# Knowledge Mapping of Lexicography Research: A Visual Analysis With VOSviewer and CiteSpace

Jingwen Dong[1]

[1] Guangdong University of Foreign Studies, Guangdong, China

## Abstract

Lexicography has seen significant advancements during the past seven decades. This paper aims to provide a thorough review of lexicography research by analysing 7302 related publications from the Web of Science Core Collection database with two visualization tools: VOSviewer and CiteSpace. The lexicography studies are initially evaluated for the heated research topics, followed by identifying the most influential references, authors, journals, institutions, and nations/regions accordingly. Furthermore, through the analysis of co-cited references and keywords co-occurrence, the forthcoming research endeavours in lexicography are believed to encompass the prioritization of the user perspective in dictionary compilation, the emphasis on online and electronic dictionary development, and the exploration of dictionary integration with online technology.

**Keywords:** lexicography, scientometric, visual analysis, research trends, hotspots

## 1. Introduction

Dictionary-making is known to have a history of at least 3,000 years (Hartmann, 1984). Although the *Latin-English Dictionary of Syr Thomas Elyot Knight* (1538) marked the initial appearance of the term 'dictionary' in a dictionary title, the study of dictionary theory did not start until the 1950s. According to Hartmann and James (2000), while 'dictionary making' refers to a craft, 'lexicography' is known as a scientific description of the act of dictionary-making.

In fact, lexicography has undergone substantial development since the 1950s, which is notable in the formation of various lexicographical associations, such as the Dictionary Society of North America, founded in 1975; the European Association for Lexicography, established in 1983; the Australasian Association for Lexicography, formed in 1990, and the Asian Association for Lexicography, established in 1997. In addition, a series of lexicography journals have been produced, including *Lexicographica*—International Annual for Lexicography (established in 1985), *International Journal of Lexicography* (established in 1988), *Lexikos*—the journal of the African Association for Lexicography (established in 1991), and *Lexicography*—Journal of ASIALEX and Equinox (established in 2014).

Over the years, systematic reviews and practice guidelines have thoroughly analysed specific lexicography research issues, with related articles published in numerous journals. Among the earlier reviews, Quemada (1972), surveyed historical lexicography in Western Europe before the late 1960s. Then, by sketching "a 'typology' of the genres found in the literature on lexicography," a more comprehensive study conducted by Hartmann (1984) has offered an overview of lexicography development from historical, regional, and interlingual perspectives. Cowie (1989) has concentrated on the overview of dictionaries published for second-language learners of English. Apart from analysing the existing research, considerable insights have been presented into the potential future for developing such dictionaries.

Recently, an increased number of studies have emerged focusing on country-specific analyses of lexicographic developments. Notable examples include investigations on Italy (Marello, 2004), Great Britain (Karpova, 2018), Turkish (Bozkurt, 2019), China (Gao, 2021), Russain (Krylova, 2021), and Ukrainian (Blynova & Davydenko, 2021), among other regions. In addition, there are studies focused on specific categories or topics within the field of lexicography, including online dictionary (Tarp, 2012), academic lexicography (De Schryver, 2012a), Shakespeare lexicography (Karpova, 1992, 2020), digital lexicography (De Schryver et al., 2019), pedagogical dictionary (Li, 2019), English medical lexicography (Gordiyenko, 2021), EURALEX congress (De Schryver, 2012b), and lexicographic processes (Lemeshchenko-Lagoda, 2022), among others. Besides, some scholars employ bibliometric methodologies to offer a thorough analysis of the evolution of lexicography (McCREARY &

Dolezal, 1998; De Schryver, 2009). Nevertheless, up until now, there has been a dearth of scholarly efforts to summarize the discipline using a scientometric approach.

Scientometrics has empowered a thorough examination of a specific area and its evolution. By gaining a holistic understanding of the research landscape, the identification of research richness, differences, trends, potential research hotspots, and limitations can be facilitated by scientometrics. Therefore, this study employs CiteSpace and VOSviewer to visually represent and analyse the scholarly literature on lexicography, offering an intuitive way to examine the development of all aspects of lexicography.

## 2. Preparation of This Studies

### 2.1 Objectives

The primary objective of this study is to comprehensively depict the progression of lexicography research and discern the noteworthy research themes through the networks of co-cited references and co-occurring keywords. Meanwhile, it also hopes to identify the research networks, including the countries/regions, institutions, authors, and journals. Ultimately, the above analysis aims to discover research gaps, developing trends, and hotspots within the field.

### 2.2 Data Collection

With the Web of Science Core Collection (WOSCC) database as the data source, the search terms employed in this study are 'dictionar*' OR 'lexicograph*'. The categories are limited to 'Language Linguistics' and 'Linguistics', publication types to 'article' and 'review', and document types to 'Article', 'Proceeding Paper', 'Editorial Material', 'Review Article', and 'Early Access', with no limitation of language and time. Among the full records with cited references published before 2 September 2023, 7303 articles have been retrieved by WOSCC, with 7302 data items obtained by CiteSpace for analysis.

### 2.3 Methodology

This study employs visual software tools, namely CiteSpace (version 6.2.R4) and VOSviewer (version 1.6.19), for the analyses. As a visual analytic tool for visualising landmarks, critical paths, and emerging trends in a field of study based on relevant scholarly publications (Chen, 2022), CiteSpace is used to extract collaboration networks (countries/regions and institutions), co-citation analyses (co-cited authors, co-cited references, and co-cited journals), co-occurrence analyses (co-occurring keywords networks), and burst analyses (countries/regions, institutions, authors, references, and keywords). On the other hand, VOSviewer is a software tool for constructing and visualising bibliometric networks, which is used to obtain network maps of most-cited journals and co-occurring keywords networks in this study.

## 3. Results

### 3.1 Publication Outputs

The ultimate dataset has a total of 7302 papers, including 5333 articles, 1763 conference papers, 132 editorial materials, and 3 early access publications. These studies were published in 34 distinct languages, spanning the time period from 1956 to 2023. In 1956, the publication of two academic works on lexicography, Pfeffer's *Bilingual Lexicography* and Malkiel's *Linguistic Problems in a New Hispanic Etymological Dictionary*, marked the beginning of lexicography study. Although the annual quantity of publications generally below ten between 1956 and 1980, the number of annual publications increased yearly from 1981 onwards. Notably, in 2020, the number of annual publications peaked at 622, as shown in Figure 1.
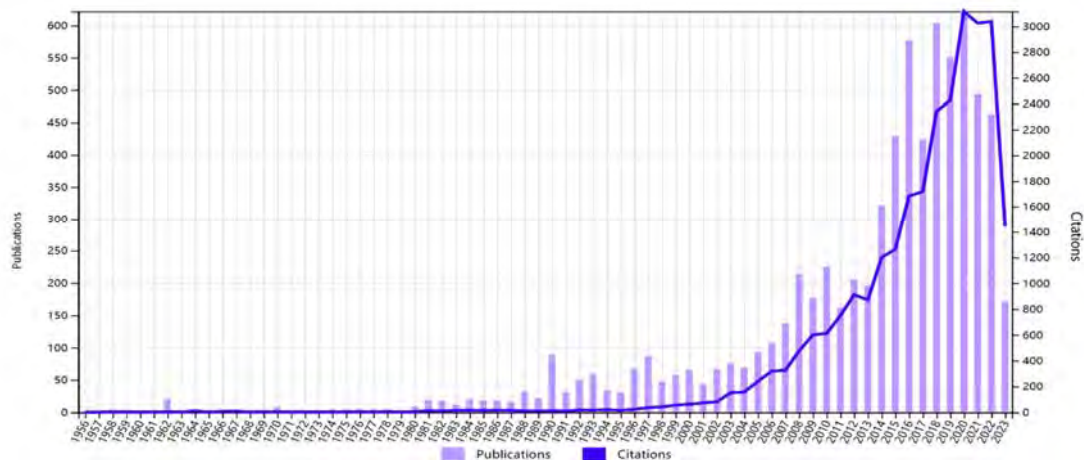
Figure 1. Times cited and publications (1956–2023)

The publication of lexicographic articles is closely related to the occurrence of academic conferences. For instance, the third international EURALEX conference took place in 1990, resulting in 89 articles published that year, marking the record before the 21st century. Besides, the significant rise in the number of articles published in 2014, 2015, and 2016 can be attributed to one or more lexicographic or linguistic conferences of that year. In addition, the inclusion of articles from the *International Journal of Lexicography* and *Lexikos* in Web of Science in 2003 and 2005 also positively influenced the subsequent number of publications and citations.

*3.2 Analysis of Co-Cited References*

The term co-citation was first introduced by Small, who argued that 'co-citation is the frequency with which two items of earlier literature are cited together by the later literature' (Small, 1973). Therefore, systematic reviews significantly benefit from co-citation networks because co-citation linkages may show how groups continuously change apart from initial publications. The research clusters, most cited papers, and turning-point papers will each be examined separately.

3.2.1 Clusters of Research

Generally, co-citation clustering refers to the formation of clusters of co-cited references (Boyack & Klavans, 2010). For papers published between 1956 and 2023, a map of reference co-citation is presented with corresponding clusters. Among the 44 clusters found, only the top 20 clusters are demonstrated based on the number of citations within. According to Figure 2(A), showing the co-citation references network, and Figure 2(B), demonstrating correspondent clustering analysis, which is obtained with CiteSpace, these clusters are considered highly credible (Q = 0.8679, S = 0.9656) (Note 1).
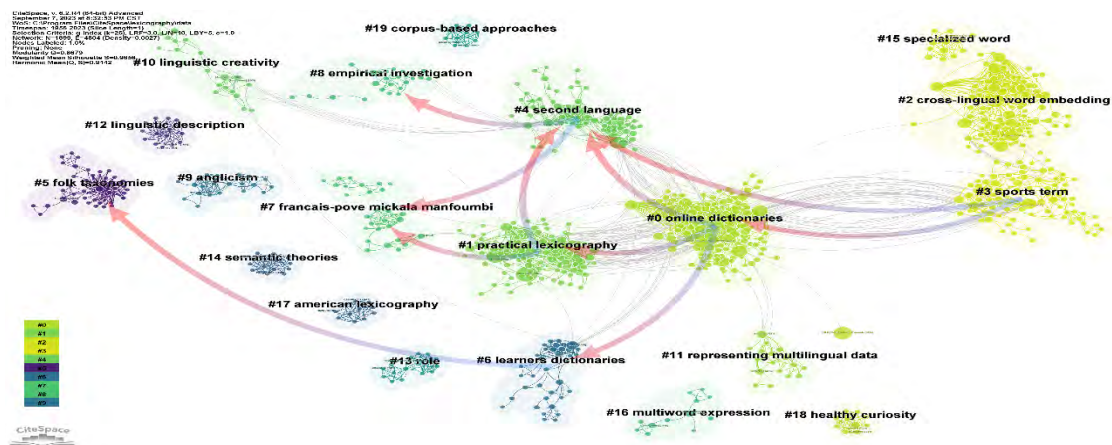
A
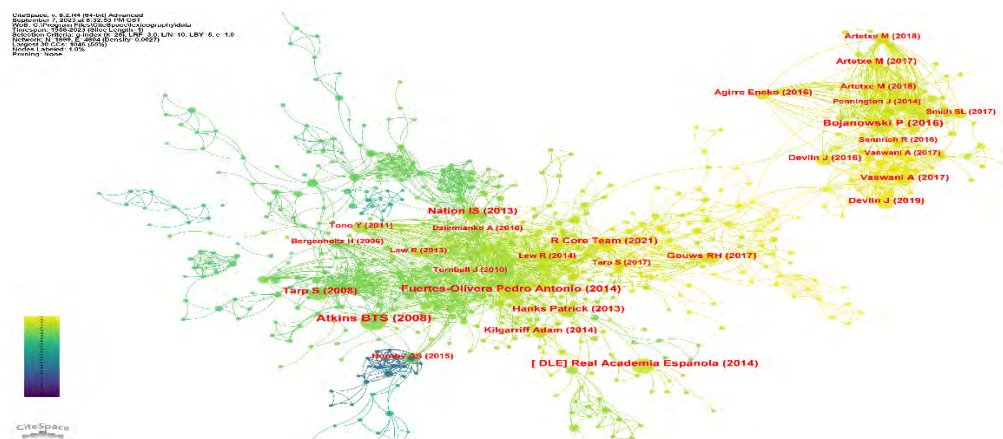
B



Figure 2. (A) Co-citation references network and (B) correspondent clusters (1956–2023)

The transition between clusters before and after the 21st century exhibits noticeable differences. Although the research clusters demonstrated a robust structure before the turn of the century, each cluster operated independently from the others with limited interconnections, such as cluster #12 ('linguistic description'; 16; S = 1; 1963) (Note 2), #9 ('anglicism'; 21; S = 1; 1978), #14 ('semantic theories'; 14; S = 1; 1975), #17 ('American lexicography'; 11; S = 1; 1973), #13 ('role'; 16; S = 1; 1988), and #19 ('corpus-based approach'; 10; S = 1; 1991). Therefore, the continuity of each research cluster is weak. However, a notable correlation between research clusters and enhanced research continuity has emerged in the new century. In this way, the evolutionary timeline of the clusters demonstrates a broad pattern in the progression of lexicography after the 21st century, with the trend starting in 2006, showcasing cluster #7 ('francais-pove mickala manfoumbi'; 29; S = 0.999; 2000), which has proved the research foundation for cluster #4 ('second language'; 68; S = 0.933; 2005) and cluster #1 ('practical lexicography'; 123; S = 0.925; 2008). Meanwhile, cluster #4 provided a research foundation for cluster #1 from 2008 to 2013, while cluster #8 ('empirical investigation'; 24; S = 0.989; 2000) also presented a research foundation for cluster #4 in 2010. In this way, clusters #4, #1, and #6 ('learners dictionaries'; 37; S = 0.999; 1987) further evolved into cluster #0 ('online dictionaries'; 175; S = 0.907; 2015). More recently, cluster #4 and cluster #0 became cluster #3 ('sports term'; 93; S = 0.964; 2018), and cluster #3 has a connection with cluster #2 ('cross-lingual word embedding'; 109; S = 0.989; 2017) in 2023.

In order to get deeper insights into current research trends, this investigation is centred on the network of co-cited literature from 2018 to 2023. As mentioned above, the reference co-citation with corresponding clusters is mapped out for articles from 2018 to 2023. Among the identified 38 clusters, only the top 20 are present based on the number of citations within each cluster. As shown in Figure 3(A), demonstrating the co-citation references network, and 3(B), depicting the correspondent clustering analysis obtained with CiteSpace, two emerging research trends are noticeable from the analysis. The first trend started from cluster #2 ('internet era'; 42; S = 0.903; 2014) with numerous lexicographical information research , and in 2018, it had evolved into cluster #3 ('lexicography-assisted writing assistant'; 34; S = 0.935; 2017), and in 2020 into cluster #4 ('virtue environment'; 32; S = 0.85; 2018). The second trend shows that 'cross-lingual word embedding' provided research foundations for cluster #0 ('neural machine translation'; 53; S = 0.872; 2018) from 2019 to 2021.
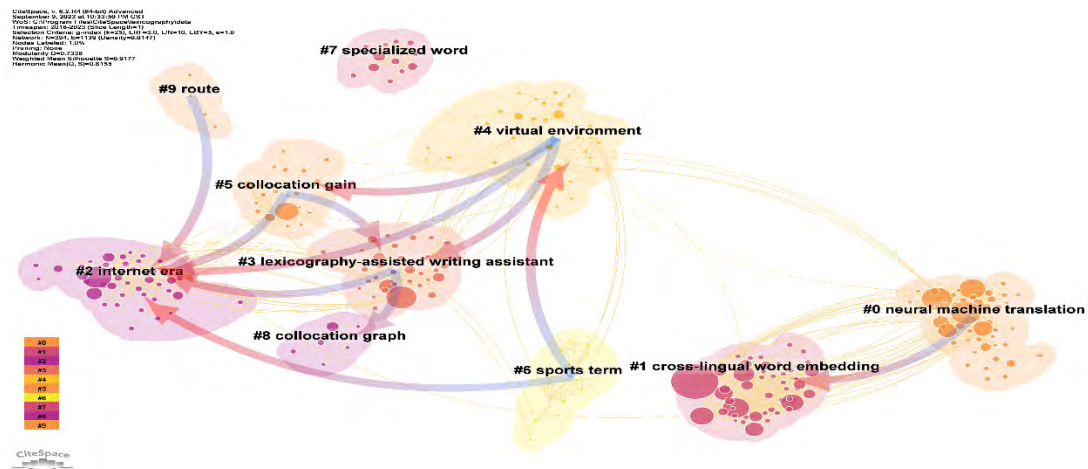
A
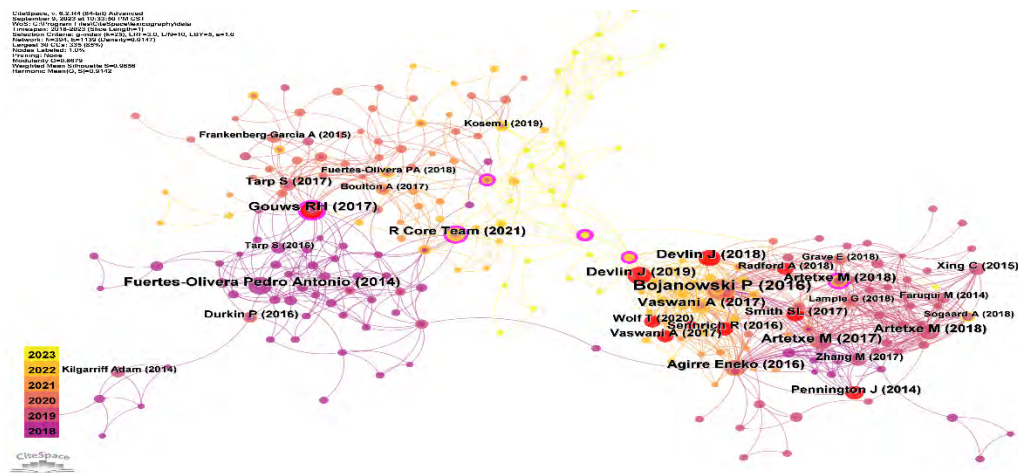


B



Figure 3. (A) Co-citation references network and (B) correspondent clusters (2018–2023)

### 3.2.2 Most Cited References And Turning-point References

Among the ten most co-cited references presented in Table 1, *The Oxford Guide to Practical Lexicography* (Atkins & Rundell, 2008) remains the most co-cited reference, with 51 citations in the mapped network. The top rang also includes *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography* (Fuertes-Olivera & Tarp, 2014), with 38 citations in the network. Besides, the paper on natural language processing published by Bojanowski et al. (2017) in the *Transactions of the Association for Computational Linguistics*, ranks third among the most cited references, demonstrating 35 citations in the network.

Articles with citation bursts are considered the center of attention at some point of time (Chen, 2022). Through a burst analysis of the co-cited literature spanning the years 1956 to 2023 and the more recent period from 2018 to 2023, the top 12 most co-cited papers are extracted, respectively. According to the burst analysis, the top three references with the most substantial citation burst are two books (Atkins et al., 2008; Tarp et al., 2008) and one dictionary (*Diccionario de la lengua española*, 2014) . However, when focusing on the past five years, the first place goes to Devlin et al.'s (2019) paper introducing a new language representation model called BERT, followed by Vaswani at el.'s (2017) paper on the new model of machine translation. As can be seen, lexicographic research has started to concentrate on fields like machine translation and neuroscience, thanks to the rapid development of information technology.

The 'turning-point' papers can be identified through the betweenness centrality. In other words, nodes with high betweenness centrality generally connect different clusters as key hubs. Firstly, Nation ISP's (2001) book *Learning Vocabulary in Another Language*, central to cluster #4 ('second language'), is not only a fundamental work in

language teaching and learning research but also further develops cluster #0 ('online dictionaries'). Similarly, Kosem et al.'s (2019) article about the example extraction tools GDEX (Good Dictionary Examples), is essential to cluster #3 ('sports term').

Table 1. The top 10 co-cited references

| Number of Citations | Centrality | Author | Year | Title | Source | Cluster ID | Type |
|---|---|---|---|---|---|---|---|
| 51 | 0.03 | B. T. Sue Atkins, Michael Rundell | 2008 | *The Oxford Guide to Practical Lexicography* | OXFORD UNIVERSITY PRESS | 1 | Book |
| 38 | 0.02 | Pedro A. Fuertes-Olivera, Sven Tarp | 2014 | *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography* | De Gruyter | 0 | Book |
| 35 | 0.01 | Piotr Bojanowski, Edouard Grave, Armand Joulin, Tomas Mikolov | 2017 | *Enriching Word Vectors with Subword Information* | Transactions of the Association for Computational Linguistics | 2 | Article |
| 35 | 0 | Real Academia Española | 2014 | *Diccionario de la lengua española* | Real Academia Española | 21 | Dictionary |
| 32 | 0.03 | Sylviane Granger (ed.), Magali Paquot (ed.) | 2012 | *Electronic Lexicography* | Oxford University Press | 0 | Book |
| 32 | 0.02 | Sven Tarp | 2008 | *Lexicography in the Borderland between Knowledge and Non-Knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography* | Max Niemeyer Verlag | 1 | Book |
| 31 | 0.07 | Nation I.S.P. | 2001 | *Learning vocabulary in another language* | Cambridge University Press | 4 | Book |
| 26 | 0.03 | Hanks Patrick | 2013 | *Lexical Analysis: Norms and Exploitations* | MIT Press | 0 | Book |
| 23 | 0.02 | Pedro A. Fuertes-Olivera(ed.), Henning Bergenholtz (ed.) | 2011 | *E-Lexicography. The Internet, Digital Initiatives and Lexicography* | Continuum | 1 | Book |
| 23 | 0.03 | R. H. Gouws, D. J. Prinsloo | 2005 | *Principles and Practice of South African Lexicography* | Stellenbosch African Sun Media | 1 | Book |

### 3.3 Co-Occurring Keywords Network

In the analysed dataset, there are 14,373 keywords, with 299 occurring at least ten times in frequency. With these 299 keywords, the research hotspots concerning microstructure, lemma, dictionary type, dictionary language, and the number of dictionary languages can be identified, as shown in Table 2. Regarding the microstructure of dictionaries, research hotspot is semantics (n = 277), followed by phraseology (n = 127), collocation (n = 119), and etymology (n = 101). The most studied lemma type is term (n = 162), followed by neologism (n = 98), idiom (n = 77), and dialect (n = 72). Meanwhile, further research also discovers that the terms' categories encompass agricultural, architectural and design, electronic, and touristic-cultural aspects. Besides, the most attention paid to the part of speech of words goes to the verb (n = 35), followed by the proverb (n = 32), noun (n = 15) and adjective (n = 14). In addition, the most studied type of dictionary is electronic dictionary (n = 100), followed by online dictionary (n = 86), historical dictionary (n = 76) and learner's dictionary (n = 57). Not surprisingly, the top two dictionary types are related to electronic tools, which proves Hanks' statement that 'Lexicography of the future will surely aim to create electronic tools' (De Schryver, 2012a).

Lexicographic languages of interest to lexicographers include English (n = 211), Russian (n = 119), Spanish (n = 58) and French (n = 51), with English taking the lead by a wide margin and Arabic (n = 13) and Chinese (n = 11), which are known as the working languages of the United Nations, receiving comparatively less attention in research endeavors. For the number(s) of languages involved, bilingual dictionaries (n = 138) have attracted the most attention, followed by monolingual dictionaries (n = 51). Even though the 'Big Five' have all launched their own monolingual English learner's dictionaries and continue to update them regularly, bilingual dictionaries remain the focus of lexicographers' research. Dictionaries in other languages number, such as multilingual dictionaries, have received less attention.

This study also covers the keyword bursts in the analyses, where keywords that experienced a surge in occurrences serve as indicators of hot topics (Chen, 2022). The results exhibit the most significant citation bursts in 'text reception' (6.95) and 'text production' (6.95) between 2005 and 2012. In particular, the keyword 'second language' demonstrates the most extended burst duration, spanning 16 years from 1996 to 2012. Comparatively, the 'Russian language' shows a significant citation burst of 5.61 in 2023. When considering the period between 2018 and 2023, 'pedagogical lexicography' ranks highest in burst strength and duration, scoring 4.24.

Table 2. Top 10 key words

| Microstructure | Freq | Lemma | Freq | Dictionary types | Freq | Language | Freq | Number of languages | Freq |
|---|---|---|---|---|---|---|---|---|---|
| Semantics | 277 | Term | 162 | Electronic dictionary | 100 | English | 211 | Bilingual dictionary | 138 |
| Phraseology | 127 | Neologism | 98 | Online dictionary | 86 | Russian | 119 | Monolingual dictionary | 51 |
| Collocation | 119 | Idiom | 77 | Historical dictionary | 76 | Spanish | 58 | Multilingual dictionary | 14 |
| Etymology | 101 | Dialect | 72 | Learner's dictionary | 57 | French | 51 | | |
| Metaphor | 82 | Loanword | 28 | Specialized dictionary | 25 | Polish | 39 | | |
| Phonology | 64 | Slang | 18 | Corpus-based dictionary | 24 | German | 29 | | |
| Polysemy | 62 | Verb | 35 | Dialect dictionary | 24 | Italian | 29 | | |
| Morphology | 45 | Proverb | 32 | Explanatory dictionary | 16 | African | 26 | | |
| Grammar | 40 | Noun | 15 | Terminological dictionary | 11 | Czech | 20 | | |
| Equivalence | 39 | Adjective | 14 | Pronunciation dictionary | 10 | Croatian | 17 | | |

Based on the average publication years, this study extracts the visualisation overlay for the co-occurring keyword networks with VOSviewer. Except for the keywords analysed above, some of the most cited keywords are corpus (n = 147), translation (n = 140), and dictionary use (n = 102).

*3.4 Analysis of Major Co-Cited Journals and Authors*

Co-cited journals can indicate high-impact journals in the field of lexicography. The ten journals with the most citations are *International Journal of Lexicography*, *Lexikos*, *Natural Language Engineering*, *Applied Linguistics*, *Lexicographica*, *Computational Linguistics*, *TESOL Quarterly*, *Lecture Notes in Computer Science*, *Language learning*, *and Modern Language Journal* (Table 3). When considering the period between 2018 and 2023, the top 5 journals with the most citations are *International Journal of Lexicography*, *Lexikos*, *Natural Language Engineering*, *Applied Linguistics*, and *Computational Linguistics*. In the past five years, the ranking of journal co-citation frequency has remained the same.

Besides, among the co-cited journals with the most publications over the past 20 years, the five journals with the most articles are *Lexikos* (n = 429), *International Journal of Lexicography* (n = 373), *Voprosy Leksikografii Russian Journal of Lexicography* (n = 140), *Vestnik Volgogradskogo Gosudarstvennogo Universiteta Seriya 2 Yazykoznanie* (n = 113), and *Lecture Notes in Artificial Intelligence* (n = 110).

Co-cited authors can signify influential figures within the lexicography field, identifying highly influential authors and those who operate as linkages between clusters and share common research interests. Overall, the most

frequently co-cited author is Tarp S, followed by Atkins BTS, Bergenholtz H, Wiegand HE, Gouws RH, Sinclair J, Lakoff G, Rundell M, and Zgusta L. At the same time, the timeline of co-cited author clusters identifies Hausmann FJ and Wiegand HE as two major authors linking clusters #0 ('online dictionaries') and #1 ('new media') . In order to get deeper insights into current high-impact authors, this investigation is centred on the network of co-cited authors from the last five years (2018–2023). The top 5 co-cited authors in the past five years are Tarp S, Atkins BTS, Lew R, Gouws RH, and Lakoff G.

Authors with a burst of occurrences indicate rising stars with spectacular productivity (Chen, 2022). In the last five years, the top three authors with the most vigorous citation bursts are Devlin J, Wolf T, and Vaswani A. Notably, Devlin J demonstrates both a high burst strength and a significant centrality of 0.11, suggesting his position as an emerging author who has been frequently cited in the past five years and a connector between cluster #2 ('frame semantics') and cluster #3 ('cross-lingual word embedding').

Table 3. Top 10 co-cited journals

| Journals with most citations (1956–2023) | Initial year | Impact factor (2022) | Total co-citations | Total articles |
|---|---|---|---|---|
| International Journal of Lexicography | 1988 | 0.5 | 948 | 373 |
| Lexikos | 1991 | 0.4 | 605 | 429 |
| Natural Language Engineering | 1995 | 2.5 | 386 | 26 |
| Applied Linguistics | 1980 | 3.6 | 377 | 27 |
| Lexicographica | 1985 | 0.3 | 302 | 56 |
| Computational Linguistics | 1974 | 9.3 | 268 | 27 |
| TESOL Quarterly | 1967 | 3.2 | 209 | 10 |
| Lecture Notes in Computer Science | 1973 | 1.27 | 200 | 33 |
| Language learning | 1948 | 4.4 | 199 | 13 |
| Modern Language Journal | 1916 | 4.9 | 194 | 19 |

*3.5 Analysis of Cooperation Networks Across Countries/Regions and Institutions*

Among the 124 countries/regions involving the analysed dataset, 79 countries exhibit over five article publications. According to the network of the co-authors' countries illustrated in Figure 4(A), each node's dimensions represent the issuance volume of the respective country or region. At the same time, the connecting lines symbolise collaborative partnerships among countries or regions. The purple outer circle around a node indicates a high centrality, signifying its significant influence within the network. Therefore, the top five countries/regions that demonstrate the highest number of publications are Russia (n = 788), Spain (n = 723), the US (n = 497), Germany (n = 433), and the People's Republic of China (n = 319), collectively representing 38% of the overall article count. Meanwhile, the top five countries/regions exhibiting the highest centrality levels are Spain, the US, South Africa, the UK, and Germany, revealing that the US is the most cited country (n = 5440), followed by Germany (n = 3059), the UK (n = 2677), Canada (n = 2276) and Spain (n = 1661).

When considering the past 5 years (2018–2023), the ranking of the top five countries/regions with the highest number of publications remain consistent, except for the People's Republic of China, which surpasses Germany in position. Yet, the top 5 countries/regions with the highest centrality have shifted notably, with the US, Russia, Germany, Spain, and South Africa leading the way. Moreover, the US remains the most cited country when it comes to the number of citations.

The examination of burstness indicates that the US had the most significant citation burst (51.65) and the most extended burst duration (1977–2007), with the top 12 countries/regions exhibiting bursts before 2013. However, further examination in the last five years (2018–2023) reveals that only Kazakhstan exhibits bursts, despite its weakness (2.1). In this way, the number of articles from each country/region has been generally steady in recent years, with no significant increase in a particular country/region.

The analysed data contains 3038 institutions, of which 671 are in partnership with each other. It is observed that 382 institutions within this dataset exhibit an article publication frequency exceeding five times. According to Figure 4(B), which illustrates the network of co-authors' institutions, the top five institutions with the most papers are: Russian Academy of Sciences (n = 144), Stellenbosch University, Centre National de la Recherche Scientifique (CNRS), UDICE-French Research Universities and Aarhus University. The top 5 institutions with the most centrality are: Austrian Academy of Sciences, KU Leuven, Aarhus University, Russian Academy of Sciences, Centre National de la Recherche Scientifique (CNRS) and Hungarian Academy of Sciences.

When considering a limited period of the past five years, the Russian Academy of Sciences (n = 108) has

maintained its position as the institution with the highest number of publications, followed by the Ministry of Education & Science of Ukraine, Stellenbosch University, the Centre National de la Recherche Scientifique (CNRS), and Tomsk State University. The Ministry of Education & Science of Ukraine has seen a significant increase in the articles published in the last five years. Moreover, the centrality of the Russian Academy of Sciences has increased, making it the highest centrality in the network of institutional collaborations in the last five years. Therefore, the Russian Academy of Sciences has occupied a leading position in terms of both publication output and influence when it comes to lexicography from 2018 to 2023.

The analysis of burstness reveals that Ghent University exhibits the most vigorous citation burst (13.81), while the University of Chicago has the longest burst duration (1974–2007). Additionally, the Saarland University demonstrates the most recent and the strongest citation burst (3.37).
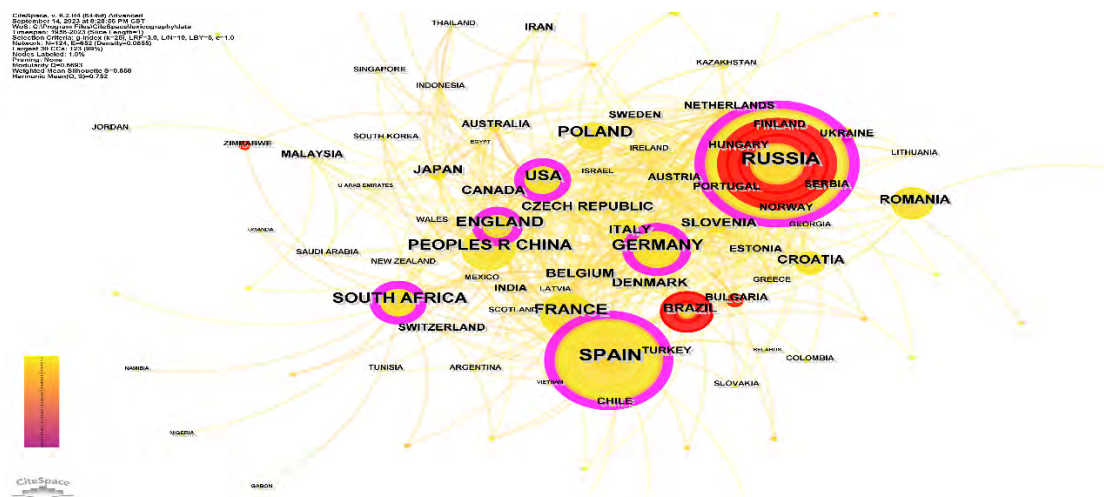


Figure 4. (A) Network of the co-authors' countries and (B) network of co-authors' institutions (1956–2023)

## 4. Discussion

### 4.1 Summary of the Main Findings

The results demonstrate significant development in lexicography over the past seven decades. The exponential rise in lexicography publications since 1980 reflects a robust and expanding research landscape, with an annual publication count stabilizing at over 100 articles by 2010 and reaching a record-breaking 622 publications in 2020.

The identified key countries/regions, institutions, authors, and journals provide valuable insights into the global distribution of lexicography research. Meanwhile, the findings of Russia leading in total publications, Spain in

centrality, and the US in citation count demonstrate the diverse and widespread nature of lexicography scholarship.

The prominent works, such as *The Oxford Guide to Practical Lexicography*, and influential authors like Tarp S, Atkins BTS, Bergenholtz H, Wiegand HE, and Gouws RH, highlight the impact of specific publications and researchers within the field.

Furthermore, the analysis of the co-cited reference network from 1956 to 2023 reveals a growing association between research clusters and increased research continuity since the beginning of the 21st century, suggesting a maturing and interconnected research landscape while fostering collaboration and sustained inquiry.

Overall, these results signify the vibrant and dynamic characteristic of lexicography research, thanks to the extensive global participation, influential contributions, and evolving research networks, all of which contribute to the rich tapestry of scholarly inquiry within this domain.

*4.2 Future Trends in Lexicography Research*

Future development trends in lexicography studies can further be analysed from the co-cited reference clusters, burst detection, and keyword co-occurrence frequency. There are three main points.

4.2.1 Prioritising the User Perspective in Dictionary Compilation

The lexicographic viewpoint can be broadly divided into two categories: the compiler's perspective and the user's perspective. A gradual transition towards the user-centred paradigm since the late 1970s can be observed in global lexicography theory and practice (Geng & Wei, 2013), leading to the increasing marginalisation of the 'compiler's perspective' in favour of the 'user's perspective'. The trend of focusing on the 'user perspective' in this study is reflected through the increase in keyword frequency for 'dictionary use,' 'students,' and 'learners'. In particular, 'students' exhibits significant burst strength, ranking in the top twelve keywords for the research categories from 1956 to 2023 and 2018 to 2023.

*Dictionaries and Their Users*, a collection compiled by Hartmann in 1979, marked the initial endeavour to tackle the challenges encountered by dictionary users (McCreary & Dolezal, 1998). In the analysed dataset, while the first article containing 'dictionary use' among the keywords appeared in 1997, the first article containing the keywords 'students' and 'learners' appeared in 1997 and 2008, respectively. Since then, the number of articles has increased. Notably, the most current research on dictionary use has adopted the empirical or experimental paradigm based on the analyses of 102 articles identified through the keyword 'dictionary use'. Meanwhile, the advancing research on 'dictionary use' has coincided with an increase in empirical studies in this field, revealing a tendency towards more reliable and convincing results and conclusions.

4.2.2 Emphasising Online and Electronic Dictionary Development

Most dictionaries can be classified with several properties, among which three categorisation methods are considered the most common. Initially, dictionaries can be classified into monolingual, bilingual, and multilingual dictionaries based on the number(s) of languages involved. Meanwhile, general dictionaries, specialised dictionaries, and learners' dictionaries can be classified based on the types of dictionary users. In addition, paper-based dictionaries and online or electronic dictionaries are categorised based on the different presentation mediums. The primary dictionary category prioritised by lexicographers is electronic and online dictionaries. According to the data in Table 2, electronic and online dictionaries are more prevalent as keywords, which are expected to be the subject of extensive research in the future. Furthermore, the analysis of co-cited references clusters spanning from 1956 to 2023 (Figure 1) recognises 'online dictionary' cluster as the biggest.

The distinction between electronic and online dictionaries is ambiguous. According to Tarp (2012), internet dictionaries are considered synonymous with electronic dictionaries. Dai and Xu (2014), conversely, posited online dictionaries as a form of electronic dictionaries, and they argued that an electronic dictionary for general users refers to a dictionary that uses a hard disk, CD-ROM, disk or chip as a storage medium and presents the information on the screen of personal computers, cell phones, and other terminals with the help of a microprocessor and related computers, networks, and other technologies for the use of general users. No matter how electronic dictionaries and online dictionaries are divided, it is indisputable that they represent the predominant direction for dictionary development.

The Oxford English Dictionary introduced its first electronic dictionary in 1988 and subsequently made the decision to discontinue its print version in 2010. Meanwhile, the print version of Macmillan Dictionary ceased publication in 2013. Electronic dictionaries and online dictionaries have developed rapidly, with major dictionaries introducing their own dictionary applications. However, it is important to note that electronic dictionaries and online dictionaries should not be regarded as mere electronic replicas of traditional paper dictionaries, and there

are big differences between the two in terms of compilation style, information storage mode and retrieval mode, etc. Hence, lexicographers are dedicated to investigate the methodologies of handling and displaying information within the context of electronic dictionaries and online dictionaries. The relative research includes: multimodal exemplification (Liu, 2015) and multimodal definition (Liu, 2017); the use of electronic dictionaries and user's needs (Lew & De Schryver, 2014); the colour of functional label (Dziemianko, 2015); the impact of corpus for dictionary makers and users (Heuberger, 2016); information and data overload (Gouws & Tarp, 2016); and the typography in online English learner's dictionaries (Hao et al., 2022).

### 4.2.3 Exploring Dictionary Integration With Modern Technology

Modern technology is gradually taking up the compilation and utilization of dictionaries. Initially, it has been seamlessly integrated into the entire dictionary compilation process. For example, the corpus has become a vital tool for compiling dictionaries, with major English dictionaries having established their own corpora to provide data support for dictionary compilation. Nevertheless, there remains to be a challenge posed by the expanding corpus size in the data screening for lexicographic compilation. According to Rundell (2002), a transition exists in corpus lexicography, away from the emphasis on corpora size and composition, which was frequently discussed in the 1980s and 1990s. Therefore, the focus should be on the recent difficulties in effectively extracting lexicographically significant information from extensive text databases.

Lexicographic technology involves several stages, including developing corpus construction and retrieval tools, extracting, analysing, and retrieving corpus, and improving the multimodal presentation of lexicographic textual information (Geng & Wei, 2017). Currently, researchers have investigated relevant technologies. For instance, Geng and Wei (2017) introduced the Verb Sketch System, a technical modal for composing verb entries in English-Chinese learner's dictionaries, as an illustrative case to explore the lexicography technology development. Then, Kosem et al. (2019) designed the first extraction tool, the GDEX (Good Dictionary Examples), to assist lexicographers in identifying high-quality examples. Furthermore, bilingual word embeddings in natural language processing has been employed in lexicography to facilitate the creation of bilingual dictionaries (Mikolov et al., 2013). However, there remains a notable lack of systematic studies focusing on the technologies for compiling dictionaries from a general viewpoint, compared to other research areas, such as dictionary usage.

Modern technology has significantly influenced the usage of dictionaries, making them more than just traditional paper-based or electronic resources for finding information. Nowadays, dictionaries are increasingly integrated into a tool designed to assist reading, writing, and translation. Based on the reference timeline for co-citation clusters over the past five years, cluster #0 neural machine translation, cluster #1 cross-lingual word embedding, and cluster #3 lexicography-assisted writing assistant have emerged as research areas particularly active between 2018 and 2023, suggesting lexicographers attention on investigating lexicographic principles integration for writing, translation, and reading platforms and programmes. For example, Frankenberg-Garcia et al. (2019) have established a project named ColloCaid (www.collocaid.uk) for those wishing to master English collocations for academic purposes. Meanwhile, by describing the functionalities of various digital writing assistants, Olivera and Tarp (2020) focus on one such tool, the Spanish-English Write Assistant. In addition, inspired by a monolingual speaker's acquisition of translation skills through reference to a bilingual dictionary, Duan et al. (2020) have proposed a novel machine translation task with a ground-truth bilingual dictionary and large-scale monolingual corpora but no available parallel sentences.

### 4.3 Limitations

Thanks to the significant number of references and the consistency of clusters, this study's analysis is considered successful. However, it is still impossible to rule out abnormal clusters. Meanwhile, since the data were only collected from WOSCC, which fails to cover all lexicographic outputs, a somewhat incomplete retrieval of the publication may affect the accuracy of the findings.

## 5. Conclusion

Since lexicographical practice and theoretical research complement one another, the development of one leads to the improvement of the other. This scientometric study presents a visual analysis of the literature about lexicographic research within the WOSCC spanning the years 1956 to 2023, presenting a comprehensive overview of the current state of lexicographic research from a macroscopic standpoint. Therefore, the emerging trends, focal areas of interest, and gaps can be identified in lexicography research, offering valuable insights to researchers seeking to comprehend the evolutionary trajectory of lexicographic research.

**Competing interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Informed consent**

Obtained.

**Ethics approval**

The Publication Ethics Committee of the Canadian Center of Science and Education.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

**Provenance and peer review**

Not commissioned; externally double-blind peer reviewed.

**Data availability statement**

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Data sharing statement**

No additional data are available.

**References**

Atkins, B. T. S., & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press. https://doi.org/10.1093/oso/9780199277704.001.0001

Blynova, I. A., & Davydenko, V. I. (2021). Contemporary Ukrainian Lexicography: Status, Typology, and Development trends. Scientific Journal of National Pedagogical Dragomanov University. Series 9. *Current Trends in Language Development*, *21*, 5–18. https://doi.org/10.31392/NPU-nc.series9.2021.21.01.

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, *61*(12), 2389–2404. https://doi.org/10.1002/asi.21419

Bozkurt, F. (2019). Monitoring Academic Studies of Turkish Lexicography: A Bibliometric Study of 84 Years. *Lexikos*, *29*(1), 288–315. https://doi.org/10.5788/29-1-1522

Chen, C. (2022). *How to Use CiteSpace*. Retrieved from http://leanpub.com/howtousecitespace

Cowie, A. P. (1989). Pedagogical Descriptions of Language: Lexis. *Annual Review of Applied Linguistics*, *10*, 196–209. https://doi.org/10.1017/S0267190500001306

Dai, Y., & Xu, H. (2014). Abstracts of Major Papers in This Issue Electronic Lexicography: Status Quo and Prospects. *LEXICOGRAPHICAL STUDIES*, *4*, 1–9, 93.

De Schryver, G.-M. (2009). Bibliometrics in Lexicography. *International Journal of Lexicography*, *22*(4), 423–465.

De Schryver, G.-M. (2012a). Trends in Twenty-five Years of Academic Lexicography. *International Journal of Lexicography*, *25*(4), 464–506. https://doi.org/10.1093/ijl/ecp027

De Schryver, G.-M. (2012b). Lexicography in the crystal ball: facts, trends and outlook. In V. F. Ruth & M. T. Julie (Eds.), *Proceedings of the 15th EURALEX International Congress, Euralex 2012 Proceedings, 7–11 August 2012* (pp. 93–163). Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

https://doi.org/10.1093/ijl/ecs030

De Schryver, G.-M., Chishman, R., & Silva, B. D. (2019). An overview of Digital Lexicography and directions for its future: An interview with Gilles-Maurice de Schryver. *Calidoscópio*, *17*(3), 659–683. https://doi.org/10.4013/ld.2019.173.13

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In B. Jill, D. Christy & S. Thamar (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2–7 June 2019* (pp. 4171–4186). Minnesota: Minneapolis. Association for Computational Linguistics.

Duan, X., Ji, B., Jia, H., Tan, M., Zhang, M., Chen, B., … Zhang, Y. (2020). Bilingual Dictionary Based Neural Machine Translation without Using Parallel Sentences. In J. Dan, C. Joyce, S. Natalie & T. Joel (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5–10 July 2020* (pp. 1570–1579). Online. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.143

Dziemianko, A. (2015). Colours in Online Dictionaries: A Case of Functional Labels. *International Journal of Lexicography*, *28*(1), 27–61. https://doi.org/10.1093/ijl/ecu028

Española, R. A. (2014). *Diccionario de la lengua española* (edición del tricentenario). Obtenido de Definición de elemento químico http://dle. rae. es.

Frankenberg-Garcia, A., Lew, R., Roberts, J. C., Rees, G. P., & Sharma, N. (2019). Developing a writing assistant to help EAP writers with collocations in real time. *ReCALL*, *31*(1), 23–39. https://doi.org/10.1017/S0958344018000150

Fuertes-Olivera, P. A., & Tarp, S. (2014). *Theory and practice of specialised online dictionaries: Lexicography versus terminography* (Vol. 146). Berlin: Walter de Gruyter GmbH & Co KG. https://doi.org/10.1515/9783110349023

Fuertes-Olivera, P. A., & Tarp, S. (2020). A window to the future: Proposal for a lexicography-assisted writing assistant. *Lexicographica*, *36*(2020), 257–286. https://doi.org/10.1515/lex-2020-0014

Gao, Y. (2021). Whither Chinese-English lexicography? – From a historical perspective. *Lexicography*, *8*(2), 107–129. https://doi.org/10.1558/lexi.20869

Geng, Y., & Wei, X. (2013). On the Paradigm Shifts in the Defining Practice of English Lexicographers: A Perspective of the Compiler's View of Language. *Foreign Language in China*, *10*(5), 95–100.

Geng, Y., & Wei, X. (2017). On the Implications of the Fourth Paradigm Shift in Scientific Studies for Contemporary Dictionary-Making. *TEFLE*, *2*, 65–70.

Gordiyenko, O. V. (2021). Dynamics and trends in the development of English medical lexicography (XVII c. – Early XXI c.). *Science and Education a New Dimension*, *248*(23), 29–31. https://doi.org/10.31174/SEND-Ph2021-248IX73-07

Gouws, R. H., & Tarp, S. (2016). Information overload and data overload in lexicography. *International Journal of Lexicography*, *30*(4), 389–415. https://doi.org/10.1093/ijl/ecw030

Hao, J., Xu, H., & Hu, H. (2022). A Multimodal Communicative Approach to the Analysis of Typography in Online English Learner's Dictionaries. *International Journal of Lexicography*, *35*(2), 234–260. https://doi.org/10.1093/ijl/ecab031

Hartmann, R. R. K. (1984). Lexicography. A Contrastive Survey. *Annual Review of Applied Linguistics*, *5*, 124–138. https://doi.org/10.1017/S0267190500001847

Hartmann, R. R. K., & James, G. (2000). *Dictionary of lexicography*. Beijing: Foreign Language Teaching and Research Press.

Heuberger, R. (2016). Corpora as game changers: The growing impact of corpus tools for dictionary makers and users: Corpus tools have created a paradigm shift in English lexicography. *English Today*, *32*(2), 24–30. https://doi.org/10.1017/S0266078415000474

Karpova, O. M. (1992). Shakespeare Lexicography. Trends of development (XVIII-XX cc.). In T. Hannu, V. Krista, S.-T. Tarja & S. Jürgen (Eds.), *The 5th EURALEX International Congress on Lexicography, Euralex 1992 Proceedings, Tampere, Finland, 4–8 August 1992* (pp. 593–600). Tampere: Tampereen yliopisto.

Karpova, O. M. (2018). MODERN LEXICOGRAPHIC PICTURE OF GREAT BRITAIN. *Bulletin of the Moscow State Regional University* (Linguistics), *6*, 28–36. https://doi.org/10.18384/2310-712X-2018-6-28-36

Karpova, O. M. (2020). A new wave of Shakespeare lexicography (with special reference to LSP dictionaries). *Lexicographica*, *36*, 241–254. https://doi.org/10.1515/lex-2020-0013

Kosem, I., Koppel, K., Zingano Kuhn, T., Michelfeit, J., & Tiberius, C. (2019). Identification and automatic extraction of good dictionary examples: The case(s) of GDEX. *International Journal of Lexicography*, *32*(2), 119–137. https://doi.org/10.1093/ijl/ecy014

Krylova, M. N. (2022). Modern Trends in the Development of Russian Lexicography. *KAZAN LINGUISTIC JOURNAL*, *5*(2), 166–179. https://doi.org/10.26907/2658-3321.2022.5.2.166-179

Krylova, O. N. (2021). *Modern Russian lexicology, lexicography and linguogeography*. Moscow: Institute for linguistic Studies Russian Academy of Sciences. https://doi.org/10.30842/265837632021

Lemeshchenko-Lagoda, V. (2022). Diachronic and Synchronic Analysis of Lexicographic. Processes of Minority Language in Scotland: a Contrastive Approach. Scientific Journal of National Pedagogical Dragomanov University. Series 9. *Current Trends in Language Development*, *21*, 76–89. https://doi.org/10.31392/NPU-nc.series9.2021.21.06

Lew, R., & De Schryver, G.-M. (2014). Dictionary Users in the Digital Revolution. *International Journal of Lexicography*, *27*(4), 341–359. https://doi.org/10.1093/ijl/ecu011

Li, J. (2022). *CiteSpace Chinese Guide*. Retrieved from http://blog.sciencenet.cn/u/jerrycueb

Li, Q. (2019). Empirical Research on Pedagogical Dictionary Use in Recent 30 Years. *Journal of Education and Learning*, *8*(6), 103. https://doi.org/10.5539/jel.v8n6p103

Liu, X. (2015). Multimodal Definition: The Multiplication of Meaning in Electronic Dictionaries. *Lexikos*, *25*, 210–232. https://doi.org/10.5788/25-1-1296

Liu, X. (2017). Multimodal Exemplification: The Expansion of Meaning in Electronic Dictionaries. *Lexikos*, *27*(1), 287–309. https://doi.org/10.5788/27-1-1404

Marello, C. (2004). Lexicography in Italy: Specific Themes and Trends. *International Journal of Lexicography*, *17*(4), 349–356. https://doi.org/10.1093/ijl/17.4.349

McCreary, D. R., & Dolezal, F. (1998). Language Learners and Dictionary Users: Bibliographic Findings and Commentary. In T. Fontenelle, P. Hiligsmann, A. Michiels, A. Moulin & S. Theissen (Eds.), *The Eighth EURALEX International Congress on Lexicography, EURALEX'98 Proceedings, Liège, Belgium, 4–8 August 1998* (pp. 611–618). Liège: Département d'anglais et de néerlandais, Université de Liège.

Mikolov, T., Le, V., & Sutskever, I. (2013). *Exploiting Similarities among Languages for Machine Translation* (arXiv:1309.4168). arXiv.

Nation, I. S. P. (2001). *Learning vocabulary in another language* (Vol. 10). Cambridge: Cambridge university press. https://doi.org/10.1017/CBO9781139524759

Quemada, B. (1972). Lexicology and lexicography. In T. A. Sebeok (Ed.), *Current trends in linguistics* (vol. 9 Linguistics in Western Europe, pp. 395–475). The Hague: Mouton. https://doi.org/10.1515/9783111561929-011

Rundell, M. (2002). Good Old-fashioned Lexicography: Human Judgment and the Limits of Automation. In M. -H. Corréard (Ed.), *Lexicography and Natural Language Processing: A Festschrift in honour of B. T. S. Atkins* (pp. 138–155). Stuttgart: EURALEX.

Small, H. (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

Tarp, S. (2008). *Lexicography in the borderland between knowledge and non-knowledge: general lexicographical theory with particular focus on learner's lexicography*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783484970434

Tarp, S. (2012). Online dictionaries: today and tomorrow. *Lexicographica*, *28*, 253–268. https://doi.org/10.1515/lexi.2012-0013

Vaswani, A., Shazeer, N., Parmar, J., Uszkoreit, L., Jones, N., Gomez, Ł., … Polosukhin, I. (2017). Attention is all you need. In von L. Ulrike, G. PictureIsabelle, B. PictureSamy, W. PictureHanna & F. PictureRob (Eds.),

*Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), 4–9 December 2017* (pp. 6000–6010). California: Curran Associates Inc.

**Notes**

Note 1. The degree to which a network may be divided into modules or clusters is measured by the network's modularity (also known as the Q score). When the cluster structure is considered significant, the Q metric—which ranges from 0 to 1—must be greater than 0.3. An approach for analysing and confirming consistency within data clusters is the silhouette (also known as the S score). When the S score exceeds 0.3, it indicates that the cluster network is homogeneous, and a value of 0.7 is considered highly reliable. The S measure goes from -1 to +1. A silhouette score of 1 means the cluster is relatively isolated.

Note 2. The information in parentheses represents the clusters' label, size, silhouette score and the average year of publication of the cluster members.

**Copyrights**