.org/10.1117/12.539245

Finch, A., Yasuda, K., Okuma, H., Sumita, E., Nakamura, S. (2011). A Bayesian Model of Transliteration and Its Human Evaluation When Integrated into a Machine Translation System. IEICE Transactions on Information and Systems, E94-D(10), 1889–1900. https://doi.org/10.1587/transinf.e94.d.1889

Fomicheva, M., & Specia, L. (2019). Taking MT Evaluation Metrics to Extremes: Beyond Correlation with Human Judgments. Computational Linguistics, 45(3), 515–558. https://doi.org/10.1162/coli_a_00356

Gambier, Y., & van Doorslaer, L. (Eds.). (2007). The Metalanguage of Translation. Target, 19(2). Portico. https://doi.org/10.1075/target.19.2

Garcia, I. (2010). Is machine translation ready yet? Target, 22(1), 7–21. Portico. https://doi.org/10.1075/target.22.1.02gar

Gile, D. (2000). The History of Research into Conference Interpreting. Target, 12(2), 297–321. Portico. https://doi.org/10.1075/target.12.2.07gil

Göpferich, S. (2013). Translation competence. Target, 25(1), 61–76. Portico. https://doi.org/10.1075/target.25.1.06goe

Guo, Y.-M., Huang, Z.-L., Guo, J., Li, H., Guo, X.-R., & Nkeli, M. J. (2019). Bibliometric Analysis on Smart Cities Research. Sustainability, 11(13), 3606. https://doi.org/10.3390/su11133606

Guzmán, F., Joty, S., Màrquez, L., & Nakov, P. (2017). Machine translation evaluation with neural networks. Computer Speech & Language, 45, 180–200. https://doi.org/10.1016/j.csl.2016.12.005

Hague, D. R. (2015). Empathic accuracy and translator training: helping students imagine other minds. The Interpreter and Translator Trainer, 9(1), 20–29. https://doi.org/10.1080/1750399x.2015.1010359

Hague, D., Melby, A., & Zheng, W. (2011). Surveying Translation Quality Assessment. The Interpreter and Translator Trainer, 5(2), 243–267. https://doi.org/10.1080/13556509.2011.10798820

Hajmohammadi, A. (2005). TRANSLATION EVALUATION IN A NEWS AGENCY. Perspectives, 13(3), 215–224. https://doi.org/10.1080/09076760508668993

Han, C. (2020). Translation quality assessment: a critical methodological review. The Translator, 26(3), 257–273. https://doi.org/10.1080/13556509.2020.1834751

Han, C., Hu, B., Fan, Q., Duan, J., & Li, X. (2022). Using computerised comparative judgement to assess translation. Across Languages and Cultures, 23(1), 56–74. https://doi.org/10.1556/084.2022.00001

Han, Y., & Meng, S. (2022). Machine English Translation Evaluation System Based on BP Neural Network Algorithm. Computational Intelligence and Neuroscience, 2022, 1–10. https://doi.org/10.1155/2022/4974579

Hassan, H., Hearne, M., Way, A., & Sima'an, K. (2006). Syntactic phrase-based statistical machine translation. Paper presented at the 2006 IEEE Spoken Language Technology Workshop. https://doi.org/10.1109/slt.2006.326799

House, J. (2002). Translation Quality Assessment: Linguistic Description versus Social Evaluation. Meta, 46(2), 243–257. https://doi.org/10.7202/003141ar

Huang, Q., & Xin, X. (2020). A bibliometric analysis of translation criticism studies and its implications. Perspectives, 28(5), 737–755. https://doi.org/10.1080/0907676x.2020.1740750

Janmaijaya, M., Shukla, A., Abraham, A., & Muhuri, P. (2018). A Scientometric Study of Neurocomputing Publications (1992–2018): An Aerial Overview of Intrinsic Structure. Publications, 6(3), 32. https://doi.org/10.3390/publications6030032

Jiang, C. (2010). Quality assessment for the translation of museum texts: application of a systemic functional model. Perspectives, 18(2), 109–126. https://doi.org/10.1080/09076761003678734

Károly, A. (2014). Translation in foreign language teaching: A case study from a functional perspective. Linguistics and Education, 25, 90–107. https://doi.org/10.1016/j.linged.2013.09.009

Karoubi, B. (2016). Translation quality assessment demystified. Babel, 62(2), 253–277. Portico. https://doi.org/10.1075/babel.62.2.05kar

Kasperavičienė, R., & Horbačauskienė, J. (2020). Self-revision and other-revision as part of translation competence in translator training. Journal of Language and Cultural Education, 8(1), 117–133. https://doi.org/10.2478/jolace-2020-0007

Ke, X.H. (2009). An Automatic Translation Evaluation System Based on Semantic Similarities and Fuzzy Neartude. Paper presented at the 2009 International Conference on Environmental Science and Information Application Technology.

Kiraly, D. (2000). A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice. St Jerome Publishing.

Kit, C., & Wong, T. M. (2008). Comparative evaluation of online machine translation systems with legal texts. Law Library Journal, 100(2), 299-321.

Korol, T. (2020). Translation assessment: Is there anything to be tested objectively? Advanced Education, 15, 32-40. https://doi.org/10.20535/2410-8286.200720

Korol, T. (2021). Digital teacher feedback as a translation assessment tool. Journal of Teaching English for Specific and Academic Purposes, 575. https://doi.org/10.22190/jtesap2104575k

Kudo, T. (2018). Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). https://doi.org/10.18653/v1/p18-1007

Lee, Y. O., & Ronowicz, E. (2014). The development of an error typology to assess translation from English into Korean in class. Babel, 60(1), 35–51. Portico. https://doi.org/10.1075/babel.60.1.03lee

Li, D. (2006). Making Translation Testing More Teaching-oriented: A Case Study of Translation Testing in China. Meta, 51(1), 72–88. https://doi.org/10.7202/012994ar

Liu, L., Fujita, A., Utiyama, M., Finch, A., & Sumita, E. (2017). Translation Quality Estimation Using Only Bilingual Corpora. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(9), 1762–1772. https://doi.org/10.1109/taslp.2017.2716195

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. Tradumàtica Tecnologies de La Traducció, 12, 455–463. https://doi.org/10.5565/rev/tradumatica.77

Maier, C. (2000). Introduction. The Translator, 6(2), 137–148. https://doi.org/10.1080/13556509.2000.10799062

Maier, C. (2014). Evaluation and Translation. Routledge. https://doi.org/10.4324/9781315759999

Martínez Melis, N., & Hurtado Albir, A. (2002). Assessment In Translation Studies: Research Needs. Meta, 46(2), 272–287. https://doi.org/10.7202/003624ar

McAlester, G. (1999). "The Source Text in Translation Assessment". In Words, Words, Words. The Translator and the Language Learner. In: Anderman G, Rogers, M. (editors). pp. 169-178. Clevedon, Philadelphia and Adelaide: Multilingual Matters.

Mei, H., & Chen, H. (2022). Assessing Students' Translation Competence: Integrating China's Standards of English With Cognitive Diagnostic Assessment Approaches. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.872025

Mellinger, C. D. (2018). Re-thinking translation quality. Target, 310–331. https://doi.org/10.1075/target.16104.mel

Moorkens, J. (2018). What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise. The Interpreter and Translator Trainer, 12(4), 375–387. https://doi.org/10.1080/1750399x.2018.1501639

Munday, J. (2016). Introducing Translation Studies. Routledge. https://doi.org/10.4324/9781315691862

Nida, E., & Taber, C. (2003). The Theory and Practice of Translation. https://doi.org/10.1163/9789004496330

Nikolaeva, S., & Korol, T. (2021). Prospective Philologists' translation assessment triangulation: screen video recording and think aloud protocol combination. Advanced Education, 8(18), 30–41. LOCKSS. https://doi.org/10.20535/2410-8286.228550

Olawumi, T. O., & Chan, D. W. M. (2018). A scientometric review of global research on sustainability and sustainable development. Journal of Cleaner Production, 183, 231–250. https://doi.org/10.1016/j.jclepro.2018.02.162

Orozco, M., & Hurtado Albir, A. (2004). Measuring Translation Competence Acquisition. Meta, 47(3), 375–402. https://doi.org/10.7202/008022ar

PACET (2019). Evolution of the Efficacy of the Translation Process in Translation Competence Acquisition. Meta: Journal Des Traducteurs, 64(1), 242. https://doi.org/10.7202/1065336ar

PACTE (2005). Investigating translation competence: conceptual and methodological issues. Volet Traduction, 50(2), 609–619. https://doi.org/10.7202/011004ar

Palumbo, G. (2009). Key terms in translation studies: Bloomsbury Publishing.

Pan, W., Jian, L., & Liu, T. (2019). Grey system theory trends from 1991 to 2018: a bibliometric analysis and visualization. Scientometrics, 121(3), 1407–1434. https://doi.org/10.1007/s11192-019-03256-z

Pollack, J., & Adler, D. (2015). Emergent trends and passing fads in project management research: A scientometric analysis of changes in the field. International Journal of Project Management, 33(1), 236–248. https://doi.org/10.1016/j.ijproman.2014.04.011

Popescu, T. (2013). A Corpus-based Approach to Translation Error Analysis. a Case-study of Romanian EFL Learners. Procedia - Social and Behavioral Sciences, 83, 242–247. https://doi.org/10.1016/j.sbspro.2013.06.048

Pranckutė, R. (2021). Web of Science (WoS) and Scopus: The Titans of Bibliographic Information in Today's Academic World. Publications, 9(1), 12. https://doi.org/10.3390/publications9010012

Pym, A. (2004). Redefining Translation Competence in an Electronic Age. In Defence of a Minimalist Approach. Meta, 48(4), 481–497. https://doi.org/10.7202/008533ar

Pym, A. (2013). Research skills in translation studies: What we need training in. Across Languages and Cultures, 14(1), 1–14. https://doi.org/10.1556/acr.14.2013.1.1

Sadat, F., & Habash, N. (2006). Combination of Arabic preprocessing schemes for statistical machine translation. Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the ACL - ACL '06. https://doi.org/10.3115/1220175.1220176

Schwenk, H., & Esteve, Y. (2008). Data selection and smoothing in an open-source system for the 2008 NIST machine translation evaluation. Interspeech 2008. https://doi.org/10.21437/interspeech.2008-676

Şerban, A., & Mureşan, V. (2019). Register and Translation Competence Acquisition. Investigating Student Attitudes. Studia Universitatis Babeş-Bolyai Philologia, 64(4), 213–230. https://doi.org/10.24193/subbphilo.2019.4.13

Si, Y.H., Shi, J.G., Chen, J.D., Wu, G.D., & Wang, H.M., (2019). Reviewing sustainable development research in China from the perspective of China Population, Resources and Environment. China population, resources and environment, 29(7), 166-176.

Su, W. (2021). Understanding rubric use in peer assessment of translation. Perspectives, 30(1), 71–85. https://doi.org/10.1080/0907676x.2020.1862260

Tingting, L., & Mengyu, X. (2019). Analysis and evaluation on the quality of news text machine translation based on neural network. Multimedia Tools and Applications, 79(23–24), 17015–17026. https://doi.org/10.1007/s11042-019-7532-5

Tymoczko, M. (2006). Trajectories of Research in Translation Studies. Séances Plénières, 50(4), 1082–1097. https://doi.org/10.7202/012062ar

Waddington, C. (2002). Different Methods of Evaluating Student Translations: The Question of Validity. Meta, 46(2), 311–325. https://doi.org/10.7202/004583ar

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in Machine Translation. Engineering, 18, 143–153. https://doi.org/10.1016/j.eng.2021.03.023

Wang, L., & Wang, X. (2020). How to evaluate literary translations in the classroom context: through error analysis or scale-based method? Current Trends in Translation Teaching and Learning E, 7, 276-313.

Williams, M. (2002). The Application of Argumentation Theory to Translation Quality Assessment. Meta, 46(2), 326–344. https://doi.org/10.7202/004605ar

Xiao, W.Q. (2012). Study on the Undergraduate Translation Major Testing. People's Publishing House.

Yang, M., Zhu, J., Li, J., Wang, L., Qi, H., Li, S., & Daxin, L. (2008). Extending BLEU Evaluation Method with Linguistic Weight. 2008 The 9th International Conference for Young Computer Scientists. https://doi.org/10.1109/icycs.2008.362

Yang, M.-Y., Sun, S.-Q., Zhu, J.-G., Li, S., Zhao, T.-J., & Zhu, X.-N. (2011). Improvement of Machine Translation Evaluation by Simple Linguistically Motivated Features. Journal of Computer Science and Technology, 26(1), 57–67. https://doi.org/10.1007/s11390-011-9415-8

Yang, Z. H. (2019). Studies on Translation Testing and Assessment. Foreign Language Teaching and Research Press.

Yao, J.M., Qu, Y.Q., Zhu, Q.M., & Zhang, J. (2006). A Visualization method for machine translation evaluation results. Proceedings (Stymne et al., 2012) of the 20th Pacific Asia Conference on Language, Information and Computation.

Yilmaz, I. (2019). Bibliometric Analysis of Bibliometric Studies on Tourism Published in Turkey. Anais Brasileiros de Estudos Turísticos - ABET, 9(1, 2 e 3). https://doi.org/10.34019/2238-2925.2019.v9.27111

Yunxing, L. (2001). Translation evaluation in a systemic perspective. Perspectives, 9(1), 15–21. https://doi.org/10.1080/0907676x.2001.9961403

Zhu, J., & Wang, H. (2006). The effect of translation quality in MT-based cross-language information retrieval. Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the ACL - ACL '06. https://doi.org/10.3115/1220175.1220250

Zlatnar Moe, M., Mikolic Juznic, T., & Žigon, T. (2015). I know languages, therefore, I can translate? Translation and Interpreting Studies, 10(1), 87–107. Portico. https://doi.org/10.1075/tis.10.1.05moe