REVIEW

The Role of Corpus Linguistics in Information and Communication Technologies Education: A Bibliometric Analysis of Research Trends

Nilufar Barnoyeva [1*] , *Gulnoza Sabirova* [2] , *Dilnoza Radjabova* [1] , *Xilola Umarxo'jayeva* [1] , *Gulchekhra Razzokova* [1] , *Sitora Daniyarova* [1] , *Dilafruz Saidvalieva* [1]

[1] Foreign Languages Department, Tashkent University of Information Technologies named after Mukhammad al Khwarazmi, Amir Temur Street, Tashkent 100084, Uzbekistan

[2] Tashkent Institute of Irrigation and Agricultural Mechanization Engineers, National Research University, Kari Niyaziy 39 str, Tashkent 100000, Uzbekistan

ABSTRACT

By improving linguistic analysis and digital literacy, corpus linguistics' incorporation into information and communication technologies (ICT) education empowers students to use standardized research techniques across languages. Thanks to ICT tools, students become active researchers instead of passive learners, encouraging linguistic inquiry. This multidisciplinary method updates corpus-based research for worldwide relevance by utilizing automation and data analysis. Using the Web of Science and Scopus databases, this study performs a bibliometric analysis of publications on corpus linguistics in ICT education from 2004 to 2024. VOSviewer, Excel, and CSV files were used for data extraction, screening, and analysis. The study provides information about the knowledge structure of the field by identifying publication trends, essential authors, journals, and keyword clusters. The study shows that corpus linguistics research among ICT students follows a cyclical pattern, with Web of Science peaking in 2014 and Scopus peaking in 2018. Publications from early research (2004–2010) were few, but after 2010, interest increased. For 20 years, key terms have not changed, highlighting a long-term scholarly focus. Research trends are heavily influenced by institutional structures, disciplinary traditions, and ICT adoption; this emphasizes the necessity of customized ICT policies in corpus linguistics education. Based on cyclical

CORRESPONDING AUTHOR:

Nilufar Barnoyeva, Foreign Languages Department, Tashkent University of Information Technologies named after Mukhammad al Khwarazmi, , Amir Temur Street, Tashkent 100084, Uzbekistan; Email: nilufar.barnoyeva@mail.ru

publication patterns in Web of Science and Scopus, the findings show a shift from a theoretical focus to ICT integration. The study emphasizes how interdisciplinary approaches are necessary to improve the sustainability and accessibility of corpus linguistics in ICT education.

Keywords: Bibliometrics; Corpus Linguistics; Educational Technology; ICT Education; Linguistic Analysis

1. Introduction

Corpus Linguistics (CL) is one of the approaches in modern linguistics that is expanding the fastest[1]. CL fosters an excellent knowledge of human language and computational models that can mimic, process, and analyze linguistic data by bridging the gap between linguistics and computer science. CL is a fast-emerging discipline of linguistics that deals with creating and analyzing corpora (Eng. corpussing, corpora- pl). It emerged due to the increasing demand for computer technology to cope with enormous arrays of language data[2]. Numerous classifications within the discipline resulted from the extensive scientific investigation made possible by this computerized phase. For linguistic analysis and study, a "corpus" (plural: corpora) is a substantial, systematic set of texts or speech data[3]. Our teaching methodology, combining theory, problem-solving, and independent research, shows that students can effectively apply standardized methods across multiple languages despite the traditional perception of corpus linguistics as language-dependent[4].

Information and Communication Technologies (ICT) are crucial for developing nations to progress and compete globally because they promote global connectivity, draw investment, increase economic activity, and advance society[5]. Many educators utilize ICT to supplement conventional teaching strategies, such as information retrieval, in which students are "passive learners of knowledge instead of active producers able to participate in the learning process." Galea (2002) describes how ICT can support teaching and learning in a document titled Teaching and Learning with ICT. She claims that there are two primary causes for the rise in ICT use in UK classrooms. First of all, ICT can alter the tempo of the lessons. She said that today's kids must grow up with the potential and abilities necessary to fully benefit from the new opportunities ICT presents. "Secondly, academic researchers in the UK are very interested in improving the quality of teaching and learning in schools so that students

can attain better results."

New technologies have also been shown to benefit students in many ways[6].

A new strategy that improves linguistic analysis and digital literacy is the incorporation of corpus linguistics into ICT education. ICT tools have helped corpus linguistics, once thought to be language-dependent, by enabling students to use standardized research techniques in various languages. ICT students go beyond passively consuming knowledge to actively participating in linguistic research thanks to this interdisciplinary connection, which promotes active learning. Language research is made more dynamic, scalable, and applicable in global contexts by utilizing ICT-driven automation, data analysis, and digital resources to improve corpus-based studies.

2. Literature Review

According to G.R. Bennett, a corpus is a collection of written or spoken texts that support linguistic study and allow for examining language in use[7]. To explore the frequency and patterns of language usage, McEnery states that a corpus is a sizable and organized collection of texts that are typically kept and analyzed electronically[8]. Many types of corpora, corpus techniques, and the analytical tools that support corpus analysis are created and utilized by qualified specialists and analysts who work for academic institutions and/or publishing houses (e.g., the multi-billion-word Cambridge English Corpus (CIC), which has restricted access and is primarily accessible by colleagues in the Cambridge University Press)[9]. Due in part to interactions with other fields like corpus linguistics and language documentation, the management and storage of sociolinguistic data have received more attention in recent years. The methods developed in corpus linguistics are rarely used by sociolinguists in their research, even though corpus linguistics and sociolinguistics share similarities regarding epistemology, focus, and purpose[10]. The perspective of corpus linguistics can

offer sociolinguists a fresh look at the use of big databases as well as assistance in resolving issues related to data management and storage[11]. However, ensuring high-quality primary data is appropriately recorded and stored, with sustainability and accessibility taking precedence over immediate research interests, is essential to building a corpus from language documentation[12]. Despite extensive annotation schemes developed by corpus and documentary linguists, they often overlook sociolinguistically relevant details such as informants' social and demographic data or interactional context, making it essential to integrate these aspects and maintain links between sound recordings, linguistic annotations, and rich metadata in database creation[13]. By P.V. Sysoev, one should comprehend "the main characteristics and signs, distinguishing some information technologies from others" under the didactic characteristics of ICTs and "external demonstration of ICT used in the educational process to achieve the set goals" under the methodological functions[14]. To develop a methodology for teaching English collocations in ICT, it is necessary to consider CL's didactic characteristics and methodological functions. The issue of selecting and representativeness of the analyzed material is undoubtedly vital in all linguistic research. The concept of "corpora of texts" (CT) has recently become more and more integrated into linguists' scientific discourse, serving as the foundation for the development of CL[15]. One of the term's founders, John Sinclair, describes corpus linguistics (CL) as a technique that uses vast databases of authentic texts to methodically analyze language data to observe language in context and comprehend linguistic patterns in everyday usage[16]. The systematic and empirical study of language based on the analysis of large corpora allows researchers to explore linguistic phenomena through authentic language use and to apply these insights to various fields, including language teaching and linguistic theory, according to V.V. Rykov, despite the linguistics research of other scientists[17]. Although collocations are widely acknowledged to play a critical role in L2 English learning and instruction, they are still not adequately included in instructional materials and receive little attention in the classroom[18]. Nesselhauf claims that not enough research has been done on collocation instruction and the challenges faced by L2 learners[19]. The application of concordance for the following objectives: examining word meanings in multiple contexts and contrasting

collocations[20]. However, Souza Hodne claims that comparatively few studies have linked textbooks, teaching, and corpora[18]. Incorporating corpora into L2 instruction is necessary, according to Chambers, who confirms that CL consultation will eventually replace textbooks[21]. The study of lexical compatibility or collocations in context, where the term "collocation" is frequently used with a more meaningful and straightforward concept than in traditional linguistics, is made possible by corpus linguistics (CL), which researchers refer to as statistical[22].

Bibliometrics is an analytical method used in a variety of settings[23]. It has become increasingly common in scientific research in recent years[24]. Bibliometric analysis offers a trustworthy way to map the scientific evolution of data, collaborate with authors and nations, and rank[13]. Bibliometric reviews seek to provide a comprehensive overview of the literature rather than assess the calibre of the study or defining words[25].

The main subjects covered are the history, ideas, methods, and annotation strategies for creating and utilizing electronic text corpora, along with comparisons of different corpus search tools[4]. Many linguists resorted to corpus data because they believed that data gathering must involve more than academics making intuitive acceptability judgments about what people may and cannot say[1]. Concordances are crucial to linguistic studies because they provide detailed information on how words are employed and in what situations within CL. They enable researchers and users (students and teachers) to examine linguistic phenomena, identify patterns, and form perceptive opinions on language structure and usage. As mentioned, concordance is akin to linguistic search engines, and corpus managers have a similar function. This section examines what the term "corpus manager" means. "A specialized text management system then presented in a user-friendly format" is how V.P. Zakharov defines a corpus manager[26]. Teachers may construct more effective, inclusive, and engaging learning environments that embrace pedagogical foundations that encourage students' success[27].

3. Materials and Methods

For this review, we looked for publications on regional knowledge from the research conducted. Using Corpus Linguistics in ICT students as the keyword, the study was con-

ducted on the most popular bibliographic online databases, Scopus and WOS Base, for 2004–2024. A CSV file, Microsoft Excel 2021, RIS, VOS, and a viewer were used to design the analysis.

3.1. Article Review and Study Eligibility Criteria

Relevant data were entered into a spreadsheet for the search process, including the keyword "Corpus Linguistics in ICT" and all English-language articles. The article type is "article," the period is 2004-2024, the subject area is "Corpus Linguistics in ICT," and the document type is "article." The subject area is "Scientific Science, Computer Science, Education, Linguistics, Multidisciplinary." The flow of the chosen research methodology is depicted in **Figure 1**.
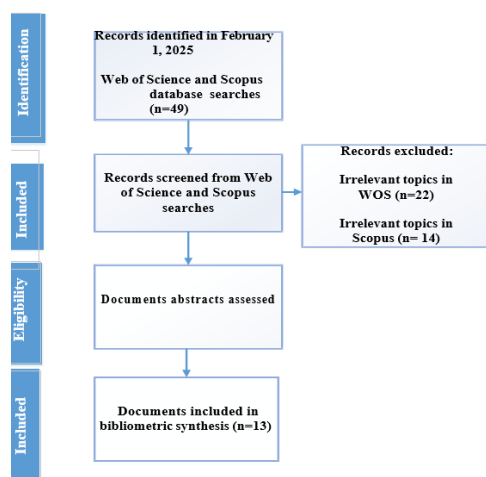

Figure 1. Data Identification Flow Chart.

During the screening process, the following exclusion criteria were used:

1. The title, abstract, and review of the article are reported in English
2. Articles related to other research areas
3. Lack of definition of search terms (Corpus Linguistics, ICT students).
4. Many articles do not have a DOI, and their search ability is limited. In general, excluding these articles using the Scopus filter options was impossible.

3.2. Bibliometric Analysis

A bibliometric analysis was conducted after the CSV data was imported into Excel. Before beginning the studies,

the data were carefully examined for mistakes. Following an analysis of the reviewed articles, the most pertinent ones were found, along with the corresponding authors, who produced the most significant number of articles. The search's articles were evaluated and categorized based on a variety of factors, including the number of papers published annually, the type of document, the top list of papers, the top list of journals, the top list of funding sponsors, the distribution of the articles by subject categories and periodicals, and affiliation by nation and institution. Lastly, by finding clusters of the most prevalent keywords in the literature, the co-authors and co-occurrence of keywords were examined to investigate the knowledge components and structure of the research domain.

4. Results

4.1. Trend of Publications on Corpus Linguistics in ICT students

In many educational areas, the impact of Corpus Linguistics on ICT students has a wide range of scientific implications. A total of 49 papers were published between 2004 and 2024 on Corpus Linguistics in ICT students (**Figure 2** and **Figure 3**).
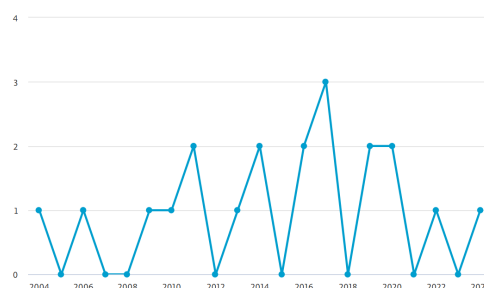

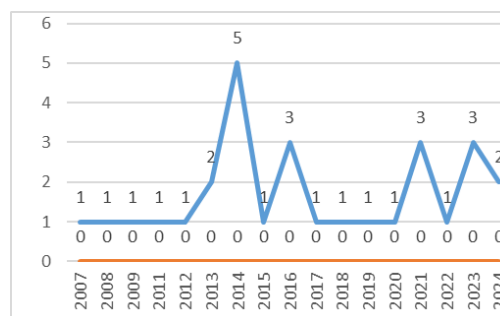**Figure 2.** The period of articles on Corpus Linguistics is in ICT students in Scopus base.

Figure 3. The period of articles on Corpus Linguistics in ICT students in WOS base.

Figure 2 shows the distribution of corpus linguistics articles about ICT students in the Scopus database from 2004 to 2024. Only one or fewer articles were published annually between 2004 and 2010, indicating low publication frequency. Following 2010, there has been a steady upward trend with sporadic increases and minor declines. The peak years for activity are 2012 and 2014. With the most articles published (three) in 2018, the publication frequency peaked in 2018. Following 2018, the quantity of articles varies once more, showing a discernible decrease. With only one article published in some years, the publication levels off at lower levels. Finally, this data demonstrates a range of interest in the subject, with a peak in research between 2016 and 2018 and a subsequent decline. The Web of Science (WOS) database's publication trends for corpus linguistics articles relevant to ICT students from 2007 to 2024 are highlighted in **Figure 3**. One article is consistently published annually. With two articles, there is a discernible increase beginning in 2013. 2014 saw the highest peak with five articles, indicating a significant increase in research interest. In 2015, the number fell to zero after 2014. With three articles, there was a slight rebound in 2016, after which there were yearly variations between one and zero articles. The publication rate was maintained after the publication of three articles in 2020, 2022, and 2023. There was a minor decline with two articles in 2024.

According to the data, research significantly increased in 2014, but it then declined and stabilized in the years that followed. Publications have been consistent since 2019, indicating a moderate but ongoing interest in this field of study.

4.2. Publication Types of AI in Academic Writing

The Scopus and WOS databases' distributions of various publication types are contrasted in **Table 1**. While roughly the same book chapters are in both databases, WOS has more articles and conference papers overall. In particular, Scopus contains only five articles, a substantial decrease from the 16 articles found in WOS. WOS also has 14 conference papers, while Scopus only has 13. Curiously, there is one book in Scopus that is absent from WOS. Only one "Early Access" publication listed by WOS is not present in Scopus. One chapter of a book is included in each database. The two sources' publication coverage differs slightly, as this

comparison demonstrates.

Table 1. A list summarizing the distribution of publication types of Scopus and WOS.

Scopus	Numbers	WOS	Numbers
Article	5	Article	16
Conference Paper	13	Conference Paper	14
Book	1	Early Access	1
Book Chapter	1	Book Chapter	1

4.3. Authors and Their Affiliated Country

Our research found that 379 authors studied AI in academic writing from 2004 to 2024. **Figures 4** and **5** show 10 authors who have published more than two papers. Among them, Nazim, M, Mohammad, T, Khan, S reigned with four publications, followed by Alzubi, A with three, Chaka, C, Bernik, A, Becker, B, Barrot, J, Ayyoub, A, and Alafnan, M with two publications. This list of the top three authors comes from Saudi Arabia.
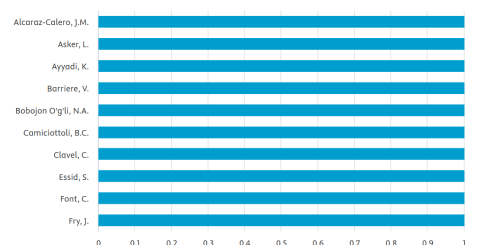

Figure 4. The list of top authors is published in Figure Corpus Linguistics in ICT students in Scopus base.
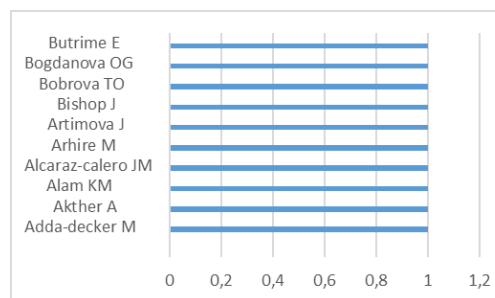

Figure 5. The List of top authors published in Corpus Linguistics in ICT students in WOS base.

Institutions are classified according to the quality of the articles they publish. From 2004 to 2024, seventy-seven different institutions cooperated to publish papers related to corpus linguistics in ICT students. Our analysis of the top 10 institutes' publications on corpus linguistics in ICT students aims to determine the influential and productive institutions in this field. As indicated in Scopus based in

Figure 6, the institutions include universities from different countries, such as The Hong Kong Polytechnic University, NIWI-KNAW, State Grid Information and Telecommunication Branch, Universidad de Murcia, Chinese University of Hong Kong, Al-Imam Muhammad Ibn Saud Islamic University, Lomonosov Moscow State University, University of Zagreb, Norges Teknisk-Naturvitenskapelige Universitet, Huizhou University. Hong Kong University published two papers compared to other universities, but other universities indicated the same percentage, which published one paper. In the WOS base in **Figure 7**, the institutions include universities from different countries, such as SINTEF, Beijing University of Technology, National De La Recherche Scientifique CNRS, Charles University Prague, Chung Ang University, Ctr Res Online Communities E-Learning Syst, Dongguk University, Drexel University, Ecole Polytechnique, and FCCN. All the institutions have published one paper. However, SINTEF published two papers. Both bases have the same indication while analyzing affiliations.
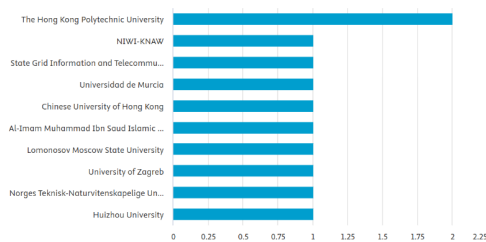

Figure 6. List of top institutions in Corpus Linguistics, in ICT students in Scopus base.
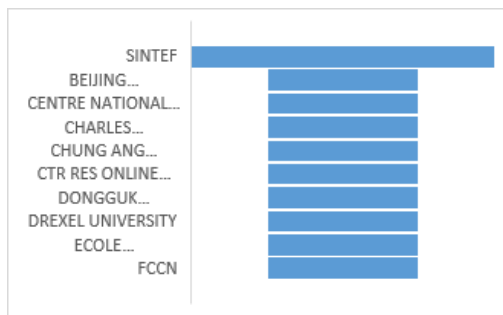

Figure 7. List of top institutions on Corpus, Linguistics in ICT students in WOS base.

4.4. Top Countries on AI in Academic Writing

The number of publications in the ten most productive countries in the field of AI in academic writing between 2004 and 2024 (see **Figure 8** and **9**). Among them, the USA dominated with 94 publications, followed by China 32, UK 31, India 22, Saudi Arabia 21, Australia 17, Malaysia and Indonesia 13, Canada 12, and Russia 7.
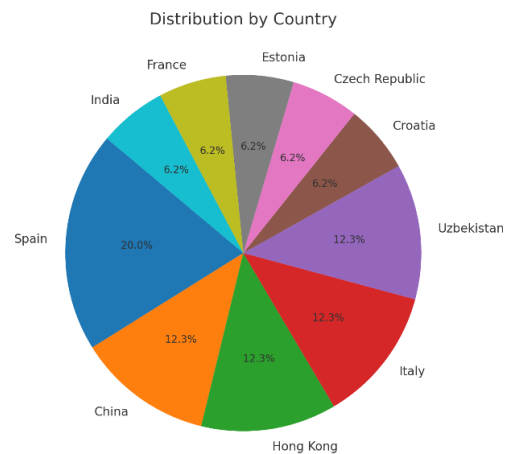

Figure 8. List of top countries on Corpus Linguistics, in ICT students in Scopus base.
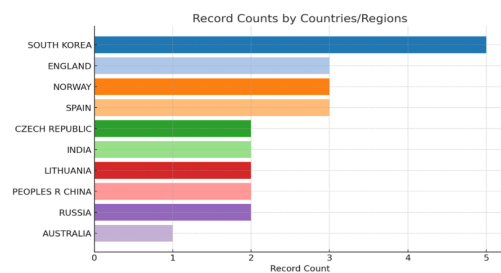

Figure 9. List of top counties on Corpus, Linguistics in ICT students in WOS base.

4.5. Top Co-authorships and Keywords on AI in Academic Writing

VOSviewer can generate co-authorship, keyword co-occurrences, citations, bibliographic coupling, and co-citation maps based on bibliographic data. File formats supported include .txt, ris, and .csv from databases such as WOS. The raw file was imported into VOSviewer, and the software created a map of co-authorship and keyword co-occurrences (shown in **Figure 7** and **Figure 8**). The co-authorship analysis resulted in a network of 84 authors. There are 84 items distributed over 28 clusters: Total Links (129).

The analysis yielded 153 keywords. After excluding the general keywords with a low relevance score and those with low occurrence (by default, a minimum of 121 occurrences of a keyword is selected to strengthen the co-occurrence re-

sults), 121 items were finally identified. Based on the total link strength, each resulting keyword is sketched in a node, creating a network map of all keywords. **Figure 8** shows the network map of the top 10 authors' keyword co-occurrence. The size of the node reflects the keyword's degree of importance. There are 121 items distributed over 19 clusters: Total link strength (325) and links (324).

Figure 10 and Figure 11 show two separate but related visualizations created with VOSviewer, a program for building and visualizing bibliometric networks. These numbers reveal the organization of academic collaboration and topic trends in research throughout time. In particular, **Figure 11** shows a network of top keywords based on total link strength, while **Figure 10** maps top co-authorships based on total link strength.
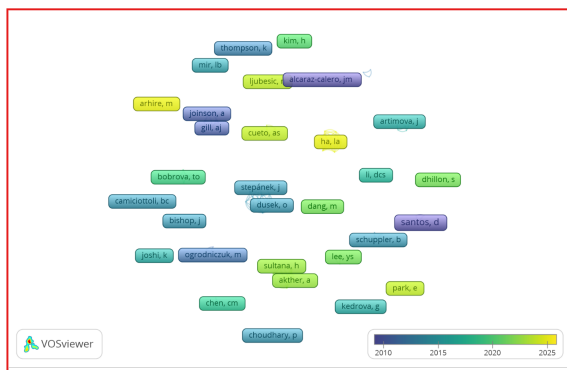

Figure 10. Network map of top co-authorships based on the total link strength.
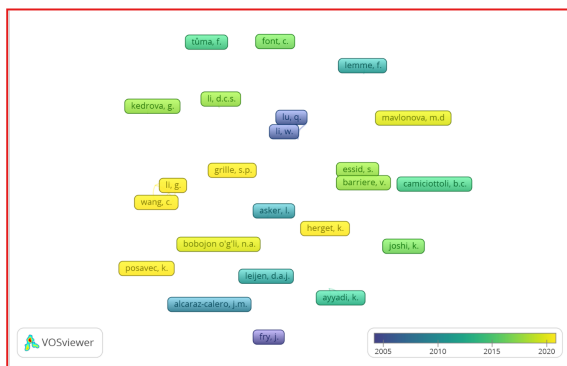

Figure 11. Network map of top keywords based on the total link strength.

The most crucial co-author relationships between scholars are depicted in **Figure 10**. The hues, which range from bright yellow (more recent years, about 2025) to deeper blue (older years, like 2010), show the average publishing year, while the nodes stand in for individual writers. Larger nodes

typically represent authors with greater collaboration (higher total connection strength). In contrast, **Figure 11** maps the most frequently occurring phrases in the dataset and concentrates on keywords rather than individuals. A term or an author often linked to particular study areas is represented by each node. The color gradient depicts the time dimension, as in the preceding illustration, with brighter tones signifying more recent usage. This picture illustrates cooperation and how research subjects have changed over time.

The structure also shows a significant difference: the keyword network (**Figure 11**) is more scattered, suggesting a greater variety of subject diversity, but the co-authorship network (**Figure 10**) is more packed, representing tighter groups. This discrepancy implies that study subjects could be more varied and related, even though collaborative interactions often create closer communities.

Thanks to the dynamic dimension provided by the temporal color gradient in both maps, readers may follow the development of partnerships and themes. Thus, recent patterns in research and cooperation can be identified, which is essential for determining a field's future course.

VOSviewer can generate co-authorship, keyword co-occurrences, citations, bibliographic coupling, and co-citation maps based on bibliographic data. File formats supported include .txt, ris, and .csv from databases such as Scopus. The raw file was imported into VOSviewer, and the software created a map of co-authorship and keyword co-occurrences (shown in **Figure 7** and **Figure 8**). The co-authorship analysis resulted in a network of 37 authors, with a total of 25 links.

The analysis yielded 38 keywords. 38 items were identified. Based on the total link strength, each resulting keyword is sketched in a node, creating a network map of all keywords. **Figure 8** shows the network map of the top 10 authors' keyword co-occurrence. The size of the node reflects the keyword's degree of importance. There are 38 items distributed over 4 clusters: Total link strength (255) and links (247).

Visualizations of keyword networks based on corpus linguistics research are shown in **Figure 12** and **Figure 13.** Both images provide insights into the topic's intellectual structure and temporal growth by using VOSviewer to show the correlations and trends among often occurring terms. Each node represents keywords, and the linkages show how

strongly co-occurrence associations exist. The color gradient makes a temporal perspective on research trends possible, which commonly ranges from purple/blue to yellow/green and corresponds to the average publication year of the articles in which the keywords appear. A network diagram of the most popular terms according to total link strength is shown in **Figure 12**. The range of the temporal gradient is from 2010 to 2020. The conspicuous placement of essential terms like "linguistics," "corpus linguistics," and "corpus" close to the centre indicates their fundamental function in the area. The green-yellow spectrum of keywords like "computational linguistics," "digital storage," and "IT (information technology)" indicates that they have just recently emerged or have become increasingly relevant. Conversely, blue and purple-colored phrases like "peer reviews," "syntactical structure," and "academic writings" are grouped, suggesting topics that were more common in past research.
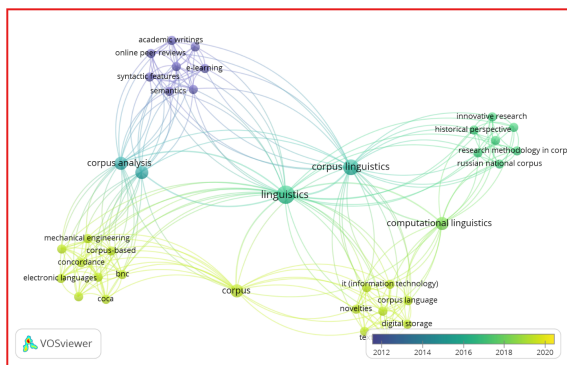

Figure 12. Network map of top co-authorships based on the total link strength.
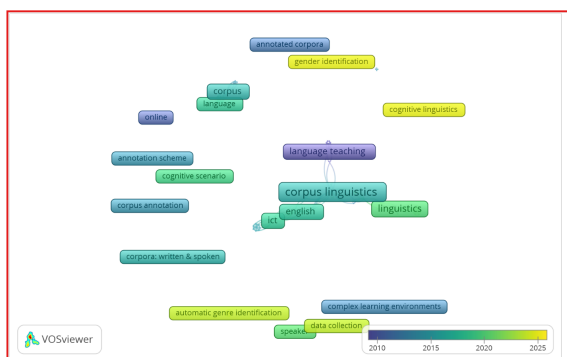

Figure 13. Network map of top keywords based on the total link strength.

On the other hand, **Figure 13** presents an alternative network map of the most popular terms. It is likewise based on the total link strength but shows a more distinct temporal history that extends into the mid-2020s. Words like "lin-guistics," "corpus linguistics," and "English" continue to be essential, reaffirming their fundamental position.

Figure 12 and **Figure 13** show the rapid development of corpus linguistics as a subject of study. While **Figure 13** offers an overview of current and emerging trends, **Figure 12** highlights the historical development and fundamental connections between important concepts. These representations point to the evolution of corpus linguistics from a mostly academic and descriptive tool to a diverse area that now includes computational techniques and practical language research.

5. Discussion

This study aimed to identify the academic disciplines and scientific publications that have had the most significant influence on corpus linguistics among ICT students by using bibliometric analysis to examine a variety of data on the topic. According to this bibliometric study, the best research on corpus linguistics in ICT students was found in publications published more than 20 years ago. The amount of change in the key terms associated with corpus linguistics among ICT students between 2004 and 2024 is demonstrated by the annual production of articles. The introduction stage shows that ICT students' interest in corpus linguistics started with the description of the approach and had only a passing connection to technology. The approach's requirements are evident, and the keywords have been used for the past 20 years.

Publication trends within major academic databases like Scopus and the Web of Science (WOS) show that the study of corpus linguistics in the context of ICT students has received varied attention over the years. The Scopus database shows a cyclical pattern in publications about corpus linguistics for ICT students between 2004 and 2024. There is little activity in the first few years (2004–2010), with hardly more than one article published annually. After 2010, there is a discernible increase in research output, with sporadic peaks in 2012 and 2014. A distinct pattern in the publication of corpus linguistics articles for ICT students can be seen in the Web of Science (WOS) database. Several significant distinctions between the two databases become apparent when comparing them. First, Scopus reached its peak activity in 2018, while WOS showed a more noticeable peak in 2014.

Applying information communication tools in linguistics can benefit language corpora analysis and its theoretical value through concordance-based searches in modern e-corpora such as BNC and COCA[28]. However, according to the findings of Jenny Fry's case study, each academic field's distinct cultural identity determines how Information and Communication Technologies (ICTs) are adopted in that field. Different disciplines have intellectual and social traditions that influence how they use ICTs, and knowledge is fragmented and diverse. Several factors, including task uncertainty, the need for collaboration, institutional structures, and field-specific traditions, influence the ICT system's suitability for a given academic community. She also indicates that the fields that do not adhere to the "big science" model of knowledge production may suffer if these cultural differences are not considered when creating ICT policies and infrastructures[29]. In addition to certain factors found in earlier research, Leijen[30] discovers that the language peers use during the feedback process impacts the efficacy and utility of asynchronous peer feedback using ICT tools. Corpus linguistics techniques are being investigated as a possible tool for examining these interactions.

Eventually, the results show that the use and efficacy of ICTs in academia vary by context, including linguistic factors, cultural identity, and disciplinary traditions. Field-specific needs, institutional structures, and communication styles influence the adoption and impact of ICTs, even though they improve language corpora analysis and peer feedback processes. This emphasizes the importance of considering these contextual differences when creating ICT policies and infrastructures.

Concordances are important in linguistic study because they offer in-depth information on how words are used and what contexts within CL. They enable scholars and users (students and instructors) to investigate linguistic occurrences, spot trends, and make insightful judgments on language usage and structure. As previously indicated, concordance is comparable to linguistic search engines, and corpus managers serve a similar purpose. This section investigates the definition of the word "corpus manager." A corpus manager is "a specialized text management system presented in a user-friendly format" by V.P. Zakharov. When teaching English collocations, a corpus manager can provide several pedagogic features that improve students' learning. A corpus manager is a program or application used for text search and retrieval that enables users to search, examine, and modify linguistic data from a corpus.

The methodical function and didactic qualities of employing CL are among the most promising, with significant potential for guiding the technique of teaching foreign languages to enhance students' collocational awareness. Due to CL's usability, students may study collocations in the context of ICT in English language instruction, analyze vast amounts of text, and spot trends in lexical unit use. Language research is expected to become more real and valid when students perform their studies and make their own "discoveries" while learning a foreign language. English language teachers received a new educational tool for students' collocational awareness, and this new tool should be actively used along with traditional ones.

Integrating Corpus Linguistics into teaching English collocations to ICT students effectively enhances their collocation awareness. This method leverages real-life language use, providing students with authentic examples and contextualized learning opportunities crucial for mastering collocations. The didactic characteristics (authenticity, relevance, variety of contexts, independent learning, CL integration, date-driven exploration) and methodological functions (enhancement of students' understanding and usage of collocations, focus on collocations to the student's field of study, professional interests, combination with other words to convey specific meanings, promotion self-directed learning) of Corpus Linguistics significantly contribute to the promotion of collocation awareness of ICT students. By utilizing Corpus Linguistics data analysis and language learning tools, students can identify and understand the subtle examples of collocations in their field of study, ultimately enhancing their language skills and communication abilities.

6. Conclusion

This study examines the scholarly influence, keyword dynamics, and publication trends over the last 20 years to demonstrate the changing role of corpus linguistics in ICT education. According to the findings, the field has changed due to the growing integration of ICT tools, whereas early research mainly concentrated on theoretical aspects. The cyclical Web of Science and Scopus publication patterns in-

dicate variable but consistent scholarly interest. To improve corpus linguistics' application and accessibility in ICT education, the study also emphasizes how important it is to match it with technological developments. To ensure sustainable and contextually relevant applications, future research should further investigate interdisciplinary approaches to integrate ICT tools in corpus linguistics.

This study offers a bibliometric analysis of corpus linguistics in ICT education, highlighting thematic shifts and long-term publication trends spanning two decades. This study fills the gap by analyzing how corpus linguistics has changed among ICT students specifically, in contrast to earlier research that concentrated on the influence of ICT on linguistics. An additional dimension is provided by the comparative study of Web of Science and Scopus publication trends, which includes information on the cyclical nature of research output in this area and disciplinary influences.

A very successful and cutting-edge educational strategy is the incorporation of corpus linguistics (CL) and corpus management into English language training, especially when teaching collocations to ICT students. Students can investigate and evaluate lexical patterns in relevant contexts using concordances and corpus tools, which offer genuine, data-driven insights into language use in everyday situations. This method encourages a better comprehension of how language works in specific academic and professional domains and supports autonomous and discovery-based learning. Teachers may significantly improve students' collocational awareness and, eventually, their overall language proficiency and communicative efficiency in academic and professional contexts by utilizing CL's methodological and pedagogic qualities.

Author Contributions

N.B.: Conceptualization, Methodology, Investigation. G.S.: Supervision, Validation. D.R.: Visualization. X.U.: Writing – Original draft preparation. G.R.: Software. S.D.: Writing – Reviewing and Editing. D.S.: Resources. All authors have read and agreed to the published version of the manuscript.

Funding

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data that support the findings of this study are available from the Scopus base.

Acknowledgments

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] Gries, S.T., 2009. What is Corpus Linguistics? Language and Linguistics Compass. 3(5), 1225-1241. DOI: https://doi.org/10.1111/j.1749-818X.2009.00149.x

[2] Podlesskaya, V.I., 2011. Modern Computer Methods in the Study and Teaching of Linguistic Disciplines: Corpus Linguistics. In Proceedings of the 18th PCSF 2018 - Professional Culture of the Specialist of the Future. Russian State University for the Humanities: Moscow, Russia.

[3] Stefanowitsch, A., 2020. Corpus Linguistics: A Guide to the Methodology. Textbooks in Language Sciences 7. Language Science Press: Berlin, Germany.

[4] Stoykova, V., 2014. Teaching Corpus Linguistics. Procedia - Social and Behavioral Sciences. 143, 437-441.

DOI: https://doi.org/10.1016/j.sbspro.2014.07.513

[5] Belloumi, M., Alharethi, A.S., 2024. Impact of ICT and financial development on environmental quality in Saudi Arabia. Energy Nexus. 100359. DOI: https://doi.org/10.1016/j.nexus.2024.100359

[6] OECD Centre for Educational Research and Innovation, 2005. E-Learning in tertiary education: Where do we stand? OECD Publishing: Paris, France.

[7] Bennett, G.R., 2010. Using Corpora in Language Learning Classroom. Corpus Linguistics for teachers. Cambridge University Press: Cambridge, UK. DOI: https://doi.org/10.3998/mpub.371534

[8] McEnery, T., Wilson, A., 2001. Corpus Linguistics: An Introduction. 2nd ed. Edinburgh University Press: Edinburgh, UK.

[9] Knight, D., Khallaf, N., Rayson, P., et al., 2024. FreeTxt: A corpus-based bilingual free-text survey and questionnaire data analysis toolkit. Applied Corpus Linguistics. 4(3), 100103. DOI: https://doi.org/10.1016/j.acorp.2024.100103

[10] Baker, P., 2010. Sociolinguistics and corpus linguistics. Edinburgh University Press: Edinburgh, UK.

[11] Macaulay, R.K., 1991. Locating dialect in discourse: The language of honest men and bonnie lasses in Ayr. Oxford University Press: Oxford, UK.

[12] Himmelmann, N.P., 2006. Chapter 1 Language documentation: What is it and what is it good for? In: Gippert, J., Himmelmann, N.P., Mosel, U. (eds.), Essentials of Language Documentation. Walter de Gruyter: Berlin, Germany. pp. 1-30. DOI: https://doi.org/10.1515/9783110197730.1

[13] Kendall, T., 2008. On the History and Future of Sociolinguistic Data. Language and Linguistics Compass. 2(2), 332-351. DOI: https://doi.org/10.1111/j.1749-818X.2008.00051.x

[14] Sysoev, P.V., 2012. Modern Information and Communication Technologies: Didactic Properties and Functions. Language and Culture. MSU Publishing House: Moscow, Russia.

[15] Leech, G., 1991. The State of Art in Corpus Linguistics. In: Aijmer, K., Altenberg, B. (eds.), English Corpus Linguistics. Longman: London, UK. pp. 8-29.

[16] Sinclair, J., 1991. Corpus, Concordance, Collocation. Oxford University Press: Oxford, UK.

[17] Rykov, R.R., 2001. Corpus. Logical Induction and Deduction as Principles of Representing a Subject Area in a Text Corpus [Логические индукция и дедукция как принципы отражения предметной области в корпусе текстов]. In Proceedings of the International Seminar Dialogue 2001 on Computational Linguistics and Its Applications, Aksakovo, Russia. pp. 15-23.

[18] Souza, H.L., 2009. Collocations and Teaching: Investigating Word Combinations in Two English Textbooks for Norwegian Upper Secondary School Students [Master's thesis]. University of Bergen: Bergen, Norway. Available from: https://hdl.handle.net/1956/3824

[19] Nesselhauf, N., 2005. Collocations in a Learner Corpus. John Benjamins Publishing: Amsterdam, Netherlands.

[20] Chan, M.K.M., 2002. Concordancers and Concordances: Tools for Chinese Language Teaching and Research. Journal of the Chinese Language Teachers Association. 37(2), 1-58.

[21] Chambers, A., 2005. Integrating Corpus Consultation in Language Studies. Language Learning & Technology. 9(3), 111-125.

[22] Sysoev, P.V., 2010. Linguistic corpus in the methodology of teaching foreign languages. Yazyk i Kul'tura [Language and Culture]. 1(9), 99-111.

[23] Kaur, J., Radicchi, F., Menczer, F., 2013. Universality of scholarly impact metrics. Journal of Informetrics. 7(4), 924-932. DOI: https://doi.org/10.1016/j.joi.2013.09.002

[24] Khan, M.A., Pattnaik, D., Ashraf, R., et al., 2021. Value of special issues in the journal of business research: A bibliometric analysis. Journal of Business Research. 125, 295-313. DOI: https://doi.org/10.1016/j.jbusres.2020.12.015

[25] Mardieva, D., Bakiyeva, G., Kannazarova, Z., 2024. A Bibliometric Review: Interventions for Enhancing Speaking Skills in non-English-Speaking Contexts. XLinguae. 17(4), 195-224. DOI: https://doi.org/10.18355/XL.2024.17.04.12

[26] Zakharov, V.P., Bogdanova, S.Y., 2013. Corpus Linguistics: A textbook for linguistics students. 2nd ed. St. Petersburg State University Press: St. Petersburg, Russia.

[27] Goyibova, N., Muslimov, N., Sabirova, G., 2025. Differentiation approach in education: Tailoring instruction for diverse learner needs. MethodsX. 14, 103163. DOI: https://doi.org/10.1016/j.mex.2025.103163

[28] Bobojon, N.A., O'g'li, 2019. ICT in linguistic studies: Application of electronic language corpus and corpus-based analysis. Test Engineering and Management. 81(11-12), 4170-4176.

[29] Fry, J., 2004. The Cultural Shaping of ICTs within Academic Fields: Corpus-based Linguistics as a Case Study. Literary and Linguistic Computing. 19(3), 303-319. DOI: https://doi.org/10.1093/llc/19.3.303

[30] Leijen, D.A.J., 2011. Investigating factors influencing the usefulness of ICT supported peer feedback using corpus analysis. In: IADIS International Conference on Cognition and Exploratory Learning in Digital Age. pp. 319-320. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84882982358&partnerID=40&md5=db8c2f1d929c01d48f8c0a6cd0c6d6cc (cited 8 April 2025).