

Bibliometrix: Science Mapping Analysis with R Biblioshiny Based on Web of Science in Applied Linguistics



Babak Daneshvar Ghorbani 

1 Introduction

Today, more scientific information is being made than in the past, and keeping up with everything published is getting more complicated. Furthermore, it is increasingly referred to as multidisciplinary science (Linnenluecke et al. 2020; Ware and Mabe 2015; Briner and Denyer 2012). Henceforth, a significant step toward cumulative scientific knowledge advancement is the synthesis of past research findings and bibliometric data (Aria and Cuccurullo 2017). Globalization has led to the disappearance of borders, information spread, and scientific knowledge advancement. Furthermore, bibliometric studies shed light on assessing the scientific literature by quantitative attitude. These investigations offer a readable, organized, and repeatable literature review.

For a decade, bibliometric analysis has been extended to all disciplines. However, the field of Applied Linguistics, which is relatively new, has witnessed a great deal of data, multiple studies, and many scientific publications. It is getting harder and more challenging for applied linguists to deal with large and fragmented volumes of data, and in light of the rapid increase in resources in the field of applied linguistics and its various strands, the concept of science mapping has become increasingly important in order to understand the current state of the art, connections, opportunities, and main players within an established community of practitioners, themes, as well as relationships between authors and institutions. (Guler et al. 2016). It is important to draw a diagram showing the interconnectedness of items like research data, locations, themes, and author-to-author and author-to-institutional ties (Şenel and Demir 2018).

B. D. Ghorbani (✉)
Iran University of Science and Technology, Tehran, Iran
e-mail: babak_daneshvar@alumni.iust.ac.ir

According to Shi et al. (2020), bibliometric analysis is a computer program that analyzes data to determine its metrological and content characteristics. Hence, it is an objective and reliable method based on statistical techniques for analyzing information (Broadus 1987; Diodato and Gellatly 2013). Bibliometric analysis is complex due to the fact that it involves several steps. For this purpose, it is necessary to use a variety of complex and varied analysis and mapping tools, some of which may only be made available under a commercial license.

Bibliometric analysis can be done with numerous software or package programs such as VOSviewer (Eck et al. 2010), CiteSpace (Chen 2006), SciMAT (Cobo et al. 2012), Bibexcel, Biblioshiny (Aria and Cuccurullo 2017). This chapter aims to provide a thorough introduction to a free, open-source R package for conducting a systematic study of scientific literature. Moreover, The Bibliometrix package offers a collection of quantitative research instruments for bibliometrics and scientometrics. To do so, Biblioshiny integrates the features of the bibliometrix package with the simplicity of the Shiny package environment for web applications. Besides, the attributes of the bibliometrix package and the convenience of the Shiny-based web application for users led to Biblioshiny's popularity (Aria and Cuccurullo 2017).

This chapter aims to provide a general and systematic overview of scientometric studies. To do this, I developed a project to map out the status of two topics published in Language and Linguistics and Education and Educational Research. In addition, comprehensive information about the English language and technology-related publications is described in detail, and PRISMA guidelines have been followed in extracting and filtering data from WoS.

Three sections have been created to carry out bibliometric research using Biblioshiny: (1) data collection, (2) data entering and (3) data visualization (Fig. 1).

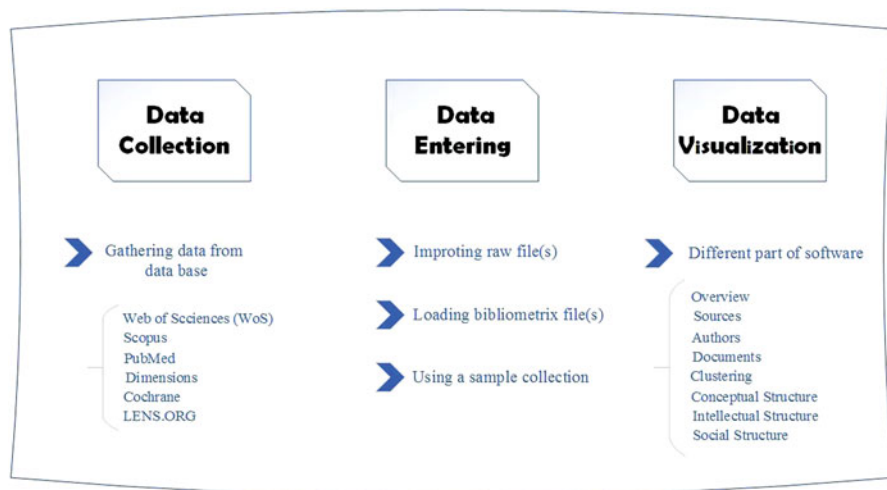


Fig. 1 Bibliometric research's steps through Biblioshiny

Data is the dynamic engine of bibliometric analysis. The clearer and more thorough our data, the more precise and accurate our results will be. According to Visser et al. (2021), “Bibliographical databases” are digital collections of references that are used to represent scientific literature in an orderly manner, such as journals, conference proceedings, patents, books, and more. In most cases, they include highly detailed descriptions of the topics covered in the form of keywords, subject categorization phrases, or abstracts. In Biblioshiny, raw files with the following formats can be imported to access major bibliometric databases: BIB, TXT, CIW, CSV, bibliometrix files: rdata, xls, and using sample collections. Bibliometric’s numerous available methods make importing bibliographic information from Scopus, Clarivate Analytics’ Web of Science, Dimensions, The Lens, PubMed, and Cochrane possible. Almost 36 million articles are available through the Web of Science (WoS), making it one of the most popular bibliographic databases. Scopus offers access to approximately 20 million articles. In terms of the quality of the data it contains, the Web of Science is superior to other databases (Aria et al. 2020; Harzing and Alakangas 2016). Despite the acceptable status of WoS metadata, three metadata items are completely missing from Scopus (Keywords Plus, Number of cited References, and Science Categories). Lastly, the data must be cleansed. In order for a calculation to be accurate, it must be based on data that is precise. A variety of pre-processing techniques can be employed, for example, to identify duplicates or misspellings. Although bibliometric data is generally reliable, there may be some discrepancies due to the fact that a single reference may include multiple editions of the same publication or different spellings of the author’s name.

2 Data Gathering

To gather the information we need, combine a list of phrases and items in the search box that may be used to locate articles in the specific field that make use of some methods that connect with the Boolean operators (“AND”, “OR”, and “NOT”). Make sure that research is limited to the target area and related fields. Next, the time period will be defined. It is also essential for the user to determine which type of document will be analyzed (article, book chapter, book review, the editorial). There is a wide range of languages used in academic papers from which to choose depending on the specifics of our investigation.

The inclusion criteria of this bibliometric study adhered to PRISMA’s guiding principles. The term PRISMA stands for preferred reporting items for systematic reviews and meta-analyses. It illustrates the process of screening in a visually appealing manner. Initially, the PRISMA records the number of articles that have been retrieved, and then it makes the selection process transparent by indicating the decisions made at different stages of the systematic review. At each stage, the number of articles is recorded. When excluding articles, researchers should include the reasons for doing so (Moher et al. 2009). The following eight WoS online indexing databases were chosen in accordance with the search strategy: Science

Scopus					WoS				
Metadata	Description	Missing Counts	Missing %	Status	Metadata	Description	Missing Counts	Missing %	Status
AB	Abstract	0	0.00	Excellent	AU	Author	0	0.00	Excellent
C1	Affiliation	0	0.00	Excellent	CR	Cited References	0	0.00	Excellent
AU	Author	0	0.00	Excellent	DT	Document Type	0	0.00	Excellent
CR	Cited References	0	0.00	Excellent	SO	Journal	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent	LA	Language	0	0.00	Excellent
SO	Journal	0	0.00	Excellent	NR	Number of Cited References	0	0.00	Excellent
LA	Language	0	0.00	Excellent	WC	Science Categories	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent	T1	Title	0	0.00	Excellent
T1	Title	0	0.00	Excellent	TC	Total Citation	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent	C1	Affiliation	5	0.57	Good
DI	DOI	11	6.01	Good	RP	Corresponding Author	5	0.57	Good
RP	Corresponding Author	16	8.74	Good	AB	Abstract	29	3.29	Good
DE	Keywords	29	15.30	Acceptable	PY	Publication Year	73	8.29	Good
ID	Keywords Plus	183	100.00	Completely missing	DE	Keywords	96	10.90	Acceptable
NR	Number of Cited References	183	100.00	Completely missing	DI	DOI	126	14.30	Acceptable
WC	Science Categories	183	100.00	Completely missing	ID	Keywords Plus	165	18.73	Acceptable

Fig. 2 Scopus and WoS metadata status

Citation Index Expanded (SCI), Social Sciences Citation Index (SSCI), Art and Humanities Citation Index (AHCI), Conference Proceedings Index – Science (CPCI-S), Conference Proceedings Citation Index – Social Science and Humanities (CPCI-SSH), Book Citation Index – Science (BKCI-S), Book Citation Index – Social Sciences and Humanities (BKCI-SSH), and Emerging Sources Citation Indexes (ESCI). Figure 2 depicts the PRISMA flow diagram utilized in the bibliometric study.

In the following query, only articles that meet the following criteria will be considered: (“Computer Assisted Language Learning” OR “CALL” OR “Mobile Assisted Language Learning” OR “MALL” OR “Robot Assisted Language Learning” OR “RALL”) AND (“Applied Linguistics” OR “ESL” OR “EFL” OR “L2 English” OR “English as a Second Language” OR “English as a Foreign Language” OR “English Teaching”) in the title, abstract or in the keyword list. Next, narrow the document topic to include only Language and Linguistics, Education and Educational Research. Ten years was chosen as the time frame for this bibliometric study (2013–2022). Therefore, only articles were selected as a document type. As a final step, English-language articles were included. Furthermore, a plain text data set was downloaded from WoS using the PRISMA flow diagram steps. According to the exclusion criteria, some records were extracted from plain text. A total of 881 articles relating to the English language and technology were analyzed in this study (Fig. 3).

3 Working with R Biblioshiny

To begin, install the most recent version of R (<https://cran.r-project.org>). Next, install the most recent version of Rstudio (<https://rstudio.com>). Having completed all these steps, open Rstudio and run the following commands in the top left window: (1) install.packages (“bibliometrix”), (2) library (bibliometrix), (3) bibliometrix:: biblioshiny (R Core Team 2014) (Figs. 4 and 5).

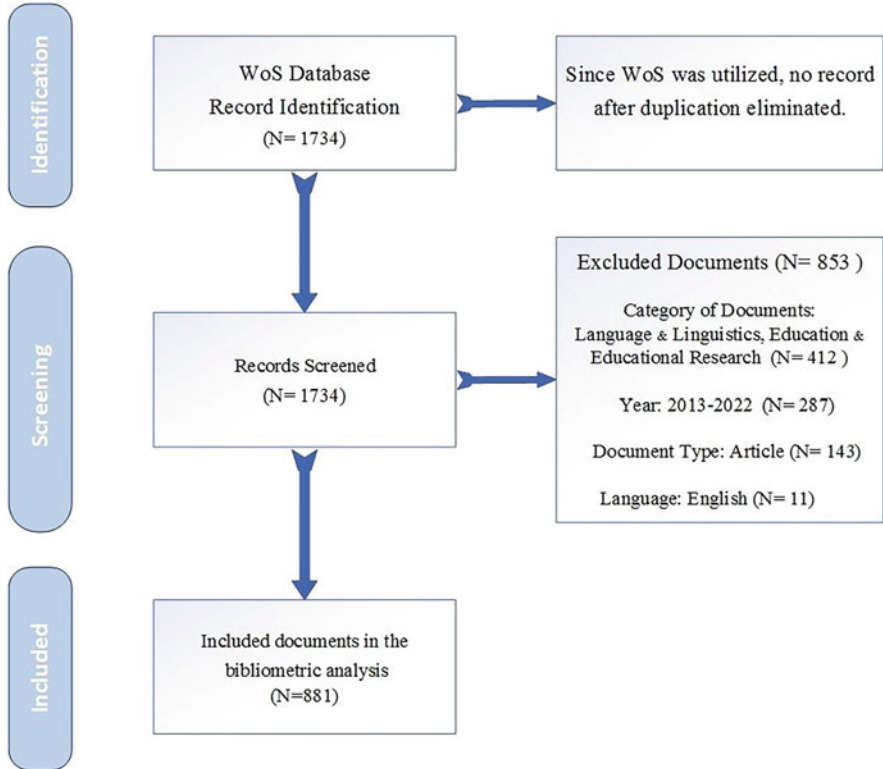


Fig. 3 PRISMA flow diagram

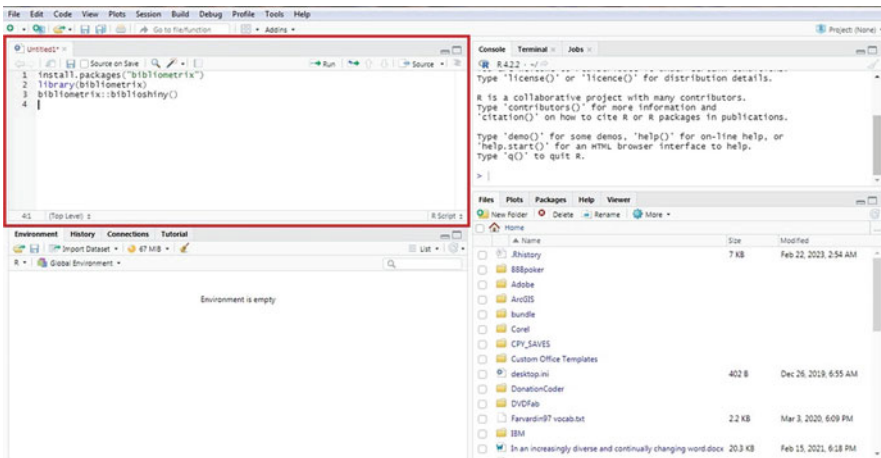


Fig. 4 Rstudio workplace

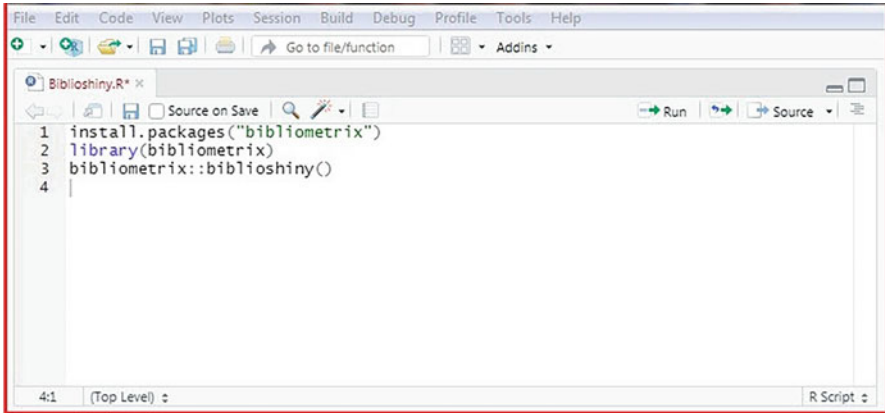


Fig. 5 Code editore





	The first icon on the screen pertains to software notifications.
	Users may be able to get package tutorials such as data conversion, import, and Bilioshiny tutorials.
	Use this donation button if you wish to contribute to the future development of the Bibliometrix package.
	With the help of this icon, user can access Bibliometrix GitHub as well as K-Synth.

Fig. 6 Basic sittings of the Biblioshiny

Once run the program, you enter to graphical user interface (GUI) of the Biblioshiny. On the top right side are four basic icons, each of which is described individually (Fig. 6).

In addition, users can add the results of specific analyses into the Export section to obtain a CSV file. Meanwhile, it should be noted that in every section of Biblioshiny, a plus icon can be used to include the results in the report. Through the system settings, which is the last icon, the user can alter general modifications of software such as resolutions of export plots as PNG as well as changing height in inches. It is possible to alter general software settings through the system settings, which is the last icon. These adjustments include adjusting the resolution of export plots as PNGs as well as changing the height in inches. It is time to explore the Biblioshiny application. The first step is to select the data format used in the study. A raw data file, a bibliometric file, or a sample collection can be used in this format. Importing and loading files using Biblioshiny’s APIs (Application Programming Interfaces) is

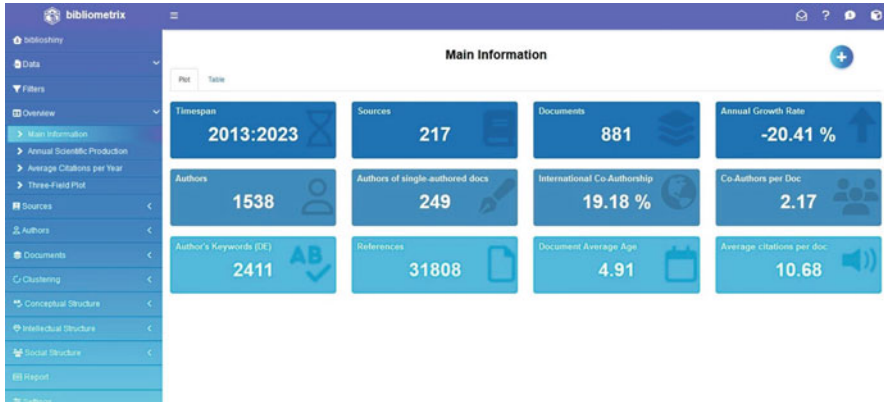


Fig. 7 Bibliometric data overview

also possible. Select the plain text file exported from WoS from the browse button, and then select “Web of Science (WoS/WoK)” as the database. Data from other databases, such as Scopus, PubMed, [Lens.org](https://www.lens.org), Dimensions, and Cochrane Library, can be analyzed. The number of documents will appear in the conversion results once the user clicks the start button. Figure 7 shows that 881 documents have been uploaded to the app with different information. An in-depth literature review of the study can be illustrated with Biblioshiny’s user-friendly interface.

3.1 Overview

Figure 7 shows the most important information about the documents in the boxes. A total of 881 articles were included in the study from 217 sources between 2013 and 2022. There were 881 articles related to the English language and technology written by 1538 authors. One author wrote 249 articles. The average number of citations per document was 10.68. Biblioshiny also offered the option to download each plot in multiple formats (CSV, Excel, PDF) and different resolutions. The plots can be exported in a resolution ranging from 75 dpi to 600 dpi.

According to Fig. 8, the highest number of publications was published in 2022 ($N = 120$). The number of documents produced during this period has fluctuated since 2013. Besides, graphs generated with Plotly are all dynamic, so you can see more information by moving your mouse over them (Table 1).

The user can get a graph illustrating the rate of citations over 10 years by clicking on the average citation per year button. Interestingly, the average number of citations per year in 2019 is the highest rate, indicating that one or more articles published in 2019 received the most citations ($N = 2.39$) (Fig. 9 and Table 2).

Three-field plots can be selected based on the objective of the researchers by choosing from authors, affiliations, countries, keywords, abstracts, references, and

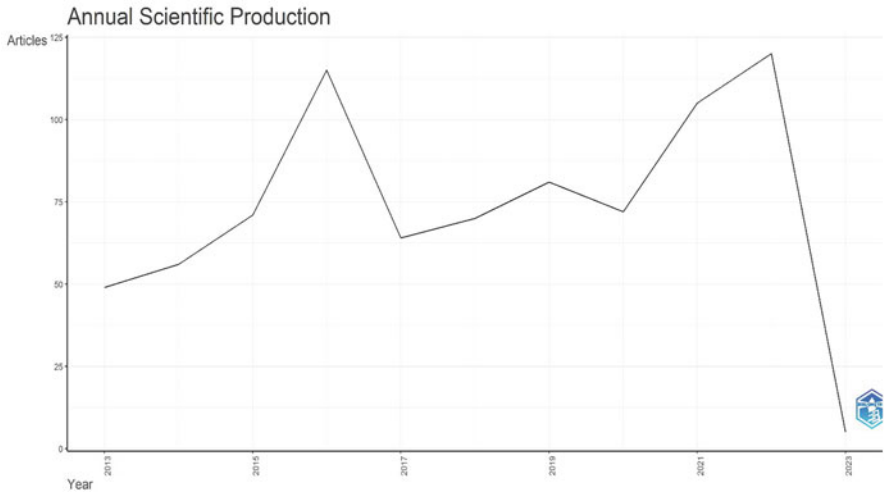


Fig. 8 Citations per year for articles

Table 1 Articles published annually

Years	Articles
2013	49
2014	56
2015	71
2016	115
2017	64
2018	70
2019	81
2020	72
2021	105
2022	120

cited sources in accordance with their objectives. Determining the variables and number of items in the middle, left, and right fields is necessary. Upon setting these conditions, clicking the apply button will produce the plot. In order to create it, three meta-data fields must be selected. In this bibliometric study, the author occupied the middle field, the country occupied the left field, and the keywords occupied the right field (Fig. 10).

3.2 Sources

The y-axis shows the number of journals and sources that have published at least one of the documents in the bibliometric repertoire, while the x-axis shows the total number of articles published. Computer Assisted Language Learning (N = 226) was

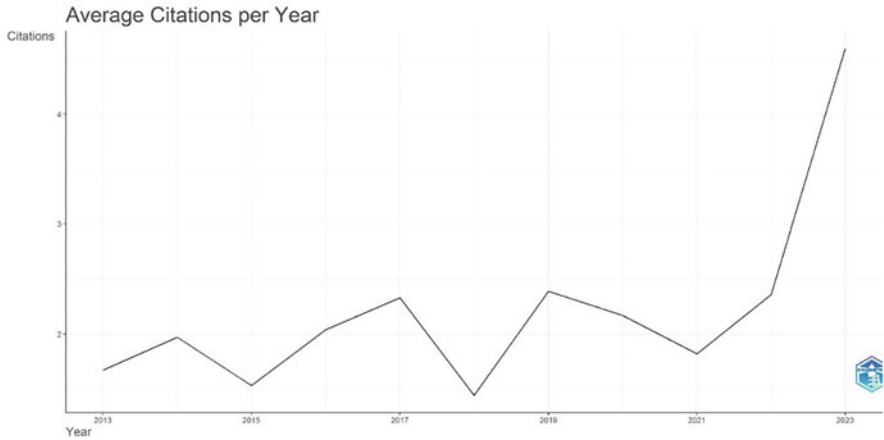


Fig. 9 Citations per year for articles

Table 2 Annual total citation per year

Year	MeanTCperArt	N	MeanTCperYear	CitableYears
2013	18.41	49.00	1.67	11
2014	19.7	56.00	1.97	10
2015	13.76	71.00	1.53	9
2016	16.3	115.00	2.04	8
2017	16.33	64.00	2.33	7
2018	8.66	70.00	1.44	6
2019	11.94	81.00	2.39	5
2020	8.67	72.00	2.17	4
2021	5.45	105.00	1.82	3
2022	4.72	120.00	2.36	2

the most commonly published journal in this research, International Journal of Computer-assisted Language Learning and Teaching (N = 66), System (N = 30), and Language Learning and Technology (N = 28). Arab World English Journal and RECALL (N = 25). Users may find it useful to click on the options button at the top of the page to find an option that allows them to add more sources to the list (Fig. 11 and Table 3).

Bradford’s law categorizes articles and journals to identify the relevant number for each topic. Its calculation is based on the sum of articles published by journals in a specific field of study. By doing so, it is possible to identify the core of the journals, which can then be used to analyze the core zone documents. Bradford defined the first zone as the center of specialized journals. In addition, Bradford’s law is also suitable for classifying journals based on their number of citations. It can be concluded that a relatively small number of journals produce a substantial amount of high-quality science based on citations (Fig. 12).

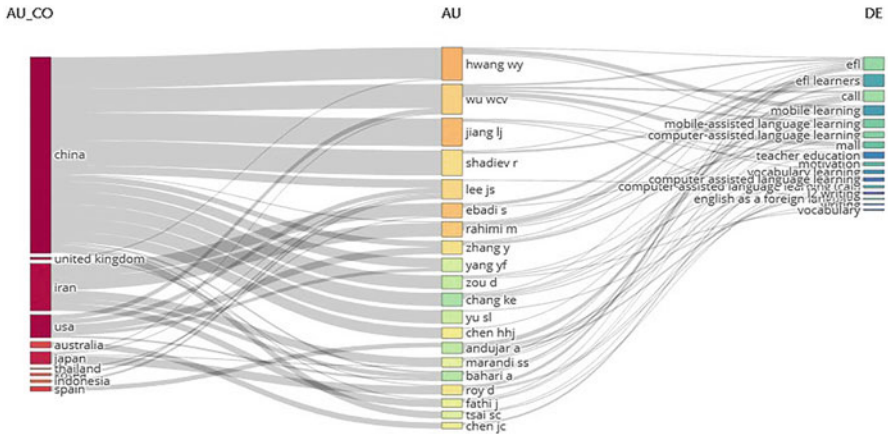


Fig. 10 Three-field plot

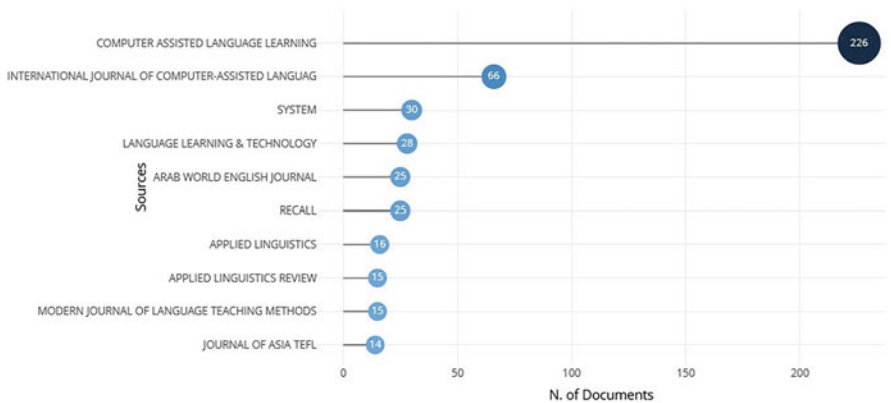


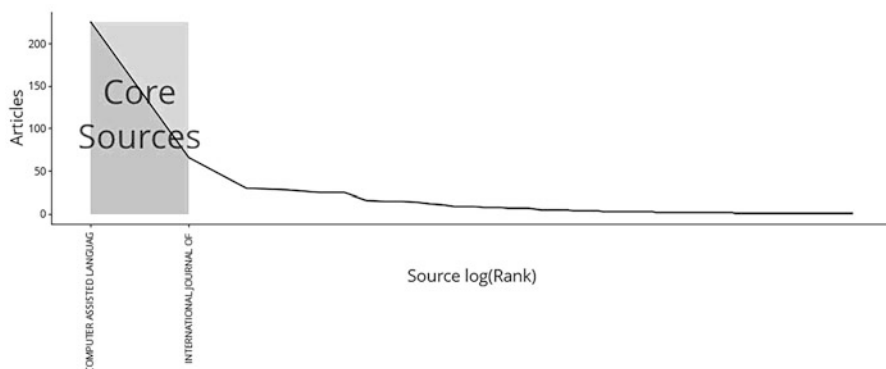
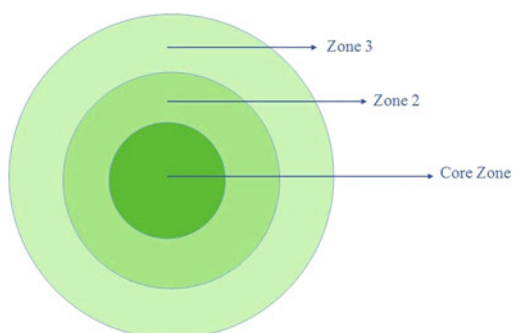
Fig. 11 Most relevant sources

Computer-Assisted Language Learning Journal and International Journal of Computer Assisted Language Learning and Teaching are two of the core zones among the 217 journals. Consequently, two journals with 292 articles are in the core zone, 26 journals with 301 articles are in the middle zone, and 189 journals with 282 articles are in the minor core (Fig. 13).

The H-index, also known as the Hirsch index, is the number of an author’s or journal’s works that have been mentioned in at least one other piece of research. Hence, this index is a more refined version of basic metrics like overall publications or citation counts. The index works best when comparing researchers in the same area due to the large variation in publications across fields. For example, if seven articles from a researcher’s total publication are cited at least seven times in other articles, the H-index of that researcher is seven. Therefore, The H-index is considered the consequence of the equilibrium between articles and citations. In addition to

Table 3 Journal article publication rate

Sources	Articles
COMPUTER ASSISTED LANGUAGE LEARNING	266
INTERNATIONAL JOURNAL OF COMPUTER-ASSISTED LANGUAGE LEARNING AND TEACHING SYSTEM	66
LANGUAGE LEARNING AND TECHNOLOGY	30
ARAB WORLD ENGLISH JOURNAL	28
RECALL	25
APPLIED LINGUISTICS	25
APPLIED LINGUISTICS REVIEW	16
MODERN JOURNAL OF LANGUAGE TEACHING METHODS	15
JOURNAL OF ASIA TEFL	14

Fig. 12 Bradford's law zone**Fig. 13** Bradford's law core sources

other science measurement indices, such as the number of publications and citations, this index distinguishes influential researchers from those who produce many articles. The m-index is calculated by dividing the scientist's H-index by the number of years (n) from his or her first publication (journal).

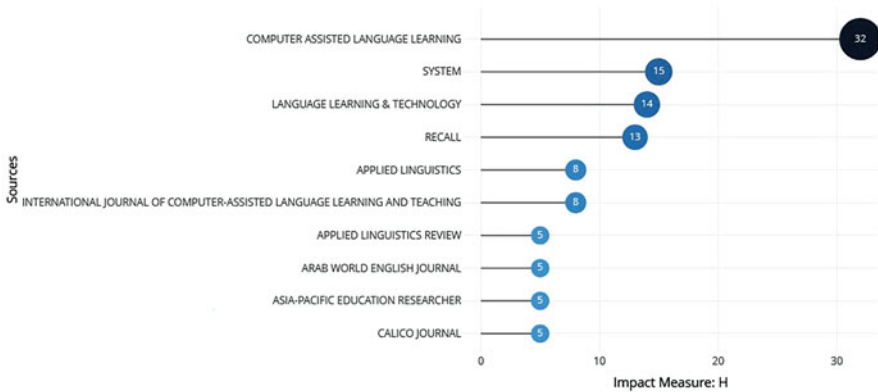


Fig. 14 The local impact of journals by H-index

Table 4 Information on local impacts of journals

Sources	h-index	g-index	m-index	TC	NY	PY-start
COMPUTER ASSISTED LANGUAGE LEARNING SYSTEM	32	46	–	3709	226	–
LANGUAGE LEARNING AND TECHNOLOGY	15	23	1.364	581	30	2013
RECALL	14	25	1.273	657	28	2013
APPLIED LINGUISTICS	13	21	1.182	468	25	2013
INTERNATIONAL JOURNAL OF COMPUTER-ASSISTED LANGUAGE LEARNING AND TEACHING	8	16	0.800	625	16	2014
APPLIED LINGUISTICS REVIEW	8	12	0.727	245	66	2013
ARAB WORLD ENGLISH JOURNAL	5	12	–	159	15	–
MODERN JOURNAL OF LANGUAGE TEACHING METHODS	5	7	0.455	72	25	2013
ASIA-PACIFIC EDUCATION RESEARCHER	5	7	–	50	8	–
CALICO JOURNAL	5	8	0.455	66	9	2013

Compared to the H-index, the g-index is superior as a worldwide indicator of an article’s citation impact. If sorted from most to least citations, the g-index is the (unique) greatest number such that the top g articles received (collectively) at least g² citations. Taking into account source local impact by H index, Computer Assisted Language Learning (H-index = 32), System (H-index = 15), Language Learning and Technology (H-index = 14), and RECALL (H-index = 13) were ranked in that order (Fig. 14 and Table 4).

The user will find the publications on the left, while developments for each year can be seen on the right. The number of articles related to English Language and Technology has increased for each journal since 2012. The graph shows that

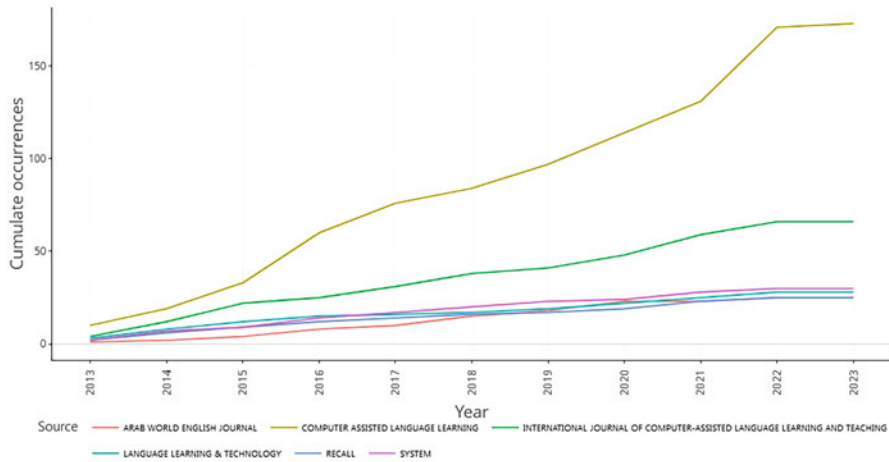


Fig. 15 Production of journals over time

Computer Assisted Language Learning Started with ten articles in 2013 and reached the highest number among other journals in 2022 ($N = 171$) (Fig. 15 and Table 5).

3.3 Authors

This section aims to illustrate bibliometric analysis through information about authors, such as the most relevant authors, most locally cited authors, top authors' production over time, Lotka's law, and authors' local impact. In regards to affiliations, the most relevant affiliations will be discussed along with their production over time. Furthermore, a bibliometric analysis of countries will be performed, including the corresponding authors' countries, the scientific output of countries, the countries' production over time, and the most cited countries. The following plots (Figs. 16 and 17) and Tables (6 and 7) illustrate the development of an author's publications and citations over time. In plots 16 and 17, it can be seen that HWANG WY had the highest number of publications ($N = 11$) and citations ($N = 65$).

Figure 18 illustrates the development of authors' production over time. The circle will likely be darker and larger if you have published more articles than in other years. As an example, LEE JS had 4 publications in 2019 and an average of 22 citations per year. As the number of documents increases, the bubble size increases. Color intensity is directly proportional to the number of citations per year and the number of citations per year. There were 3 articles authored by RAHIMI, M in 2022, and the number of citations he received per year was 22.5. The light pink line indicates the authors' timeline. Among all the ten-year activities, HWANG WY has the longest timeline (2013–2022).

Table 5 Annual production overview

Year	COMPUTER ASSISTED LANGUAGE LEARNING	INTERNATIONAL JOURNAL OF COMPUTER-ASSISTED LANGUAGE LEARNING AND TEACHING	SYSTEM	LANGUAGE LEARNING & TECHNOLOGY	ARAB WORLD ENGLISH JOURNAL	RECALL
2013	10	4	2	3	1	2
2014	19	12	7	8	2	6
2015	33	22	9	12	4	9
2016	60	25	14	15	8	12
2017	76	31	17	16	10	14
2018	84	38	20	17	15	16
2019	97	41	23	19	18	17
2020	114	48	24	22	23	19
2021	131	59	28	25	23	23
2022	171	66	30	28	25	25

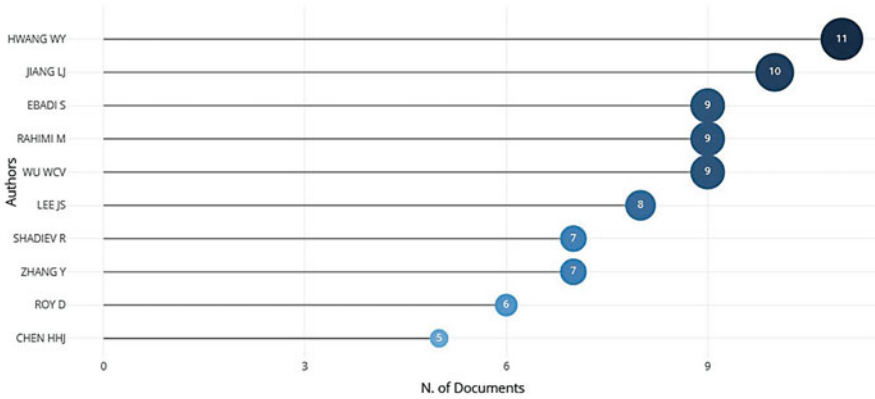


Fig. 16 Most relevant authors

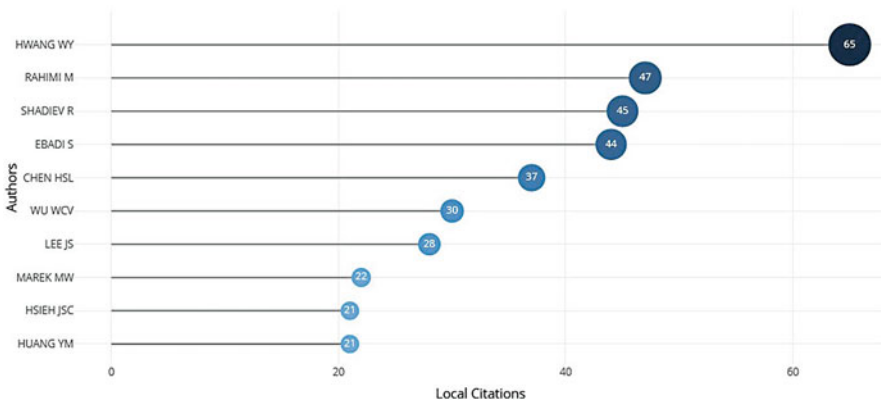


Fig. 17 Most locally cited authors

Table 6 The number of articles per author

Authors	Articles	Articles Fractionalized
HWANG WY	11	2.88
JIANG LJ	10	4.58
EBADI S	9	4.00
RAHIMI M	9	4.25
WU WCV	9	3.50
LEE JS	8	5.33
SHADIEV R	7	1.90
ZHANG Y	7	5.33
ROY D	6	4.83
CHEN HHJ	5	2.58

Table 7 Citation information of authors

Author	Local Citations
HWANG WY	65
RAHIMI M	47
SHADIEV R	45
EBADI S	44
CHEN HSL	37
WU WCV	30
LEE JS	28
MAREK MW	22
HSIEH JSC	21
HUANG YM	21

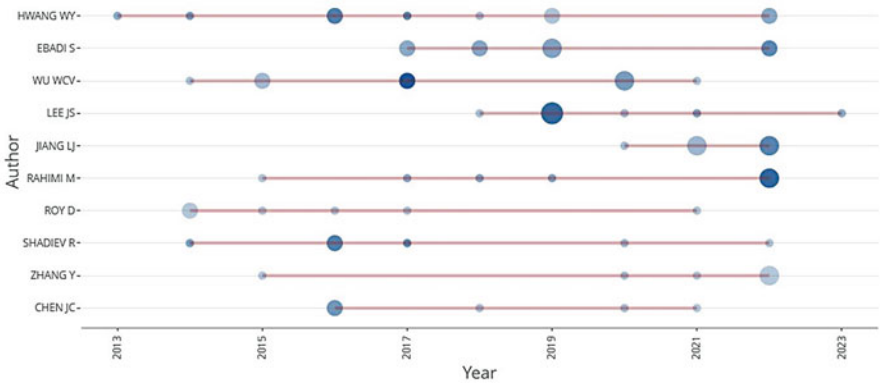


Fig. 18 Production of authors over time

Lotka’s law describes the frequency of publication by authors in different fields. There is a fixed ratio between the number of authors publishing a certain number of articles and the number of authors publishing a single article. Lotka’s law can be mathematically represented using a formula showing how it works. Where A represents the number of publications, B is the relative frequency of authors with A publications, n and C are constants that vary depending on the field ($n \approx 2$).

$$A^n B = C$$

Lotka’s law states that as the number of articles published increases, the number of authors who produce many publications will decrease. A total of 1321 authors (85.1%) have only published one article. This baseline represents authors who have published more than one article, which reveals that 142 authors (0.1%) have 2 publications. Based on this baseline, which represents the number of core authors who have published at least two articles, it can be seen that 142 authors (0.1%) have published at least two articles. On the other hand, the term occasional authors refers

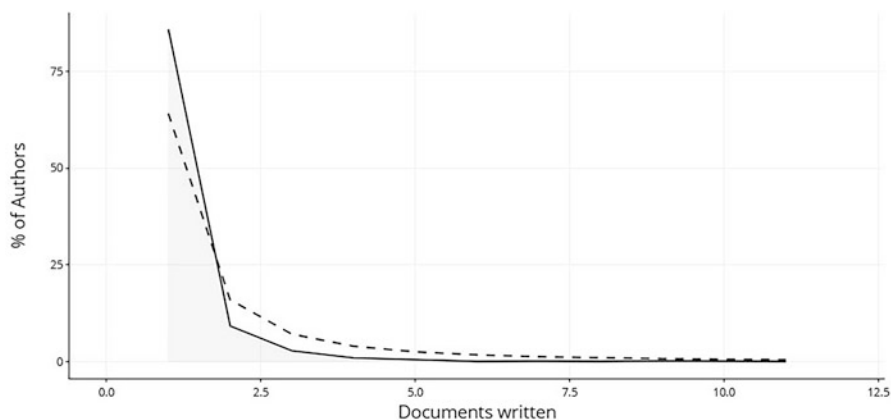


Fig. 19 Lotka's law of author productivity

Table 8 Author productivity based on Lotka's law

Documents Written	N. of Authors	Proportion of Authors
1	1321	0.859
2	142	0.092
3	43	0.028
4	15	0.010
5	8	0.005
6	1	0.001
7	2	0.001
8	1	0.001
9	3	0.002
10	1	0.001
11	1	0.001

to 1321 authors who have published only one publication (86%) (Fig. 19 and Table 8).

According to Fig. 20, authors are categorized based on the H-index of their publications. Three authors (EBADI S, HWANG WY, LEE JS) in the top tier have recorded an H-index of 7.

3.3.1 Affiliations

As found in this bibliometric analysis, the majority of the authors were from Islamic Azad University in Iran ($N = 36$), National Taiwan Normal University in China ($N = 26$), Education University in Hong Kong ($N = 20$), and Iowa State University in the USA ($N = 19$) (Fig. 21 and Table 9).

As shown in Fig. 22, the production of the affiliation has increased over time. Iowa State University, among other affiliations, had the highest rate ($N = 5$) in 2013,

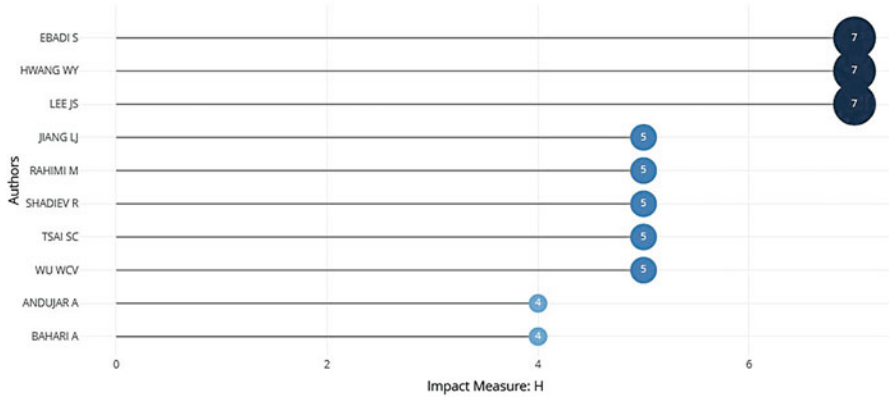


Fig. 20 Local impact of authors

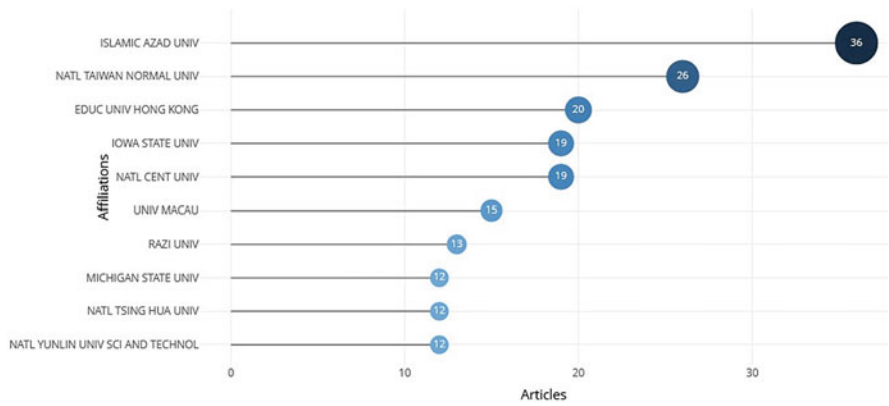


Fig. 21 Highly affiliated institution

Table 9 Institutional rate of publication

Affiliation	Articles
ISLAMIC AZAD UNIV	36
NATL TAIWAN NORMAL UNIV	26
EDUC UNIV HONG KONG	20
IOWA STATE UNIV	19
NATL CENT UNIV	19
UNIV MACAU	15
RAZI UNIV	13
MICHIGAN STATE UNIV	12
NATL TSING HUA UNIV	12
NATL YUNLIN UNI SCI AND TECHNOL	12

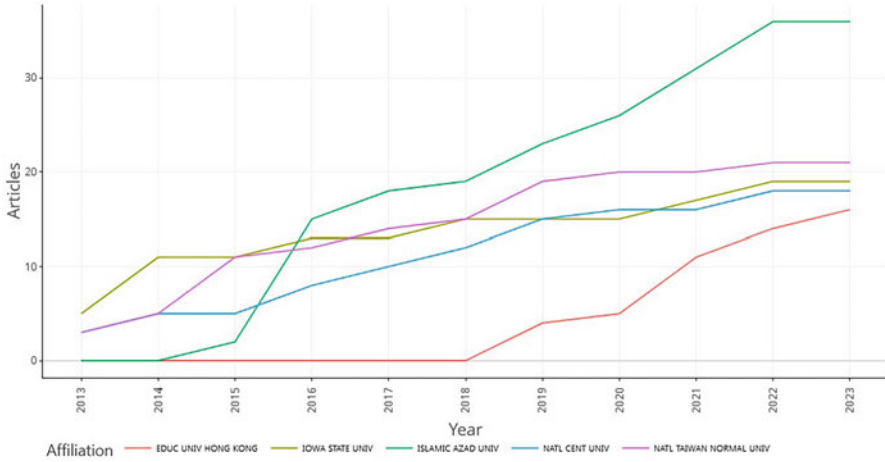


Fig. 22 Production over time by institutions

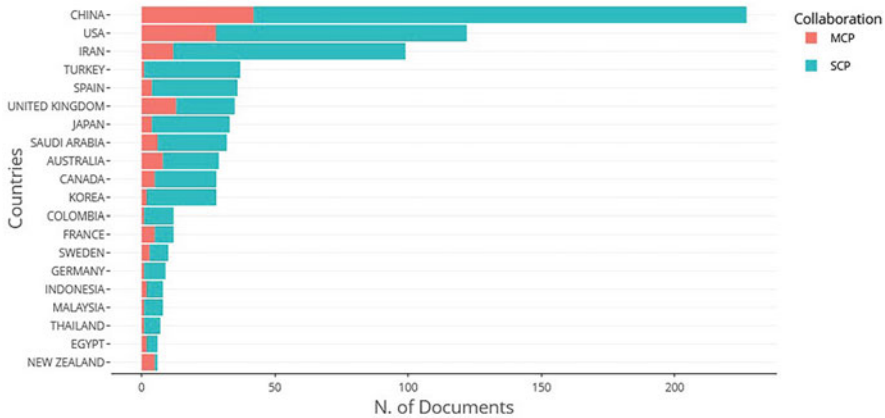


Fig. 23 Countries of corresponding authors

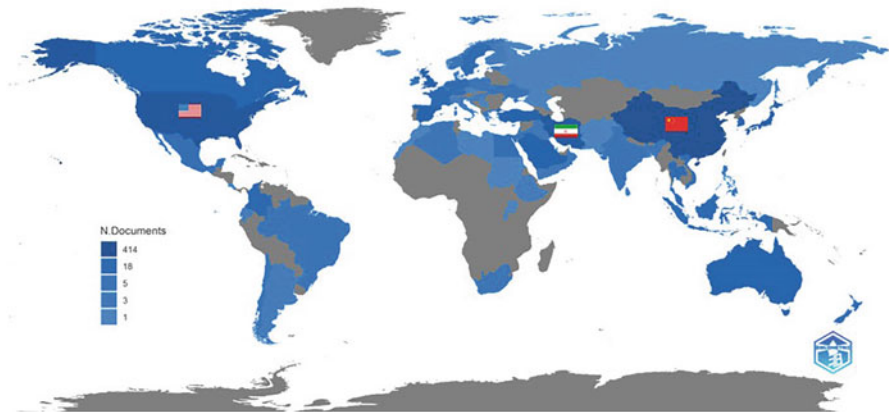
whereas Islamic Azad University, which has been growing steadily over the years, hit peak levels in 2022 (N = 36).

3.3.2 Countries

According to Fig. 23, the authors’ countries of the articles were given, the corresponding authors’ countries indicated multiple country publications (MCPs), and the turquoise bar indicated single country publications (SCP). Accordingly, MCP indicates for each country the number of documents in which at least one coauthor is from another country. MCP measures a country’s international

Table 10 Information on corresponding authors' countries

Country	Articles	SCP	MCP	Freq	MCP-Ratio
CHINA	227	185	42	0.258	0.185
USA	122	94	28	0.138	0.230
IRAN	99	87	12	0.112	0.121
TURKEY	37	36	1	0.042	0.027
SPAIN	36	32	4	0.041	0.111
UNITED KINGDOM	35	22	13	0.040	0.371
JAPAN	33	29	4	0.037	0.121
SAUDI ARABIA	32	26	6	0.036	0.188
AUSTRALIA	29	21	8	0.033	0.276
CANADA	28	23	5	0.032	0.179

**Fig. 24** Scientific production of the world

collaboration intensity. There were 227 corresponding authors from China. There is a total of 200 authors from the United States and 99 from Iran. However, most of the authors' collaborations were limited to contributions from a single country. New Zealand had a high level of international collaboration ($N = 6$, $MCP = 5$) (Table 10).

Every author's nationality contributed to the anthology is considered in this map. There is a direct correlation between the number of articles and the magnitude of the color. As the number of documents increases, the color of countries will be darker. For instance, with 414 documents, China is darker than India, with five documents. Moreover, the USA and Iran, which had a darker color than other countries, were the two countries that ranked second and third after China in terms of the frequency of publications (Fig. 24 and Table 11).

Let's take a closer look at the results of our study regarding countries' production since 2013, where China ($N = 344$) was the leading country, the USA, with

Table 11 Frequency of publication per country

Region	Freq
CHINA	414
USA	226
IRAN	146
TURKEY	75
SPAIN	66
UNITED KINGDOM	58
JAPAN	57
SAUDI ARABIA	54
AUSTRALIA	51
CANADA	46

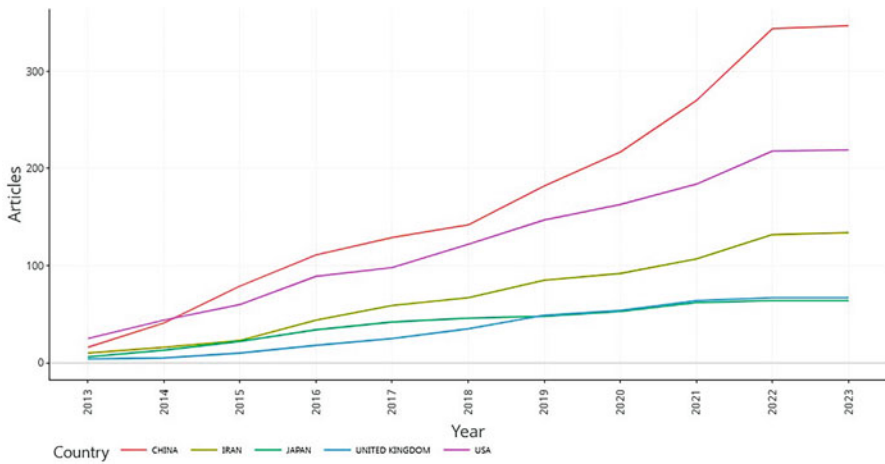


Fig. 25 Country production over time

218 documents, was the second country, and Iran (N = 132) was chasing them as the third (Fig. 25).

The most cited countries were China (N = 2962), the USA (N = 1547), Iran (N = 719), and Spain (N = 641), respectively. Although the total number of citations in New Zealand, which is in the twelfth place, was approximately one-seventeenth of that of the leading country, China, the average number of article citations in New Zealand had the highest value with 28.67 (Fig. 26 and Table 12).

3.4 Documents

The next section of the software will provide insight into some of the interesting features of the Biblioshiny by analyzing documents such as documents with the greatest local and global citations, references with the greatest local citations,

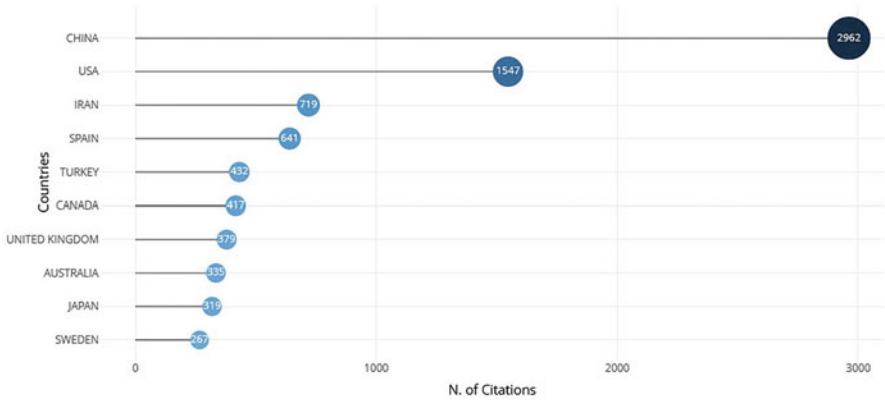


Fig. 26 Frequently cited countries

Table 12 Information on countries' citation

Country	TC	Average Article Citation
CHINA	2962	13.05
USA	1547	12.68
IRAN	719	7.26
SPAIN	641	17.81
TURKEY	432	11.68
CANADA	417	14.89
UNITED KINGDOM	379	10.83
AUSTRALIA	335	11.55
JAPAN	319	9.67
SWEDEN	269	26.70

reference spectrums, the most frequently used words, a word cloud, a tree map, and the frequency of words over time are all displayed (Fig. 27).

3.4.1 Cited References

Cenoz J.'s 2014 masterwork in the journal *Applied Linguistics* garnered the highest number of citations globally (N = 260). Additionally, the article published by HWANG WY in the Computer Assisted Language Learning journal received the most local citations in 2013 (N = 20). Furthermore, the global Citation metric measures the number of citations a document has received from all documents in the database (e.g., WoS or Scopus). Global citations indicate a document's impact across the entire bibliographic database. For many documents, a significant portion of global citations may originate from a discipline other than the one in which the document is published. In contrast, local citations are a document's citations from the studied collection.

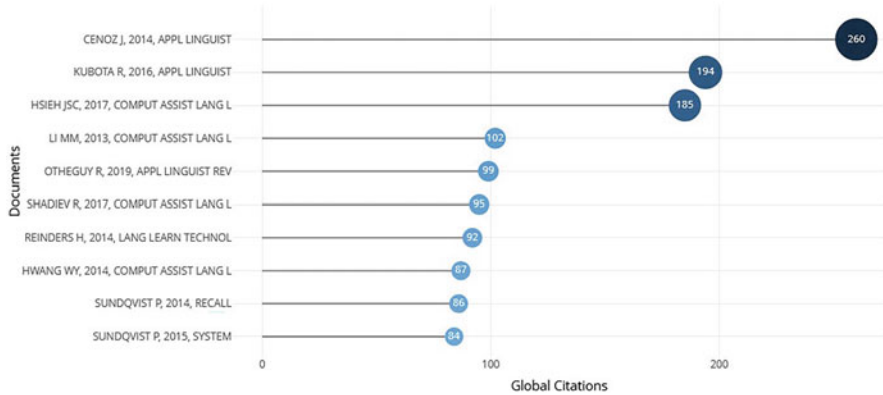


Fig. 27 Documents with global citation

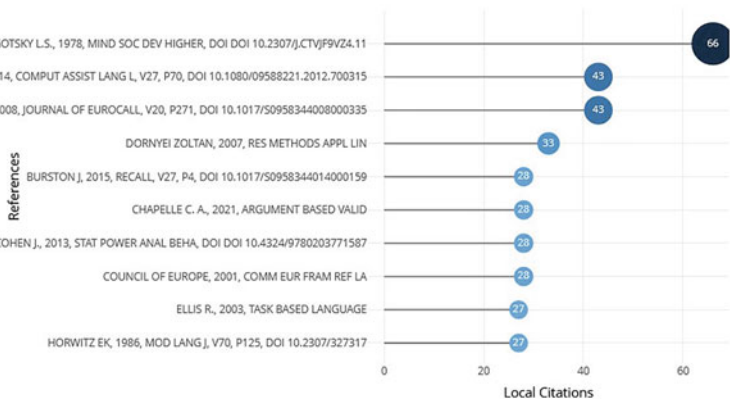


Fig. 28 The most cited local references

Similarly, the following figure illustrates the local references that have received the most citations. With 66 citations, VYGOTSKY L.S., 1978, Mind in Society: Development of Higher Psychological Processes, had the highest number of local citations. The second place went to GOLONKA, EM et al. for their computer-assisted language learning work with 43 citations (Fig. 28).

A method for determining the historical roots of research themes and topics is known as Reference Publication Year Spectroscopy (RPYS). RPYS highlights years with important findings in a paper’s cited reference profile. By doing so, it is possible to identify the historical roots of a discipline (Marx et al. 2014). A black line represents the number of citations per year, while a red line represents deviations from the median over the past 5 years (Fig. 29).

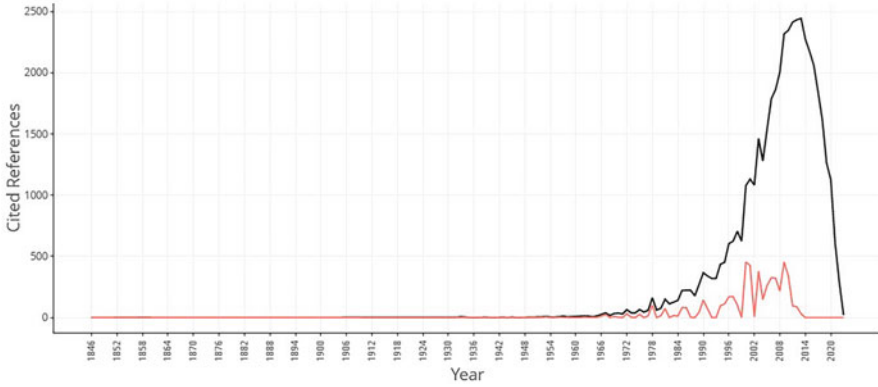


Fig. 29 Reference publication year spectroscopy

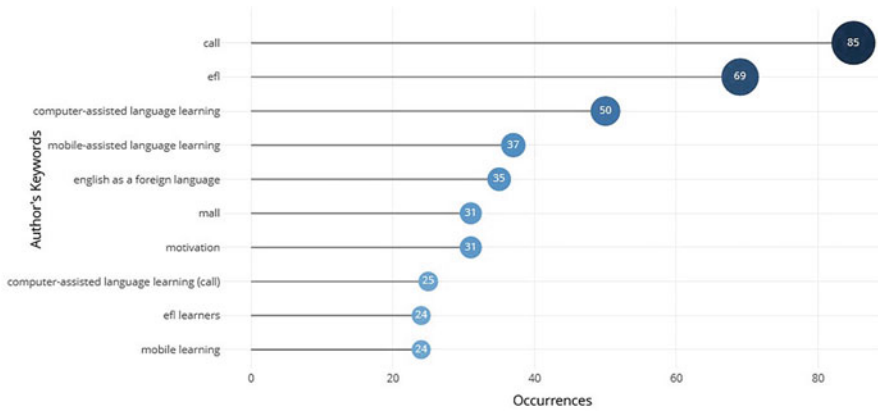


Fig. 30 Most frequent keywords

3.4.2 Words

If the user conducts keyword analysis, he or she may encounter some irrelevant words, such as conjunctions, adverbs, and plurals, that are irrelevant to the search. However, it is possible to eliminate words with the help of Bibliometrix for the time being. This issue can be avoided by creating a stop word list and uploading a TXT or CSV file that contains the words the user wishes to remove. Commas or semicolons should be used to separate or arrange items on a table. Additionally, margination analysis is plausible when plausible. This method is similar to the previous one: upload a synonym list in TXT or CSV format. Please note that the first word in the list will be replaced with the following (Fig. 30).

CALL was the most frequent word, with 85 occurrences, followed by EFL (N = 69) and Computer-assisted language learning (N = 50) in the top three most



Fig. 31 WordCloud derived from keywords

frequent words. Various parts of the document may contain keywords that can be extracted and analyzed.

In WordCloud, the top keyword plus, subject categories, author keywords, title, and abstract words are displayed. Additionally, by adjusting parameters in the options, we are able to change some default settings, such as the shape, font type, and font color. Different scale transformations can be selected for a more vivid picture of word distribution, such as frequency, square root, log, and log10 (Fig. 31).

TreeMap uses color blocks to display how frequently different words occur visually. In a similar vein, you may analyze the title, abstract, and keywords of a document. Based on the results of TreeMap, we identified concepts such as CALL (N = 85, 10%), EFL (N = 69, 8%), computer-assisted language learning (N = 50, 6%), mobile-assisted language learning (N = 37, 4%), and English as a foreign language (N = 35, 4%) (Fig. 32).

It is possible to determine the most frequent word, year, and frequency by passing the mouse over the graph. The term “CALL” had 5 times the frequency at the beginning of the study, and by the end of 2022, it had increased to 83 times the frequency. In the beginning, there were similar occurrences of computer-assisted language learning and EFL concepts (N = 5), but EFL reached 64, and computer-assisted language learning was ranked third with 48 occurrences (Fig. 33).

Trend-topic graphs are scatter diagrams in which the x-axis represents time, and the y-axis represents the topic. Each topic’s reference year is determined based on the median of occurrences during the period considered. On the graph, each bubble represents a topic. There is a direct correlation between the size of the bubble and the number of occurrences of the words. Thus, the bubble’s size indicates the term’s frequency and the length of the lines indicates how long it has been studied (Fig. 34).



Fig. 32 TreeMap derived from keyword

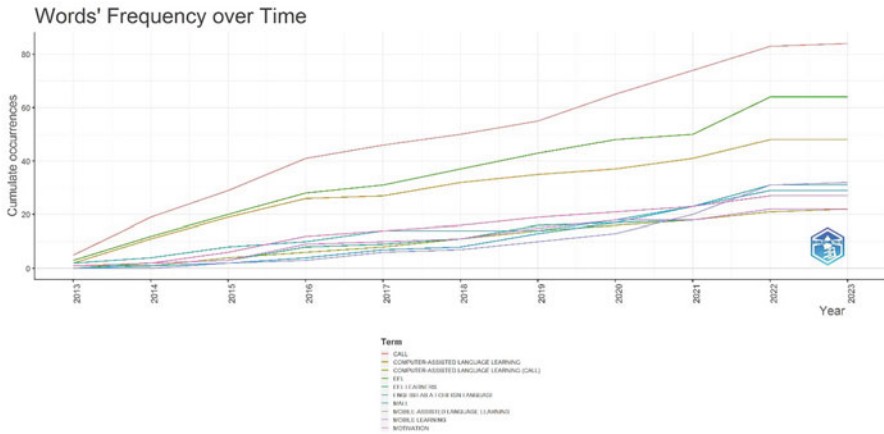


Fig. 33 Frequency of words over time

The author’s keyword indicates that English as a foreign language has been a trending topic. Based on the analysis, the other trending topics were English as a foreign language, computer-assisted language learning, and CALL.

3.5 Clustering

The concept of bibliographic coupling can be viewed as the opposite of co-citation. In other words, it refers to a relationship between two or more documents that cite one another. Bibliographic coupling assumes that two documents can be significantly correlated even if they do not directly cite each other. Assuming that the two documents share at least one bibliographic reference, this is a reasonable

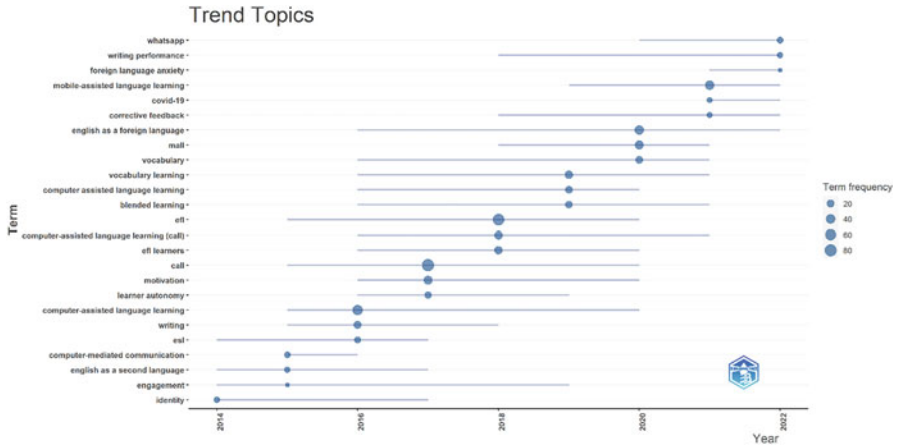


Fig. 34 Trends topic based on author’s keyword

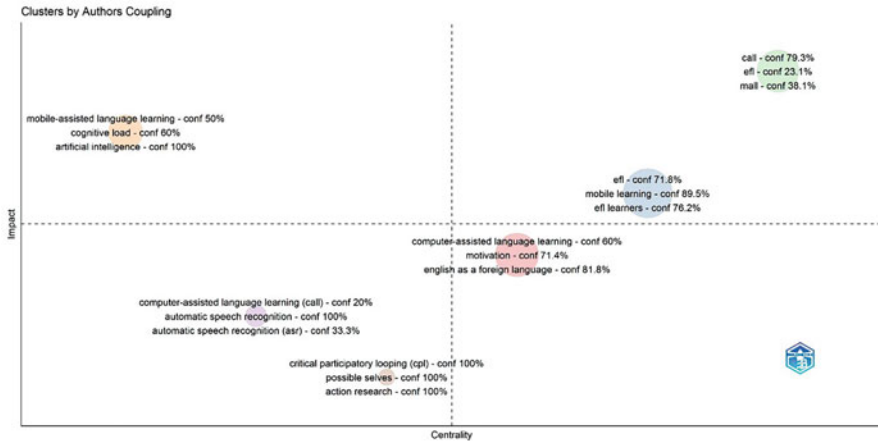


Fig. 35 Clustering by authors’ coupling

assumption. Based on their symmetrical alignment and similar references, these documents are similar. The next few figures will illustrate how network clustering, maps, and clusters work by coupling them. It should be noted that some parameters can be altered based on our bibliographic needs, such as the units of analysis (documents, authors, sources), the impact of measures (local citation score, global citation score), and attributes (cluster labeling, coupling measure) (Figs. 35 and 36 and Table 13).

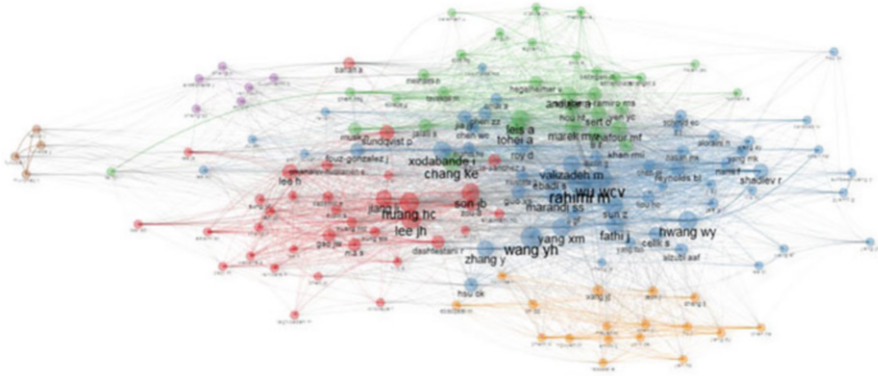


Fig. 36 Network map based on authors

Table 13 Information of clusters based on keywords

Label	Group	Freq	Centrality	Impact	Color
Computer-assisted language learning – conf 60% motivation – conf 71.4% English as a foreign language – conf 81.8%	1	35	0.59	1.73	#E41A1C80
EFL – conf 71.8% mobile learning – conf 89.55% EFL learners – conf 76.2%	2	56	0.74	1.78	#377 EB880
call – conf 79.3% EFL – conf 23.1% mall – conf 38.1%	3	32	0.78	2.45	#4DAF4A80
Computer-assisted language learning (CALL) – conf 20% automatic speech recognition – conf 100% automatic speech recognition (ASR) – conf 33.3%	4	6	0.43	1.43	#984EA380
Mobile-assisted language learning – conf 50% cognitive load – conf 60% artificial intelligence – conf 100%	5	17	0.43	1.90	#FF7F0080
Critical participatory looping (CPL) – conf 100% possible selves – conf 100% action research – conf 100%	6	4	0.58	0	#A6562880

3.6 Conceptual Structure

This conceptual structure establishes a map of the scientific field by assessing correspondences, multiple correspondences, and clustering of terms in a 2-dimensional network using a vertices network of terms obtained from the keyword, title, and abstract fields (Hubert 1980). Also, it is composed of three main parts: a co-occurrence network, a thematic map, and a thematic evolution chart. The probable link between two bibliographic items appearing in the same research is evaluated in co-occurrence network analysis. Figures 45 and 46 illustrate the co-occurrence network and degree plot analysis of the author’s keyword. According to the size of the bubbles in the figure, CALL, EFL, and computer-assisted language

learning are the most frequent keywords, ranked accordingly, as well as in the degree plot diagram (Figs. 37 and 38).

According to the thematic map, there were a total of 12 clusters identified. Based on the author’s keywords, the basic (developing) themes were (English, language, student), (call, knowledge, communication), and (competence, second-language acquisition, and negotiation). Modality was an emerging or declining theme. In terms of the motor (developed) theme, no parameters were available. Among the niche topics that were covered were (ESL, world English), (Japanese, words, speech perception), (organization, repair, conversation analysis), (attention, discourse, patterns), and (technology acceptance). A few invariance clusters are located between two of the basic quadrants and the motor quadrants, such as (technology, education, performance) and (accuracy, complexity, and performance). The video, input, and environment clusters were also located among the basic, motor, and emerging themes (Fig. 39).

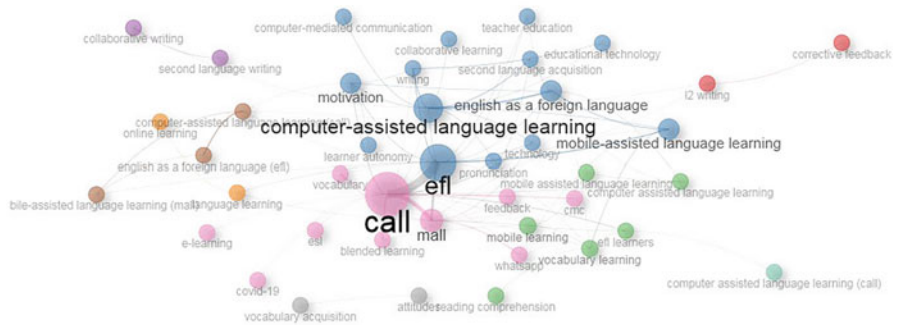


Fig. 37 Co-occurrence network based on keywords

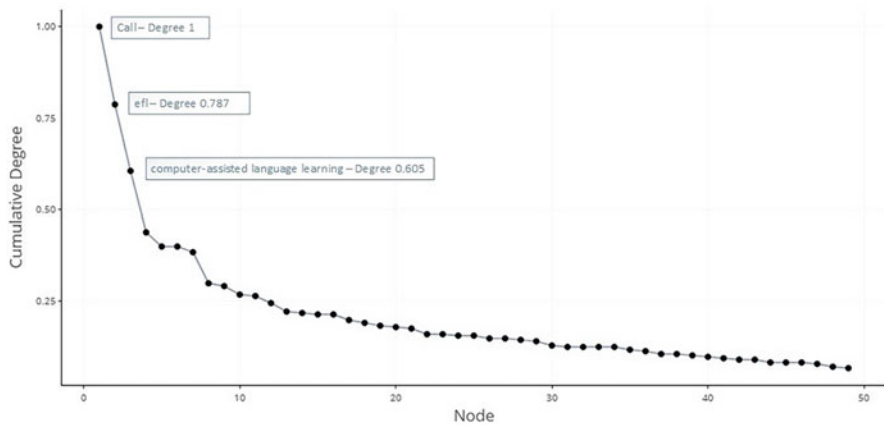


Fig. 38 Degree plot of co-occurrence network based on keywords

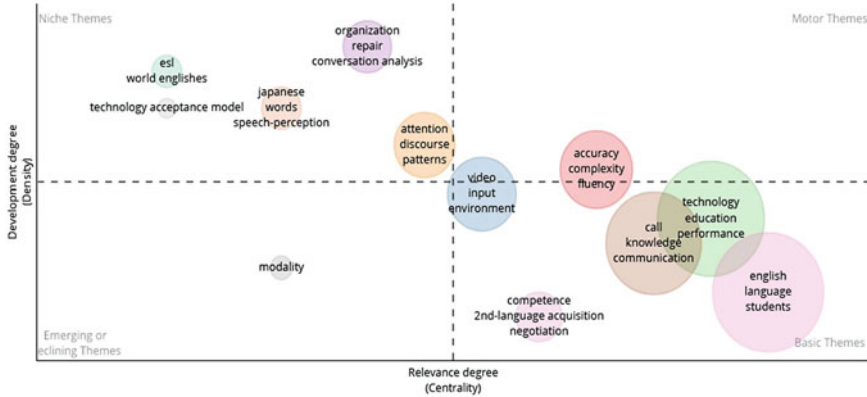


Fig. 39 Thematic map of keywords

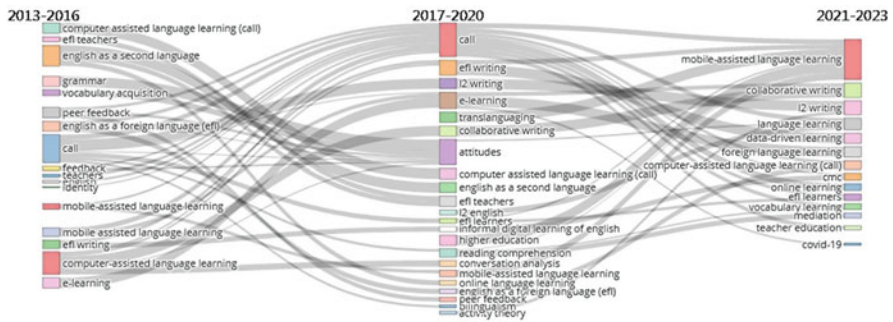


Fig. 40 Thematic evolution of keywords

The importance of identifying changes in terminology and the evolution of research fields in disciplines such as bibliometrics and scientometrics cannot be overstated. Thematic evolution analysis is a method used to reveal hidden key elements of the study, such as topics, by looking at the theme of the study about a specific topic but from different periods. Figure 47 depicts the evolution of keywords over three distinct periods (2013–2016, 2017–2020, and 2021–2023). It is important to note that the keywords “computer assisted language learning”, “call”, and “mobile assisted language learning” are important keywords as they are present in all three stages; however, covid-19 is only presented in the third stage. Most of the studies focused on using various technologies to improve language learning (Fig. 40).

3.6.1 Factorial Analysis

After conducting a factorial analysis of the author’s keywords of the articles about the English language and technology, it was determined that the following concepts have a high factor load in the first dimension when examined in this study: computer-

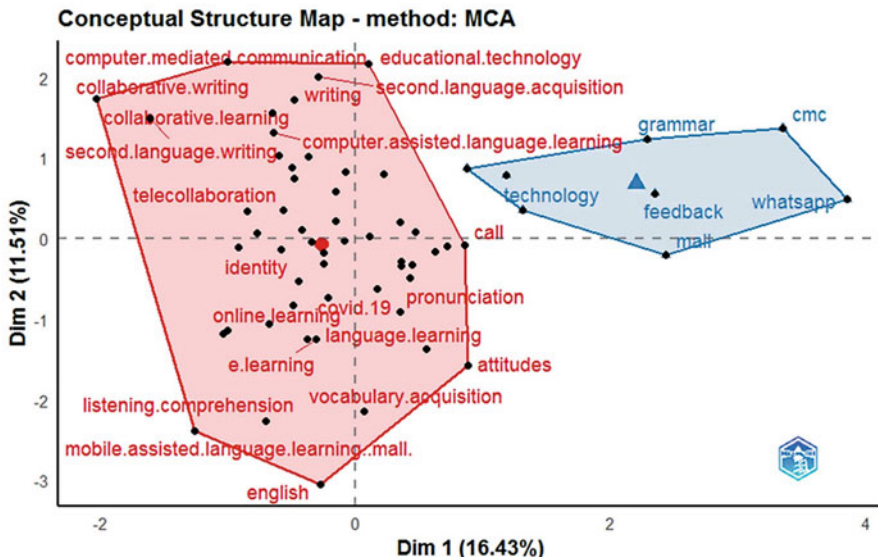


Fig. 41 Conceptual map and keyword clusters

mediated communication, collaborative writing, writing, second language acquisition, collaborative learning, second language writing, computer assisted language learning, telecollaboration, identity, call, online learning, pronunciation, language learning, e-learning, listening comprehension, attitudes, vocabulary acquisition, mobile assisted language learning, and English. Likewise, in the second cluster of keywords, you can find the following terms: technology, grammar, feedback, WhatsApp, mall, and CMC, among others (Fig. 41).

Dendrograms are diagrams that depict hierarchical relationships between objects, illustrated by their hierarchical arrangement. By contrast, hierarchical clustering illustrates the arrangement of the clusters determined by the corresponding analyses through a diagrammatic representation. As the name suggests, a dendrogram is primarily used to determine the best way to allocate objects to clusters based on their characteristics. Furthermore, the distance between the clusters can be seen on the Y-axis of the graph, which is the distance between the clusters, and on the X-axis, there are the subject concepts for which the data points of the clusters are grouped (Fig. 42).

3.7 Intellectual Structure

An intellectual structure is a methodological technique for identifying what authors, documents, or sources have had a major impact on the academic field (Kessler 1963; Small 1973). Several basic concepts describe a field, such as the major themes,

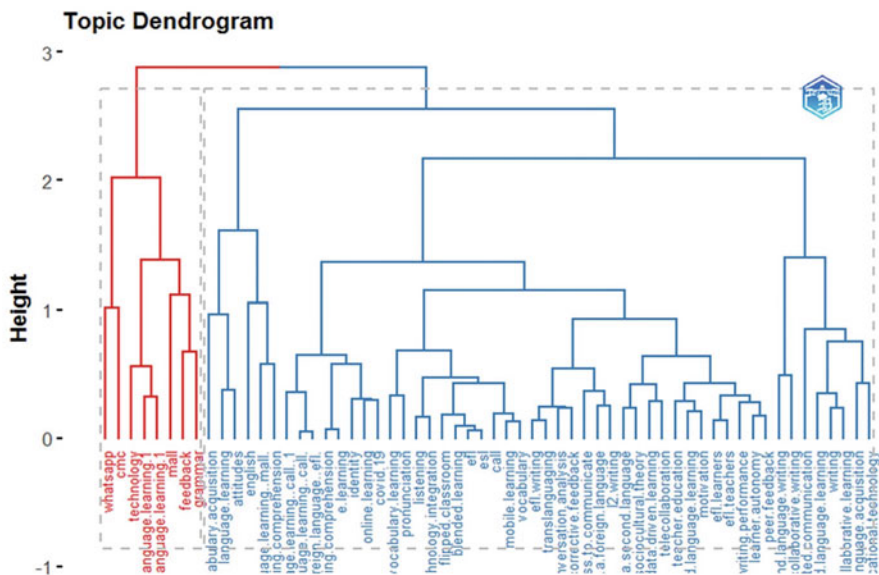


Fig. 42 Topic dendrogram

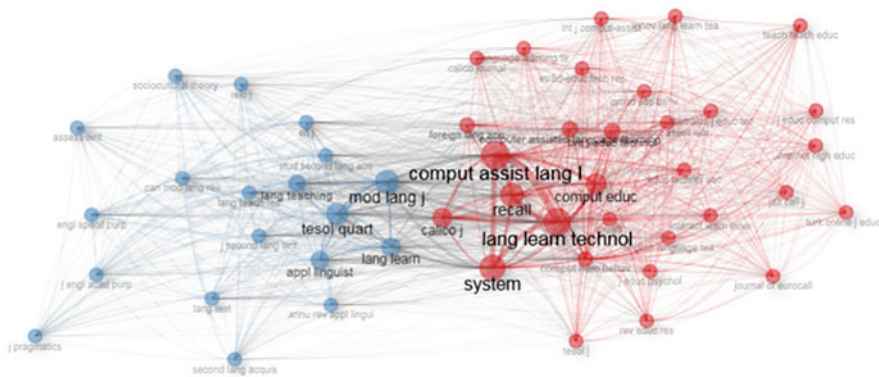


Fig. 43 Co-citation network of sources

subthemes, and patterns that are associated with that field. Co-citation network data shows that the result is a result of the impact of different sources on each other, which is clear from this graph (Fig. 43).

In addition, the degree plot diagram shows the cumulative degree of the sources based on the order in which they appear. Using the co-citation network, it is possible to see how the sources influence one another and how they are interconnected. The degree plot diagram represents the sources' cumulative degree in order of occurrence (Fig. 44).

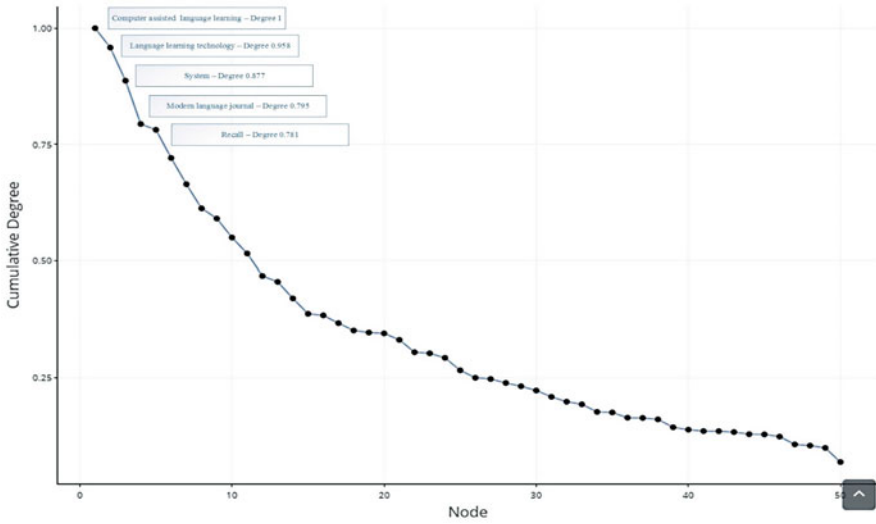


Fig. 44 Degree-plot of co-citation network of sources

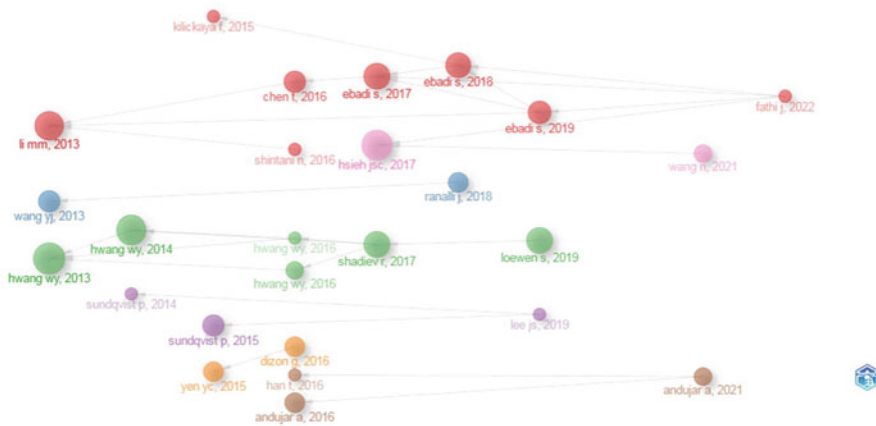


Fig. 45 Historiograph of first authors

Hence, as shown in Fig. 45, the historiograph of the study gives a full view of the document, keyword, and first author’s collaboration, along with the year in which it was completed.

3.8 Social Network

Collaboration networks can be used to describe the social structure at the various levels of scientific cooperation (Glänzel 2002). In the last section, we are able to see

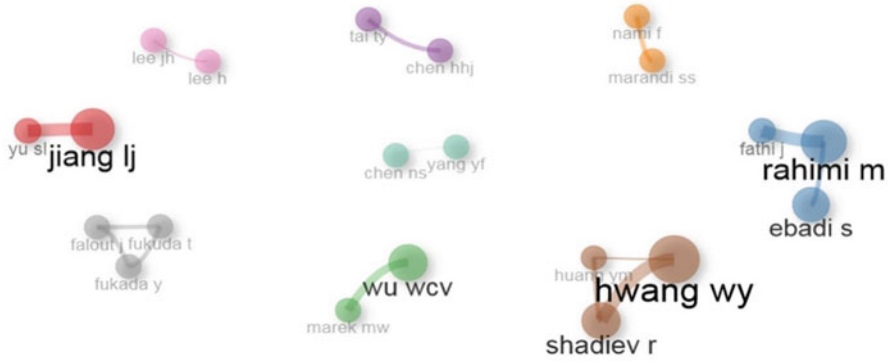


Fig. 46 Collaboration network of authors

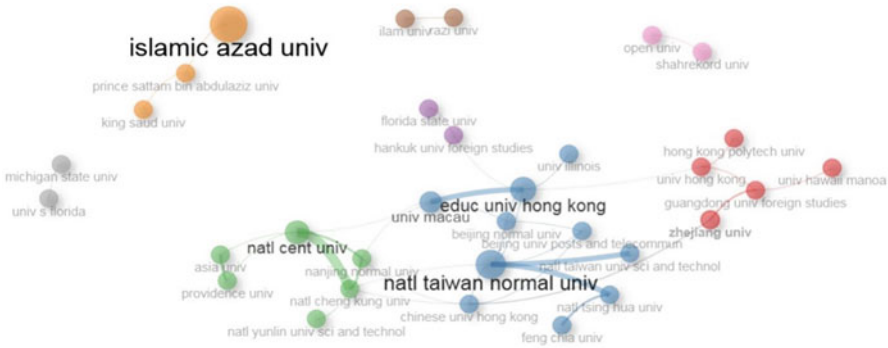


Fig. 47 Collaboration network of institutions

how authors, institutions, and countries collaborate, which illustrates an easier understanding of the issue. China was regarded as the number one country, followed by the USA and Iran. It is worth noting that Islamic Azad University in Iran was considered a leading institution in the world, followed by the National Normal University of Taiwan and Education University of Hong Kong. Among the authors with the most collaborations, HWANG WY, RAHIMI M, JIANG LJ, and WU WCV had the highest collaboration rates, respectively (Figs. 46, 47, and 48).

Additionally, this multi-purpose software allows users to visualize the collaboration between countries on the map. The frequency of the number of publications counted determines the thickness of each curve on the map of international collaborations. Thus, there is a very dense network between the USA and China. In other words, the USA has become CHINA's first international co-authorship partner with 30 frequencies. The United Kingdom became China's second international partner with 8 frequencies (Fig. 49).

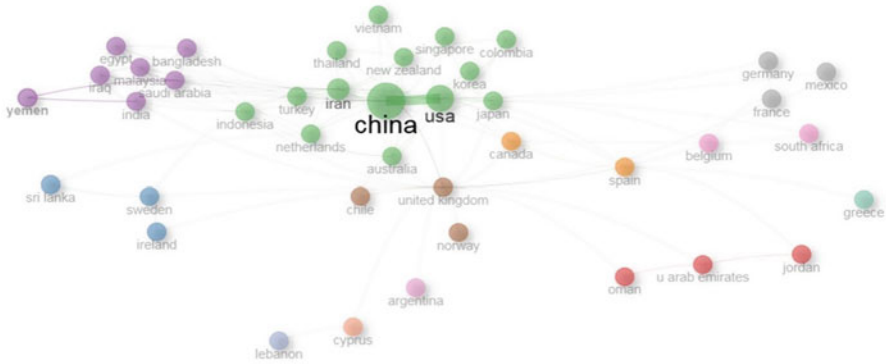


Fig. 48 Collaboration network of countries

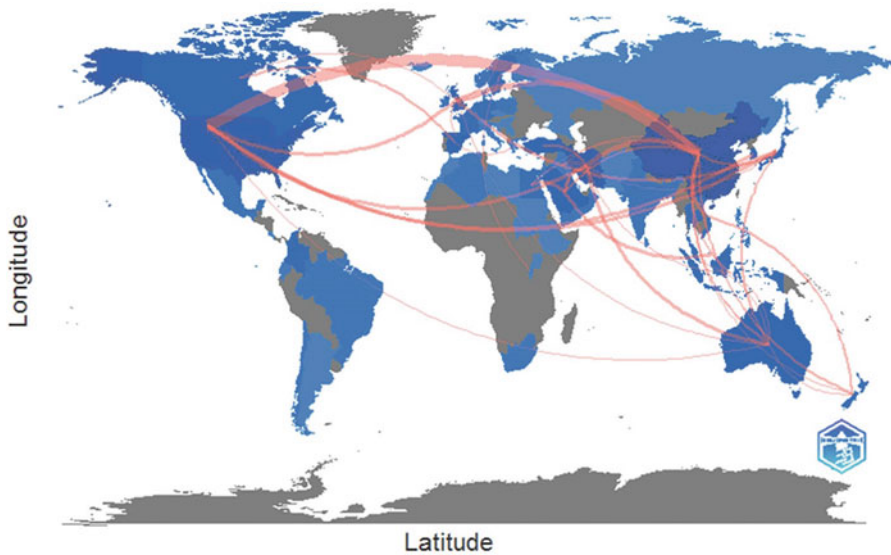


Fig. 49 Countries' collaboration across the world

4 Conclusion

It is very necessary and intriguing for academic professionals to be aware of the most popular research topics, such as books, journals, and other works that have made significant contributions to their field. Hence, scholars are encouraged to stay updated with research trends in their profession to make more educated judgments about the topics they will investigate and remain on top of emerging trends (Lei and Liu 2019). Furthermore, the significance of science mapping as a vital activity is becoming increasingly obvious to researchers working in every field of the scientific

discipline. There is no doubt that the number of publications is expanding at an accelerating rate and that many of these articles are emerging in a vague manner, which complicates the process of knowledge accumulation (Aria and Cuccurullo 2017). Meanwhile, policy and practice rely upon that, and it helps determine the intellectual framework and research frontiers of scientific subjects. Generally, it is common to use specialized software tools to perform only certain steps in a science mapping analysis. In fact, only a few of them are available to scholars that permit them to follow the complete workflow from start to finish. A comprehensive science mapping analysis of scientific literature can be implemented using the open-source software Bibliometrix. It is possible to gain information about a particular work's intellectual structure and conceptual framework due to bibliometric analysis, a data-driven approach to analyzing the literature. In this way, the user can gain insight into the progress of research on certain topics (Zupic and Čater 2015, 9).

Science mapping is becoming an essential activity for scholars of all scientific disciplines. However, in bibliometrics, the use of scientific workflows is still in its early stages. Moreover, it has become increasingly difficult to accumulate knowledge as the number of publications continues to grow exponentially (Aria and Cuccurullo 2017; Guler et al. 2016). Software tools that are specialized in science mapping analysis are commonly used only to perform certain steps of the process. In fact, only a few of these software allow scholars to follow the entire workflow in detail. This article has demonstrated how scientific workflow managers such as Biblioshiny, a powerful tool for managing bibliometric analyses that allows users to integrate online databases, statistical analysis, and data visualization, can facilitate scientific workflow management. Based on the above, identifying the intellectual structure and research frontiers of scientific domains has become vital to research, policy-making, and practice as a whole.

In summary, this study was conducted in order to introduce a bibliometric analysis using the Bibliometrix package and to introduce the Biblioshiny interface, which is easily implemented by the R programming language, and to perform a bibliometric analysis using the Biblioshiny. This study focused on English language and technology-related publications published in the fields of Language and Linguistics, Education and Educational Research. It is worth mentioning that between 2013 and 2022, 881 articles were searched in the WoS database.

References

- Aria, Massimo, and Corrado Cuccurullo. 2017. Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 11 (4): 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>.
- Aria, Massimo, Michelangelo Misuraca, and Maria Sabrina Spano. 2020. Mapping the evolution of social research and data science on 30 years of social indicators research. *Social Indicators Research* 149 (3): 803–831. <https://doi.org/10.1007/s11205-020-02281-3>.

- Briner, Rob B., and David Denyer. 2012. *Systematic review and evidence synthesis as a practice and scholarship tool*. Oxford University Press EBooks. <https://doi.org/10.1093/oxfordhb/9780199763986.013.0007>.
- Broadus, Robert N. 1987. Toward a definition of 'bibliometrics'. *Scientometrics* 12 (5–6): 373–379. <https://doi.org/10.1007/bf02016680>.
- Chen, Chaomei. 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology* 57 (3): 359–377. <https://doi.org/10.5555/1115657.1115670>.
- Cobo, Manuel, Antonio Gabriel López-Herrera, Enrique Herrera-Viedma, and Francisco Herrera. 2012. SciMAT: A new science mapping analysis software tool. *Journal of the Association for Information Science and Technology* 63 (8): 1609–1630. <https://doi.org/10.1002/asi.22688>.
- Diodato, Virgil, and Peter Gellatly. 2013. *Dictionary of bibliometrics*. Routledge. <https://doi.org/10.4324/9780203714133>.
- Eck, Van, Nees Jan, and Ludo Waltman. 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84 (2): 523–538. <https://doi.org/10.1007/s11192-009-0146-3>.
- Glänzel, Wolfgang. 2002. Coauthorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends* 50 (3): 461–473. <https://dblp.uni-trier.de/db/journals/libt/libt50.html#Glanzel02>.
- Guler, Arzu Tugce, Cathelijn J.F. Waaijer, and Magnus Palmblad. 2016. Scientific workflows for bibliometrics. *Scientometrics* 107 (2): 385–398. <https://doi.org/10.1007/s11192-016-1885-6>.
- Harzing, Anne-Wil, and Satu Alakangas. 2016. Google scholar, scopus and the web of science: A longitudinal and cross-disciplinary comparison. *Scientometrics* 106 (2): 787–804. <https://doi.org/10.1007/s11192-015-1798-9>.
- Hubert, J.J. 1980. Linguistic indicators. *Social Indicators Research* 8 (2): 223–255. <https://doi.org/10.1007/bf00286478>.
- Kessler, M.M. 1963. Bibliographic coupling between scientific papers. *American Documentation* 14 (1): 10–25. <https://doi.org/10.1002/asi.5090140103>.
- Lei, Lei, and Dilin Liu. 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>.
- Linnenluecke, Martina K., Mauricio Marrone, and Abhay K. Singh. 2020. Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management* 45 (2): 175–194. <https://doi.org/10.1177/0312896219877678>.
- Marx, Werner, Lutz Bornmann, Andreas Barth, and Loet Leydesdorff. 2014. Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology* 65 (4): 751–764. <https://doi.org/10.1002/asi.23089>.
- Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* 6 (7): e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
- R Core Team. 2014. R: A language and environment for statistical computing. *MSOR Connections* 1 (1) <https://www.r-project.org/>.
- Şenel, Emrah, and Emre Demir. 2018. Bibliometric and scientometric analysis of the articles published in the journal of religion and health between 1975 and 2016. *Journal of Religion & Health* 57 (4): 1473–1482. <https://doi.org/10.1007/s10943-017-0539-1>.
- Shi, Jiangang, Kaifeng Duan, Wu Guangdong, Rui Zhang, and Xiaowei Feng. 2020. Comprehensive metrological and content analysis of the public-private partnerships (PPPs) research field: A new bibliometric journey. *Scientometrics* 124 (3): 2145–2184. <https://doi.org/10.1007/s11192-020-03607-1>.
- Small, Henry. 1973. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science* 24 (4): 265–269. <https://doi.org/10.1002/asi.4630240406>.

- Visser, Martijn S., Nees Jan Van Eck, and Ludo Waltman. 2021. Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft academic. *Quantitative Science Studies* 2 (1): 20–41. https://doi.org/10.1162/qss_a_00112.
- Ware, Mark, and Micheal Mabe. 2015. *The STM report: An overview of scientific and scholarly journal publishing*. International Association of Scientific, Technical and Medical Publishers. https://www.stmassoc.org/2009_10_13_MWC_STM_Report.pdf.
- Zupic, Ivan, and Tomaž Čater. 2015. Bibliometric methods in management and organization. *Organizational Research Methods* 18 (3): 429–472. <https://doi.org/10.1177/1094428114562629>.

Babak Daneshvar Ghorbani is an MA holder in Applied linguistics from Iran University of Science and Technology. His research interests primarily lie on meta-research, quantitative studies, scientometric and bibliometric analysis, argumentative writing, and computer assisted language learning (CALL). He has published in different journals, including *System*, *Mextesol*, *International Journal of Technology in Education*, and *BMC Public Health*. He has presented at international conferences such as AERA 2023 (American Educational Research Association) and the 2022 Mixed Methods International Research Association conference.