Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

91

Olga V. Dekhnich[1] 🆔
Tatyana A. Litvinova[2] 🆔

| | **A keystroke analysis to study writing: a bibliometric review using R and VOSviewer** |

[1] Belgorod State National Research University,
85 Pobedy St., Belgorod, 308015, Russia
E-mail: dekhnich@bsu.edu.ru
ORCID: 0000-0001-6088-2656

[2] Voronezh State Pedagogical University,
86 Lenin St., Voronezh, 394043, Russia
E-mail: centr_rus_yaz@mail.ru
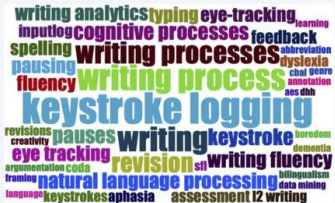ORCID: 0000-0002-6019-3700

Abstract: Data obtained by means of keystroke logging software, which records all writing events, i.e., all keypresses with their time stamps, are widely used in different research fields and practical applications – from early detection of cognitive impairments to user authentication. Keystroke logging software have become especially popular in writing studies since it allows researchers to observe writing processes unobtrusively and increases their understanding of what happens "behind the scenes" of text production. This methodology has been used in writing studies since the advent of affordable computers in the 1990s, however, there has still been no bibliometric review that would provide a holistic picture of this field and identify hot topics and trends, perform citation analysis and identify the most productive authors and countries in the field of writing research using keystroke data. This paper is pioneer research to conducting a bibliometric analysis on the field "keystroke analysis to study writing" aimed at filling in the gap. A search was conducted in the bibliographic database Scopus on July 11, 2024. We included studies published in English post-2000 that discuss the use of keystroke data and methodology to study writing process. This review followed the guidelines of the PRISMA protocol to perform the study search and selection. The search yielded 336 documents of which 273 met out inclusion criteria. The records retrieved were analysed using the bibliometrix R-package and VOSviewer software. Using these tools in combinations, we implemented both performance analysis which examines the contributions of research constituents to a given field and science mapping which is aimed at revealing the relationships between research constituents. The contribution of this study is twofold. Firstly, it provides an in-depth bibliometric analysis of an actively developing field of research which is rather limited in terms osf the countries and institutions involved (and the languages of the texts analysed), hopefully saving time and effort for

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

92

researchers new to the field. Secondly, we provide an example of using bibliometrix and VOSviewer for a bibliometric analysis which could be easily replicated in further research.

УДК 81'3

DOI: 10.18413/2313-8912-2024-10-3-0-5

Дехнич О. В.[1]
Литвинова Т. А.[2]

Анализ клавиатурного почерка в исследованиях письма: библиометрический обзор с использованием инструментов R и VOSviewer

[1] Белгородский государственный национальный исследовательский университет
ул. Победы, 85, Белгород, 308015, Россия
E-mail: dekhnich@bsu.edu.ru
ORCID: 0000-0001-6088-2656

[2] Воронежский государственный педагогический университет
ул. Ленина, 86, Воронеж, 394043, Россия
E-mail: centr_rus_yaz@mail.ru
ORCID: 0000-0002-6019-3700

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

93

Аннотация: Данные клавиатурного почерка, то есть длительность и количество нажатий клавиш, которые регистрируются с использованием программ-кейлоггеров, широко используются в различных областях знаний, в том числе для решения прикладных задач – от ранней диагностики когнитивных нарушений до идентификации пользователя. Также академические кейлоггеры и извлекаемые с их помощью данные активно используются в такой междисциплинарной области, как исследования письма (*writing research*), поскольку позволяют наблюдать за процессом создания текста, не вмешиваясь в него. Подобные исследования ведутся начиная с появлением общедоступных персональных компьютеров, то есть с конца 1990-х годов, однако библиометрический анализ таких работ, то есть анализ, направленный на выявление общей картины рассматриваемого предметного поля, определение его концептуальной структуры, наиболее активно исследуемых тем, наиболее цитируемых работ и авторов, до настоящего времени отсутствовал. Данная статья представляет собой первый опыт библиометрического анализа научных работ, направленных на анализ процесса создания текста с использованием данных кейлоггеров. Мы анализировали работы, индексируемые в базе данных Scopus и опубликованные в период с 2001 по 2024 год (273 работы было отобрано с использованием протокола PRISMA). В качестве инструментов исследования нами были выбраны библиотека на языке R bibliometrix и программа VOSviewer. С использованием данных инструментов нами были определены наиболее продуктивные и цитируемые исследователи, институты и страны, а также наиболее значимые работы. На основе различных методов анализа совместной встречаемости ключевых слов, реализованных в данных инструментах, нами была построена концептуальная структура предметного поля, а также определены исследовательские тренды. Значимость настоящей работы определяется двумя факторами. Во-первых, она содержит результаты первого библиометрического анализа важной и активно развивающейся предметной области, ограниченной, однако, с точки зрения географической принадлежности ведущих исследователей и языков, на материале которых проводятся такие исследования. Подобный анализ, как представляется, позволить ознакомить с состоянием работ в данной области тех ученых, для которых она является новой. Во-вторых, данное исследование представляет собой пример комбинации методов автоматического анализ научной литературы, которая может быть применена для анализа иных предметных областей.

Ключевые слова: Исследования письма; Клавиатурный почерк; Библиометрический анализ; Библиотека bibliometrix на языке R; VOSviewer; Анализ цитирований; Концептуальная структура предметного поля; Тематическое развитие предметного поля; PRISMA; Scopus; Ключевое слово.

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма…
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using…

94

АНАЛИЗ КЛАВИАТУРНОГО ПОЧЕРКА В ИССЛЕДОВАНИЯХ ПИСЬМА: БИБЛИОМЕТРИЧЕСКИЙ ОБЗОР С ИСПОЛЬЗОВАНИЕМ ИНСТРУМЕНТОВ R И VOSVIEWER

Проблема

Появление кейлоггеров, то есть программ, позволяющих следить за процессом письма, не мешая ему, существенно обогатило исследования письма (*writing research*) – междисциплинарную область со своей методологией

Отсутствуют систематические и библиометрические обзоры таких исследований

Решение

База данных *Scopus*, поиск (**keystroke** + **writing**) по названию + аннотации + ключевым словам (2001-2024)

протокол PRISMA

273 аннотации статей + метаданные

Библиотека на R bibliometrix + VOSviewer

ПЕРВЫЙ БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ ПРЕДМЕТНОГО ПОЛЯ

✓ Выделены наиболее продуктивные и самые цитируемые **страны/институты/авторы**
✓ Проанализированы наиболее релевантные
✓ **источники**
✓ Построена сеть **коллабораций**
✓ Определена **концептуальная структура** области
✓ Выявлены **основные** и **«нишевые»** темы
✓ Установлены **тренды** в развитии предметного поля

Область активно развивается:

✓ Начиная с 2019 года - резкий рост числа статей цитирований
✓ Всего **несколько стран – лидеров** по числу статей и цитирований, Бельгия – лидер по цитированиям, США – по числу статей
✓ Среди авторов – сотрудники EFL (TOEFL и др. продукты)
✓ **Низкий процент коллабораций**
✓ **Мало языков**, на материале которых проводятся исследования

БОЛЬШЕ ИССЛЕДОВАНИЙ + СТРАН + ЯЗЫКОВ ДЛЯ ДАЛЬНЕЙШЕГО РАЗВИТИЯ ОБЛАСТИ

1. Introduction

Typing has been the most comfortable and typical way to produce texts for many people, that is why studying typed texts and the process of their creation has become one of the important research directions in modern linguistics. The development of technical devices such as keystroke loggers, which timestamp each event (i.e. keystroke) during writing, has provided researchers with previously unavailable data reflecting different aspects of writing processes.

The use of keystroke logging data is typical not only to study writing process in linguistics but also for many practical applications, especially for user authentication and identification. The search for the keyword "keystroke" in the Scopus database, carried out in July 2024, yields more than 1700 papers – search by title, keyword and abstract – more than 4000 papers published in journals and proceedings from various subject areas – from computer science (2755 papers) to arts and humanities (365) and medicine (306).

There are several review papers which analyze different aspects of keystroke behavior: as a tool for cognitive impairments detection (Lee and Park, 2024), a digital biomarker for understanding neurocognitive functioning (Nguyen et al., 2023), motor decline (Alfalahi et al., 2022), as a tool for emotion recognition (Maalej and Kallel, 2020; Yang and Qin, 2021), user authentication (Quraishi and Bedi, 2018; Raul et al., 2020).

However, despite the rise of the number of papers which use keystroke data to study writing, to the best of our knowledge, there are no systematic or bibliometric reviews of these papers which would consider this actively developing research area from a holistic perspective with identification of the main topics and research trends.

The aim of this paper is to map the literature devoted to the study of writing using keystroke data through a bibliometric approach in order to reveal knowledge structures, priority topics and trends in this area of research.

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

95

We performed a bibliometric analysis of the papers considering the keystroke logging method for study writing searched in Scopus using the R package bibliometrix (Aria and Cuccurullo, 2017) and the standalone application VOSviewer (Van Eck and Waltman, 2014). This bibliometric analysis of 273 articles published over the last 24 years (2001-2024) is the first one to present comprehensive and statistical insight into the research status and trends on the use of the keystroke methodology to study writing.

The study's findings may offer broad perspectives on the current state of this research area and provide some guidance for future research.

Therefore, this study was driven by the following research questions.

1. What was the annual research trend in the considered subject field?

2. Which journals, countries, institutions and authors were the major research contributors?

3. What were the main research topics of the topical collection?

4. What are the trends in this research area?

2. Research methods

2.1. Transparency Statement

To address our research questions, we undertook a bibliometric review to map out the existing literature on our selected topic – using keystroke data and software to study the writing process. We adhered to the PRISMA protocol for conducting this review. A protocol was outlined before commencing the study. We ensure that this manuscript presents a truthful, precise, and complete report of the research conducted.

2.2. Eligibility Criteria

This bibliometric review was executed through a structured process that included: 1) formulating the research question; 2) identifying relevant literature; 3) selecting appropriate studies for inclusion; 4) extracting key data from these studies and data cleaning; 5) summarising and presenting the findings.

The selection criteria for the literature were divided into several categories for framing research questions in bibliometric reviews (detailed in Table 1).

Table 1. Eligibility criteria
Таблица 1. Критерии отбора

Criterion	Inclusion	Exclusion	Justification
Population	All the studies which use keystroke data and software to study the writing process	All the studies outside the defined field	As the research topic we analyze is multidisciplinary, we include papers from different subject areas related to the topic of writing and its analysis using keystroke data/software
Concept	Studies (journal articles, book chapters, conference proceedings) which discuss different aspects of the writing process based on keystroke data	Studies which do not consider the concept of the writing process	The purpose of the review is to analyze literature which is related to the keystroke data/software for the writing process understanding
Context	Studies that relate to	Studies outside	Focus of this review is

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

96

	writing research	writing research	on the use of the keystroke data/method to study writing
Language	English	Non-English	English is a lingua Franca of scientific communication
Time period	2001-2024	Before 2001	Keystroke logging software has become widely used in writing studies after the appearance of available personal computers
Types of sources	1. Journal articles 2. Book chapters 3. Conference papers	1. Non peer-reviewed sources 2. Reviews (all types) 3. Editorials	Only peer-reviewed sources and original research papers were included in our analysis
Geographical location	Any location	None	Getting international perspective

We included all study designs from any discipline found in our search, reported in English. Special attention was given to the quality of the sources included in review: we considered only original research and peer-reviewed papers.

2.3. Information Sources and Search Strategy

Scopus, which is considered the largest citation and abstract database covering a wide range of subjects was chosen as the data source for this study. Its particular suitability for bibliometric analysis (consistently reliable format, and ease of data extraction) has been highlighted by many researchers (see ref. in Rejeb et al., 2023). It is also more comprehensive than, for example, PubMed or Web of Science (WoS) and more precise than Google Scholar because of its superior control over content, that is why it is widely used for bibliometric analysis (see ref. in Rejeb et al., 2023).

Access to Scopus was performed using the personal account of one of the authors who was granted access to Scopus as part of the reviewer recognition process.

Figure 1 illustrates our compliance with the PRISMA protocol.

The following search string was used in the article's title, abstract and keywords fields to extract data from Scopus: "keystroke*" AND "writing".

According to the information of Scopus, the first article which meets the search criteria was published in 1974. However, a wide use of the keystroke methodology for writing research started after the development of freely available keystroke loggers, i.e., after 1990s (Wengelin and Johansson, 2023). In addition, only since the early 2000s, the academic landscape has been reshaped by a global trend toward requiring publications in prestigious journals indexed in international citation databases, therefore, we decided to refine our search by excluding papers published before 2001.

The data collection process for this study started on July 11, 2024; thus, we set the scope of this research is from 2001 to 2024.

From this search, 336 studies were retrieved (as shown in Figure 1).

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

97

Figure 1. PRISMA flow diagram
Рисунок 1. Диаграмма отбора работ для анализа по протоколу PRISMA
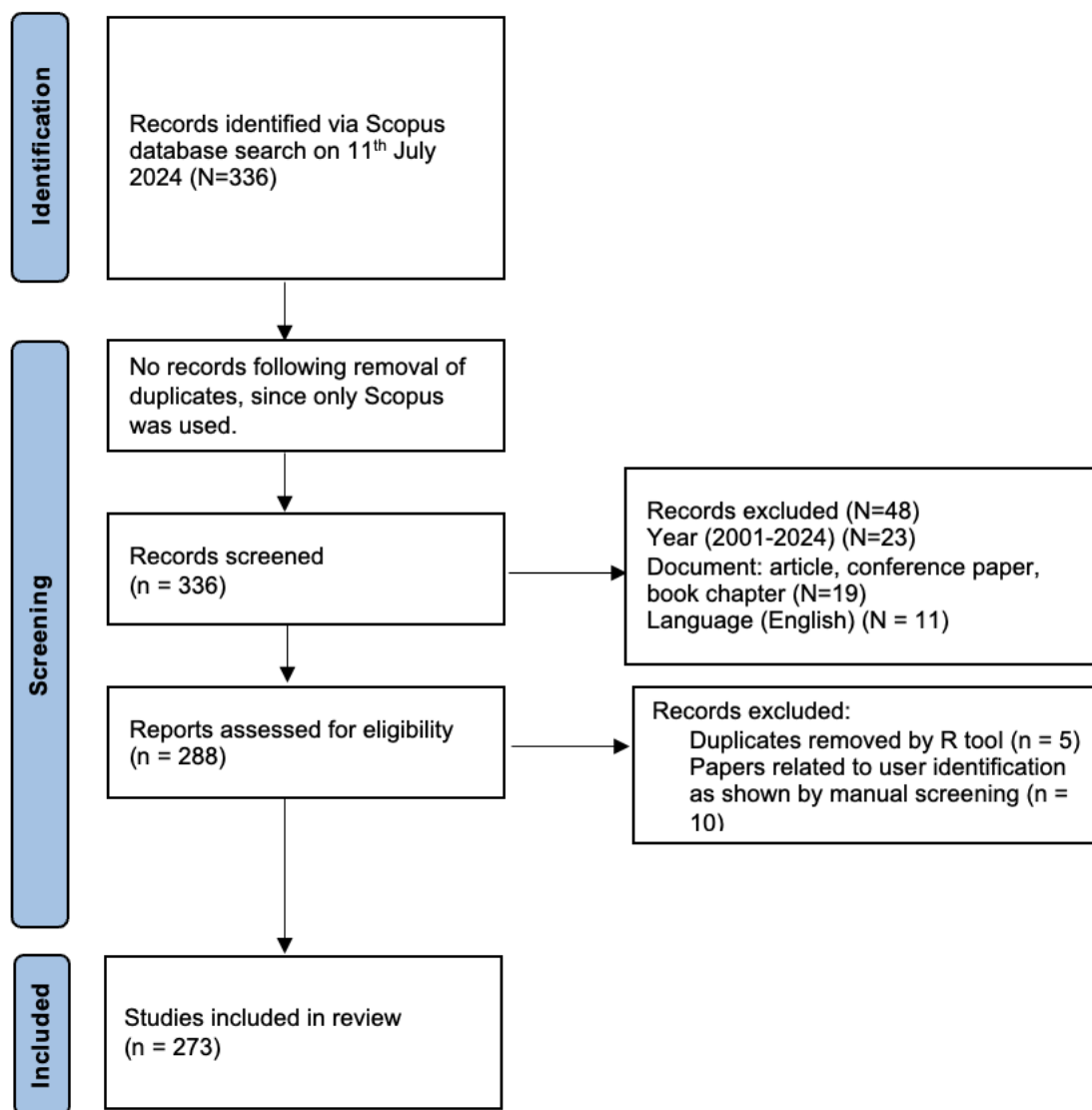
2.4. Selection of Sources of Evidence

For further analysis, all available bibliographic information for 336 retrieved records were extracted from Scopus and exported in CSV format. Through the initial screening of titles and abstracts, performed by two reviewers, 48 studies were excluded based on the predefined inclusion criteria: we excluded papers published before 2001 (N=23), written in languages other than English (N=11), and of any other types besides "article", "conference paper", "book chapter".

Then we removed duplicates detected by bibliometrix (N=5). Next, remaining titles and abstracts of the publications were manually evaluated for relevance in terms of context and concept compliance. After this screening, 10 records were removed since they did not contribute to the field of writing studies, i.e., their findings were not related to the writing process understanding (we noticed that they are related to user authentication).

Thus, 273 papers were selected for the final analysis (see Figure 1).

Data Charting Process

Data extraction was conducted by two independent reviewers, with each being assigned all the selected documents for initial analysis. The complete data set extracted by

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

98

one reviewer was then cross-checked by the other to ensure accuracy and consistency. Any differences between reviewers were addressed and resolved through consensus meetings.

Data analysis and visualisation

The metadata of the articles included in the review were processed using VOSviewer, a software tool designed for constructing and visualising bibliometric networks and the bibliometrix R-package.

Bibliometric analysis involves two main approaches: performance analysis and science mapping. The goal of performance analysis is assessing the contributions of research constituents to a given field (it relies upon such metrics as total number of publications, citations, etc.), while science mapping examines the relationships between research constituents via such methods as co-citation analysis, bibliographic coupling, co-word analysis (Donthu et al., 2021).

Bibliometrix is far less popular than VOSviewer, however, it provides a wide range of statistical techniques and visualizations useful for determining and visualizing the conceptual structure of the field. Bibliometrix enables users to conduct both performance analysis and science mapping, whereas users are only able to perform science mapping in VOSviewer. However, we combined VOSviewer with bibliometrix to get a deeper understanding of the research area under consideration, since they provide different types of analysis and visualizations for science mapping. Many researchers have used them simultaneously (Saini et al., 2022; Wang and Gu, 2022).

3. Results and Discussion

3.1. Performance analysis

3.1.1. General performance characteristics of the research area

Our dataset consists of 273 papers written by 520 authors and published in 152 sources. The annual growth rate is 15.22 %, i.e., the research in the analyzed field has attracted continuous attention of scholars. Figure 2 shows a strong increase in the number of papers from 2019 onwards (from a maximum of 15 to 26 in 2019, and at least 26 after 2019), which is an important index indicating a development trend of a research topic in a specific period of time (Farrukh et al., 2023).

Figure 2. Annual scientific production
Рисунок 2. Распределение числа публикаций по годам

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

99
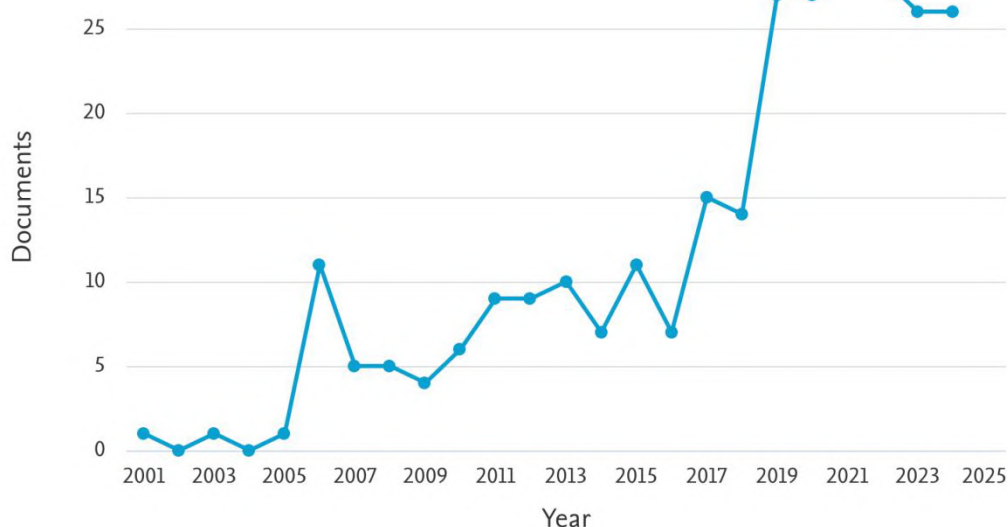
It is noteworthy that by mid-2024, 26 papers published in 2024 were already recorded in Scopus, so we can expect a further increase in the number of papers this year.

Most of the works in our collection (64 %) are articles, conference papers constitute about 24 %, the rest of the papers are book chapters.

The average citations per doc is 14.7, the average citations per year per doc is 1.906, the overall number of citations of works in our collection is 4110.

Figures 3-4 show that the burst of average citation count per article per year falls on the papers published in 2013 (they were cited around 6 times). This year, the most cited paper in our collection has been published – *Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes* (Leijten and Van Waes, 2013) (319 citations) which presents the results of analysis of the writing process with one of the most popular keyloggers designed specially for writing research – Inputlog with other observational techniques: thinking aloud protocols and eye-tracking data.

Figure 3. Average article citations per year
Рисунок 3. Усредненное число цитирований статей, изданных в данном году (на статью)
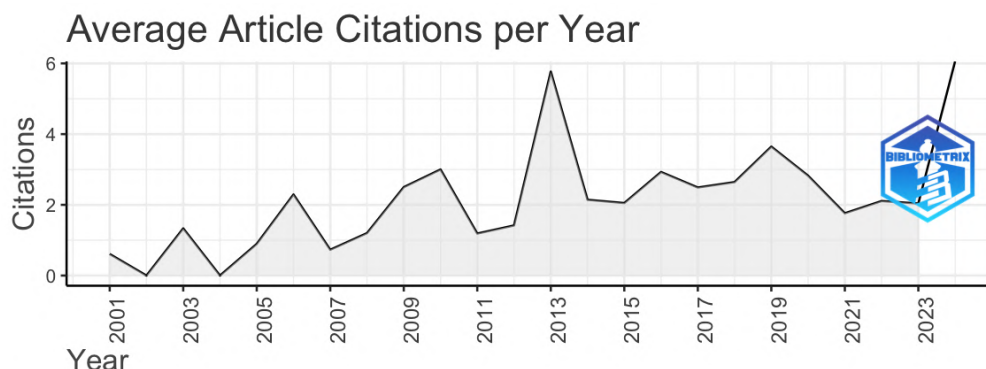
The mean total article citations per year at the highest level again can be seen for the papers published in 2013 (N= 63), which indicates that the articles published in 2013 were cited the maximum number of times although the number of articles in 2013 was rather low (N= 9). As we mentioned earlier, in 2013 the most cited paper of the collection was published. Along with it, in the top-10 most cited papers of the collection (Figure 5) there are two papers published during this year with the total of 159 citations. These papers are related to the constructing of documents using multiple digital sources (Leijten et al., 2013b) and detecting boredom and engagement during writing (Bixler and D'Mello, 2013).

The other peaks in the average total citations per year fall on 2006 and 2010 when the other papers from the top-10 most cited were published: the 2006 paper *Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL* (Stevenson et al., 2006) with 106 citations, 2006 paper *Examining Pauses in Writing: Theory, Methods and Empirical Data* (Wengelin, 2006) with 91 citations and a 2010 paper *Cognitive illusions of authorship reveal hierarchical error detection in skilled typists* with 109 citations (Logan and Crump, 2010) (2nd most cited paper in the collection). It is also should be noted that the papers published in the last years have also been actively cited.

The documents in our collection have been increasingly cited since 2019 (see Figure 6).

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

100

Figure 4. Average total article citations per year
Рисунок 4. Усредненное общее число цитирований статей, изданных в данном году (по годам)

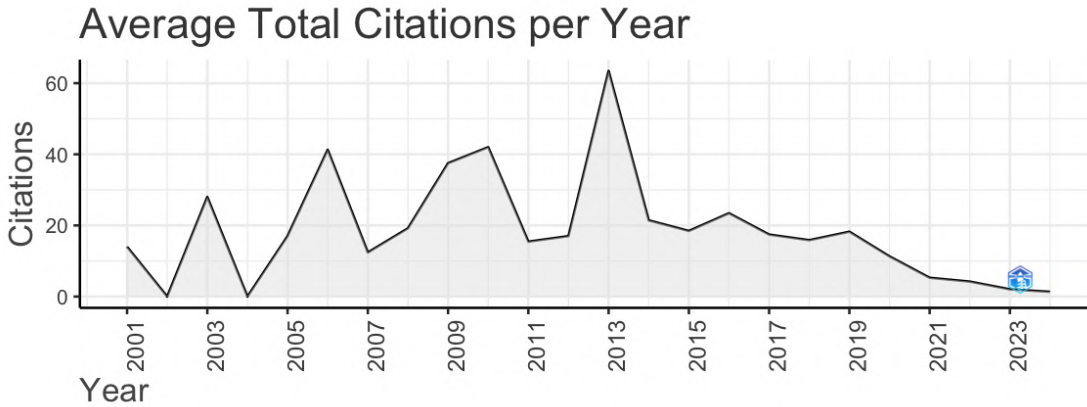**Figure 5.** Most cited papers of the collection
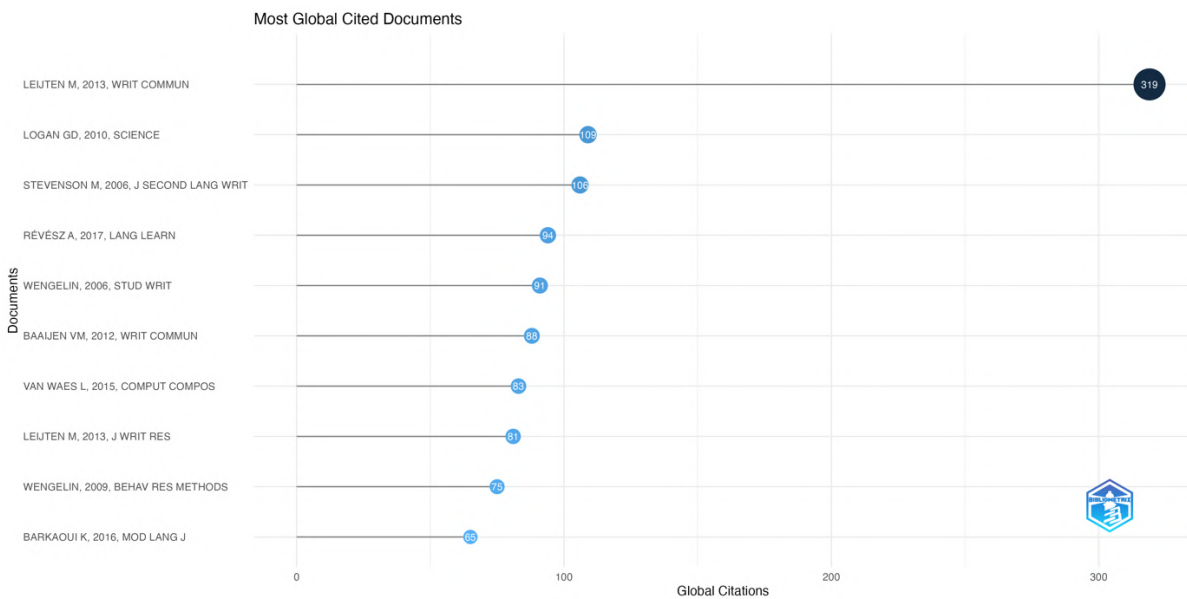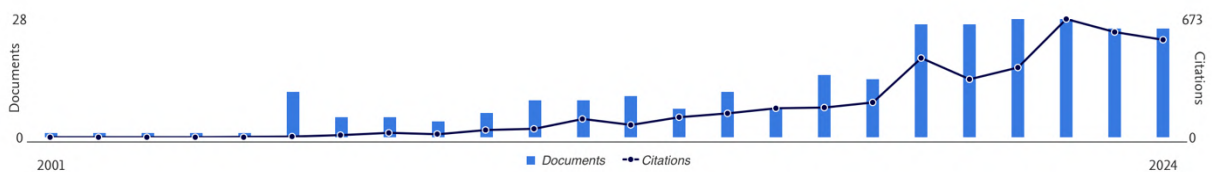**Рисунок 5.** Наиболее цитируемые работы

Figure 6. Number of papers and total number of citations by year
Рисунок 6. Число статей и цитирований по годам публикации

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

101

Overall, we see a general trend of increasing numbers of papers from 2019 onwards, and the papers in our collection have been actively cited since this year. We can safely conclude that the use of keystroke analysis in the study of writing is an area of research which is in a state of active development.

3.1.2. Most productive countries, institutions and authors Figure 7 shows the most productive countries in terms of the number of corresponding authors affiliated with these countries (this visualisation is constructed using the bibliometrix package).

Figure 7. Most productive countries (according to the corresponding author)
Рисунок 7. Страны с наибольшим числом статей (по корреспондирующему автору)

Country Scientific Production
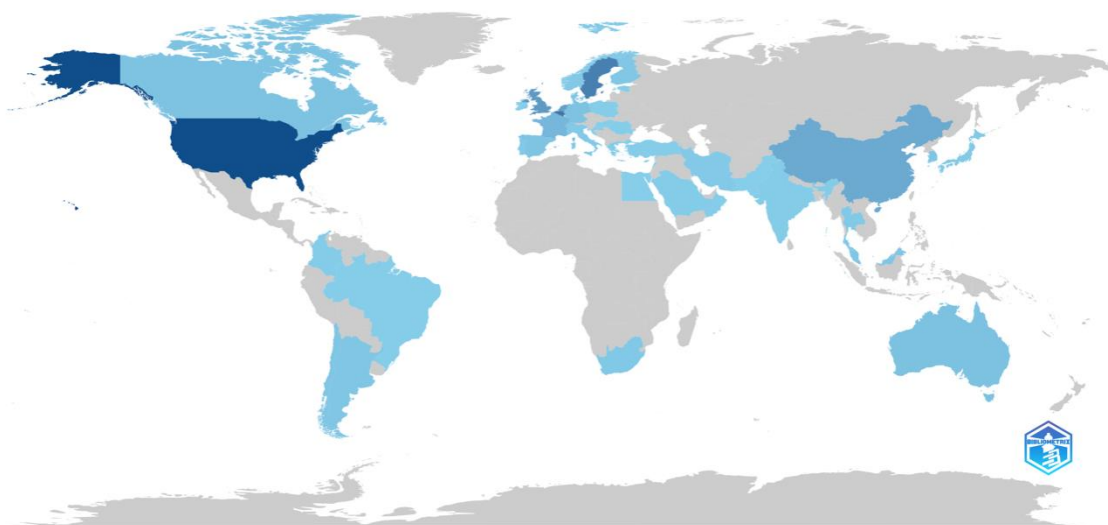
The USA is the leading country in terms of total number of papers (38 papers) and percentage of single-country publications (about 80%). **Sweden** is the second most productive country with 23 papers, 75% of which are single-country publications. **Belgium** is the third most productive country with 19 papers (12 of which are single-country). Authors from **the Netherlands** published 14 papers, but most of them (12) are multi-country publications. Researchers from **China** and the b published 13 papers, of which 9 and 7 were single-country papers respectively. The number of papers related to other countries is rather low (**Canada** – 5, **France** – 5, **Spain** – 4, **Switzerland** – 4, **Thailand** – 4, others – 3 or less).

Figure 8 shows the network of the country collaboration.

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

102

Figure 8. Country collaboration in the research area under consideration
Рисунок 8. Сеть стран, имеющих коллаборации друг с другом в рамках обозреваемого предметного поля
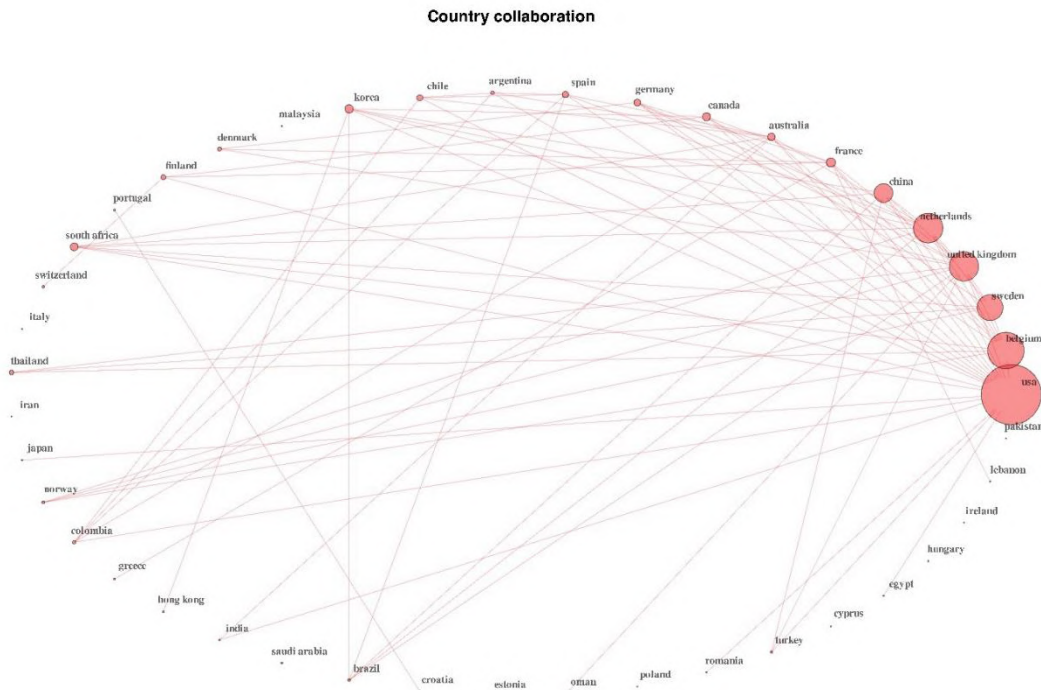
The network consists of the nodes representing a country, and the connection between the nodes represents a cooperative relationship between the countries, and the size of the nodes represents the numbers of cooperation with other countries. The larger the node is, the more international the cooperation the country participates in is. We can see that the USA has the largest diversity in terms of cooperation with other countries. Belgium, the Netherlands, the UK and Sweden are the countries with a moderate level of diversity of international cooperation.

Figure 9 shows the most productive institutions, with University of Antwerp (**Belgium**) being the leader. The second most productive organization is the Educational Testing Service (ETS), which is the world's largest private educational testing and assessment organization which developed TOEFL and other products with its headquarters in the US. The third most productive organization is Lund University (Sweden).

The analysis of the dynamics of affiliation production shows (not shown due to lack of space) that the University of Antwerp has the most pronounced trend in increasing the number of papers starting from 2009, with a sharp increase in 2021. A continuously high number of papers is affiliated to Lund University from 2010 onwards.

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

103

Figure 9. Most productive institutions (according to the corresponding author)
Рисунок 9. Организации с наибольшим числом статей (по корреспондирующему автору)
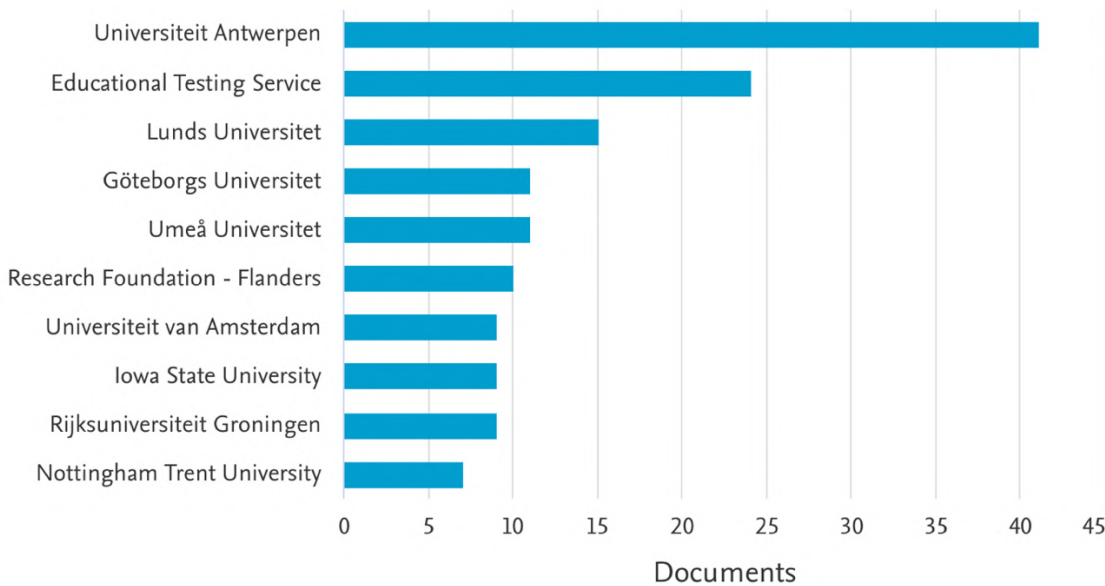
If we look at the country production over time (see Figure 10), we can easily observe a general trend of an increase in the number of papers for all the most productive countries. In the US, this increase occurred earlier (from 2015) than in the other countries (since 2019) and is a sharper one.

Figure 10. Country production over time
Рисунок 10. Распределение числа статей по странам (по годам)

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма…
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using…

104
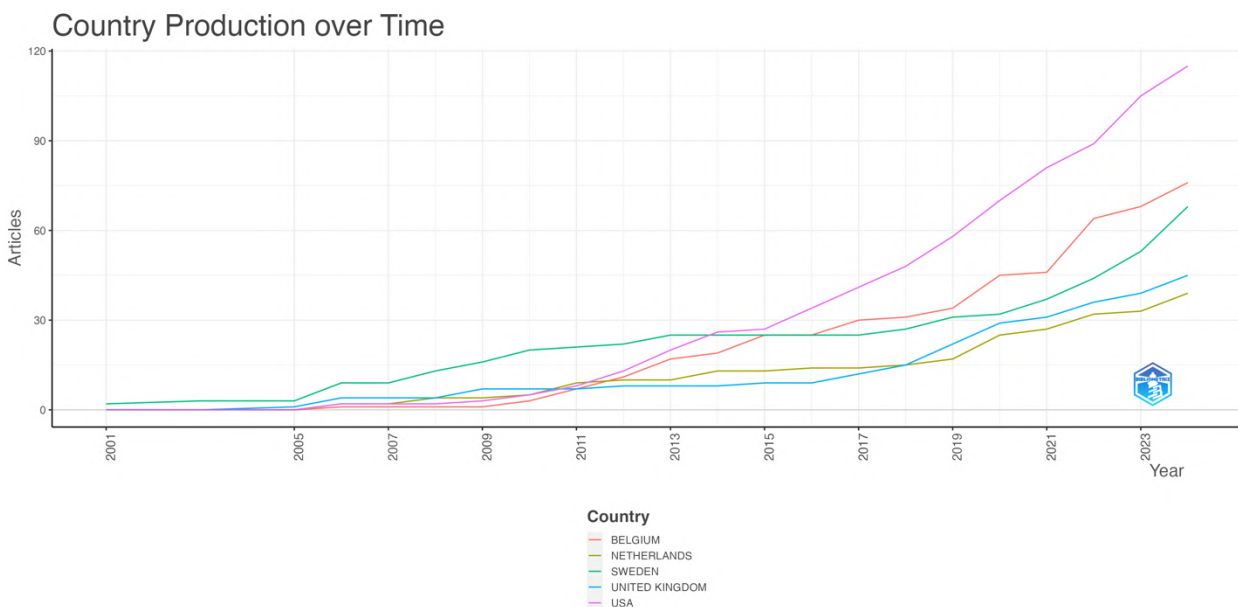
Although the USA is the most productive country, it ranks 2nd in terms of the number of citations (580), with Belgium being the leader with 666 citations. The Netherlands is the third one in this rank (Figure 11).

Figure 11. Most cited countries
Рисунок 11. Страны с наибольшим числом цитирований

Most Cited Countries
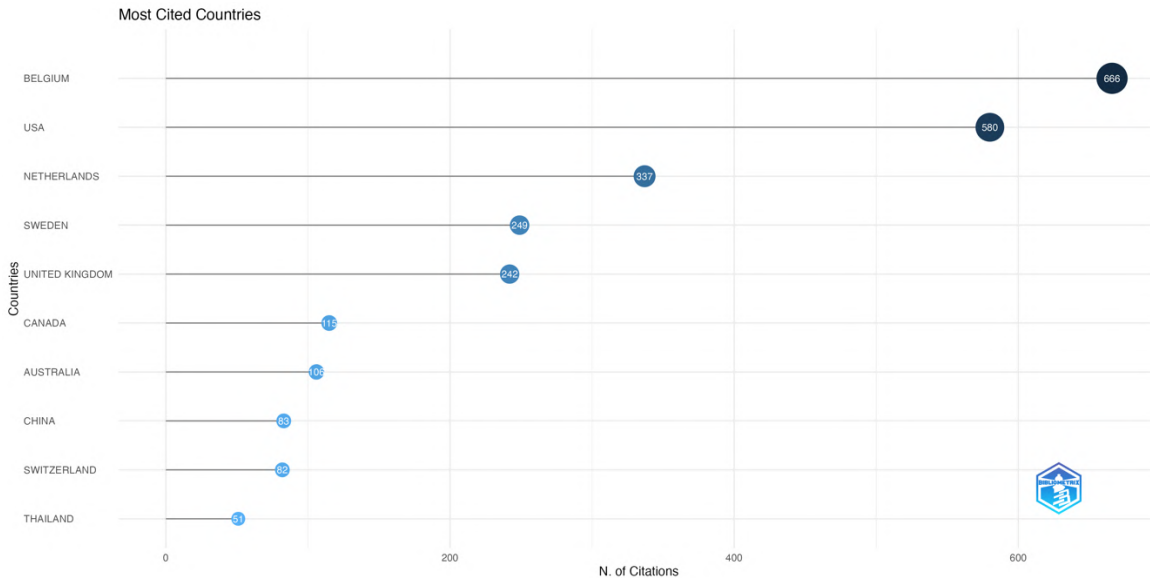
There are 522 different authors in our collection, with 36 authors of single-authored document and 866 author appearances. There are 44 single-authored papers, on average, there are 3.16 co-authors per document.

The most productive author, who published 27 papers, is **Luuk Van Waes** from the University of Antwerp, Belgium. The second most productive author, **Mariëlle Leijten**, is affiliated with the same university. **Åsa Wengelin**, the third productive author, is affiliated with the University of Gothenburg (Sweden). Mo Zhang, the author of 17 papers, is affiliated with the Educational Testing Service (USA), just like Paul Deane, her co-author. Luuk Van Waes and Mariëlle Leijten also have the highest h-index among the most productive authors (14), Paul Deane and Åsa Wengelin have h-index=10, Mo Zhang – 9. We can see that the productivity correlates with h-metrics in our collection.

3.1.3. The analysis of the most relevant sources

Figure 12 shows the distribution of sources by subject area. We can see that social sciences, arts and humanities, computer science and psychology are the leading subject areas of articles and proceedings in which papers from our collections have been published.

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

105

Figure 12. Paper distribution over a subject area
Рисунок 12. Распределение публикаций по предметным полям
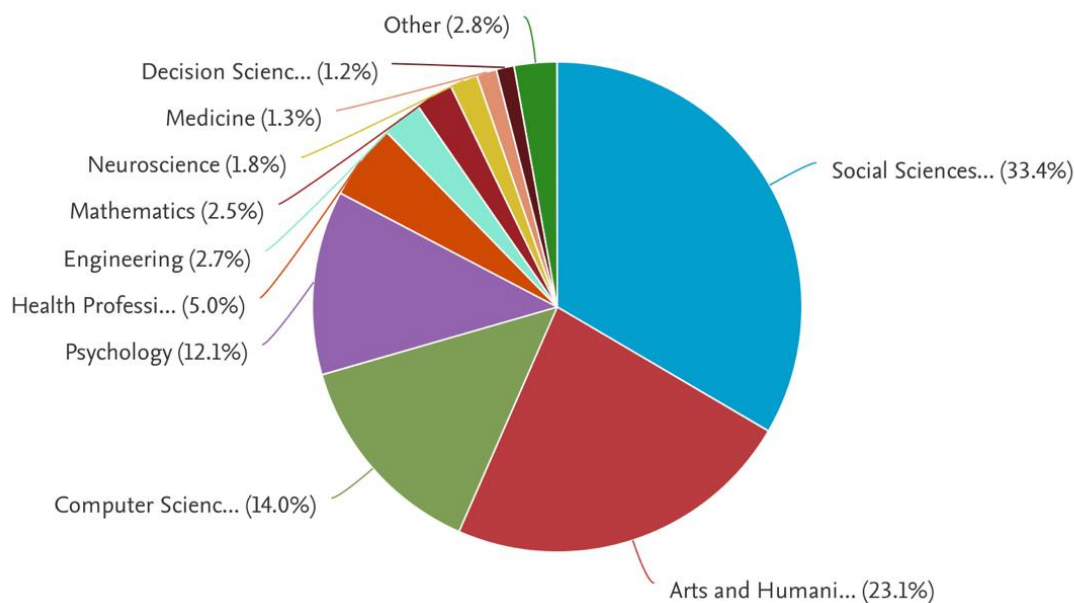
The most relevant sources in terms of the number of published papers from our collection is Reading and Writing (17 papers), Studies in Writing (17 papers), Journal of Writing Research (13), Written Communication (8), Frontiers in Psychology (7), Lecture Notes in Computer Science (7), Computers and Composition (6), ETS Research Report Series (5), Journal of Second Language Writing (5), Languages (5).

All these journals are highly reputable. The journals with the highest impact (h-index) are Studies in Writing (11), Journal of Writing Research (8), Reading and Writing (6).

Looking at the production of sources over time (Figure 13), we could see that Studies in Writing has been continuously publishing papers on the study of writing using keystroke since 2007, while the other most productive sources show a general trend of an increasing number of such papers after 2019.

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

106

Figure 13. Source production over time (over the years)
Рисунок 13. Распределение статей по источникам (по годам)

Sources' Production over Time

Source
FRONTIERS IN PSYCHOLOGY
JOURNAL OF WRITING RESEARCH
LECTURE NOTES IN COMPUTER SCIENCE (INCLUDING SUBSERIES LECTURE NOTES IN ARTIFICIAL INTELLIGENCE AND LECTURE NOTES IN BIOINFORMATICS)
READING AND WRITING
STUDIES IN WRITING
WRITTEN COMMUNICATION
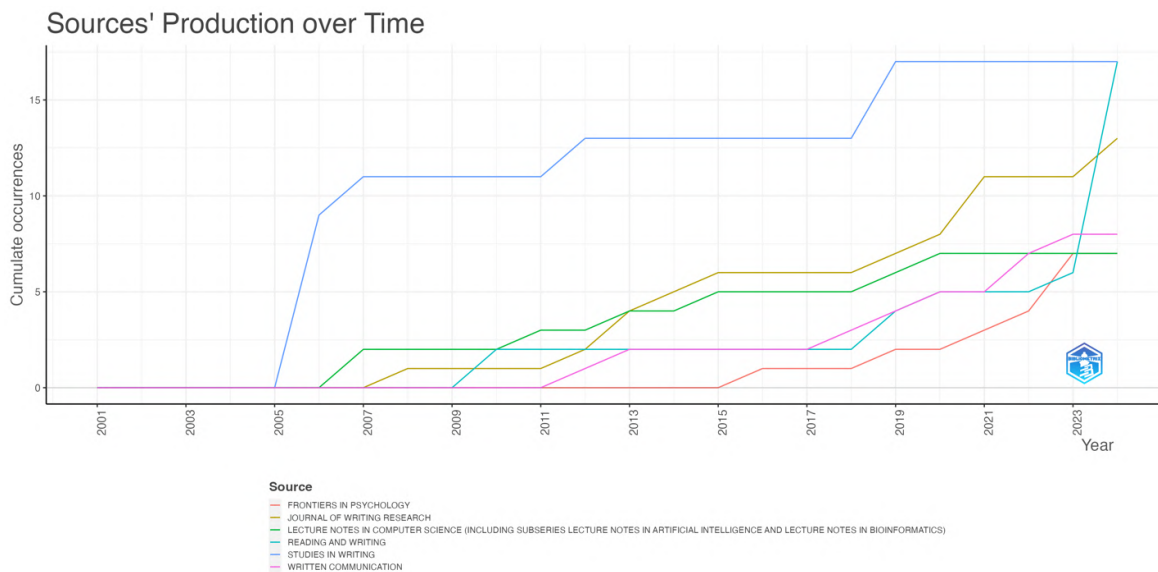
3.2. Science mapping

3.2.1. Keyword co-occurrence analysis

Keyword analysis is important in a bibliometric review as it could provide many insights about the conceptual structure of the field. Keywords are divided into authors' keywords and those which are assigned automatically (they are called "keyword plus" in Scopus). There are 801 "keyword plus" in our dataset (the most frequent ones are WRITING – 53 papers, HUMAN – 28, MALE – 26, FEMALE – 25, WRITING PROCESS – 19, STUDENTS – 18), 750 author's keywords (the top 100 most frequent ones are presented in the form of a word cloud in Figure 33). As a rule, authors' keywords are more specific, so they will be used in the further analysis.

Figure 14 gives us a snapshot of the content of our collection (the bigger the size of the circles is, the more frequent the keyword is), with KEYSTROKE LOGGING, WRITING PROCESS, WRITING, COGNITIVE PROCESSES, REVISION, PAUSES, NATURAL LANGUAGE PROCESSING are the most relevant author's keywords.

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

107

Figure 14. Word cloud of 100 most frequent author's keywords of the collection
Рисунок 14. Облако слов, отражающее встречаемость 100 самых частотных ключевых слов автора в коллекции
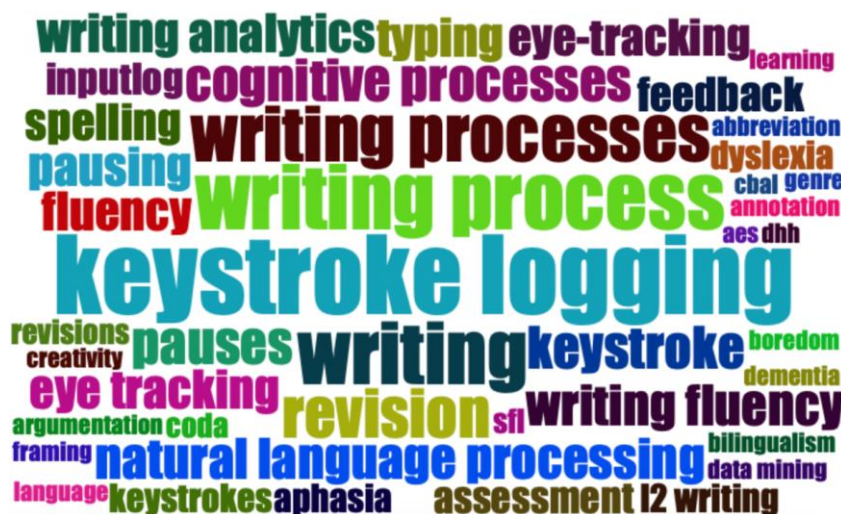
Co-word analysis, which reveals words that frequently co-occur in documents, can be used to reveal important themes (both established and emerging) within the focal domain. Using VOSviewer, we analyzed co-occurrences of authors' keywords and visualized them in the form of a network (Figure 15) (we only included words that appeared at least 10 times in our dataset). Keywords with similar colors belong to the same cluster. The size of each circle in the cluster shows the proportion of citations for that authors' keywords. Larger circles and map labels indicate a greater significance and relevance of the keywords.

We used association strength as a method of normalization and set a minimal cluster size to 10.

Cluster 1 (**red**) is the largest one in terms of the number of terms (24). It is related to the automated text assessment including second language writing using keystroke logs and various modern approaches such as scenario-based assessment (SBA), which have recently been used to assess writing skills.

Cluster 2 (**green**, 22 terms) is related to the analysis of typed texts produced by people with mild to severe (Alzheimer, dementia) cognitive problems based on keystroke analysis, especially revision process analysis,

sentence history analysis, assessment of motor and typing skills using techniques for automated scoring for writing quality based on the deep learning method and AI in general. Methods of mixture models have widely been used in such papers to distinguish between the healthy and the ill respondents in terms of pauses duration.

Cluster 3 (**dark blue**, 21 items) is related to the study of academic writing, including writing in a second language and using an analysis of revision via keystroke logs in combination with eye-tracking. The problems of writing tasks complexity and writing performance are discussed in such papers.

Cluster 4 (**greenish yellow**, 21 items) is associated with the study of such aspects of writing as pause behavior, revision behavior in people with reading and writing difficulties (including those suffering from aphasia).

Cluster 5 (**purple**, 16 items) is related to the study of cognitive load and cognitive process in L2 writing (including writing from multiple sources and workplace writing) via analysis of revisions and pauses analysis using the Inputlog software. The problems of error correction are also discussed in such papers.

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

108

Cluster 6 (**light blue**, 16 items) is related to the problem of the analysis of the writing process using a set of methods (keystroke logging, screencasting, eye tracking) with different purposes including tracking the states of writers (boredom, engagement) with possible writing intervention depending on the detected state. The problems of writing analytics are discussed.

Cluster 7 (**orange**, 12 items) discusses the problems of the characteristics of written product via an analysis of typing of people with dyslexia, deaf and hard of hearing (DHH), children of deaf adults (CODA).

Figure 15. Clusters of authors' keywords
Рисунок 15. Сеть совместной встречаемости ключевых слов автора
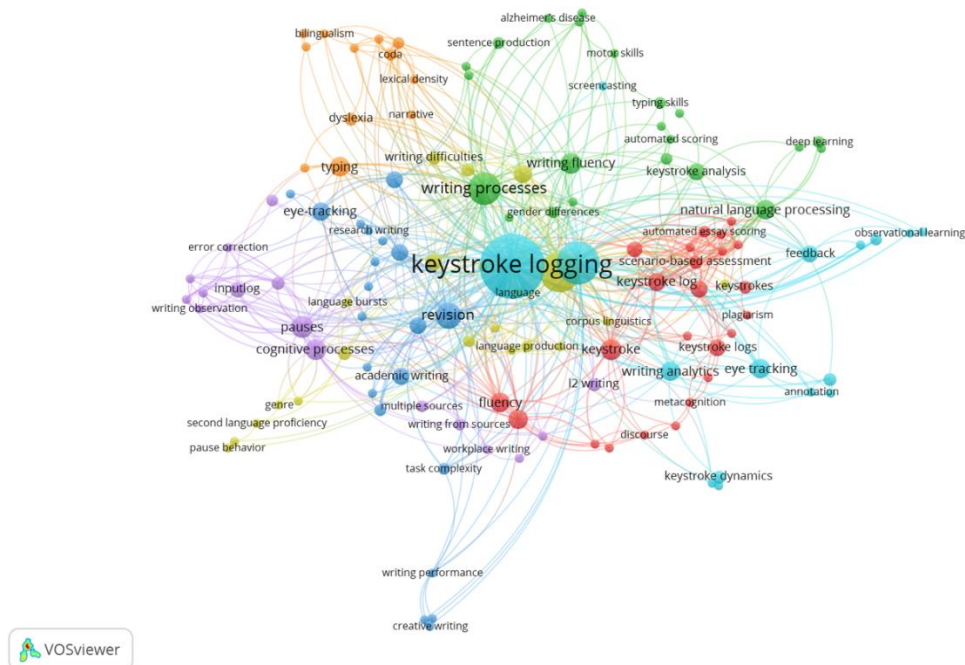
However, this picture is static. The bibliometric tools used in this study allowed us to reveal the research trends using a different set of methods. The basic method is the analysis of dynamics of keywords occurrences over time (Figure 16). For example, we can see that natural language processing methods have been actively used over the last ten years, with a peak in the last year. Writing analytics and writing assessment are the recent topical trends. Inputlog was often mentioned in the papers published in 2010-2015, with a peak in 2013.

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

109

Figure 16. Dynamics of keywords over the years
Рисунок 16. Динамика частотности ключевых слов по годам
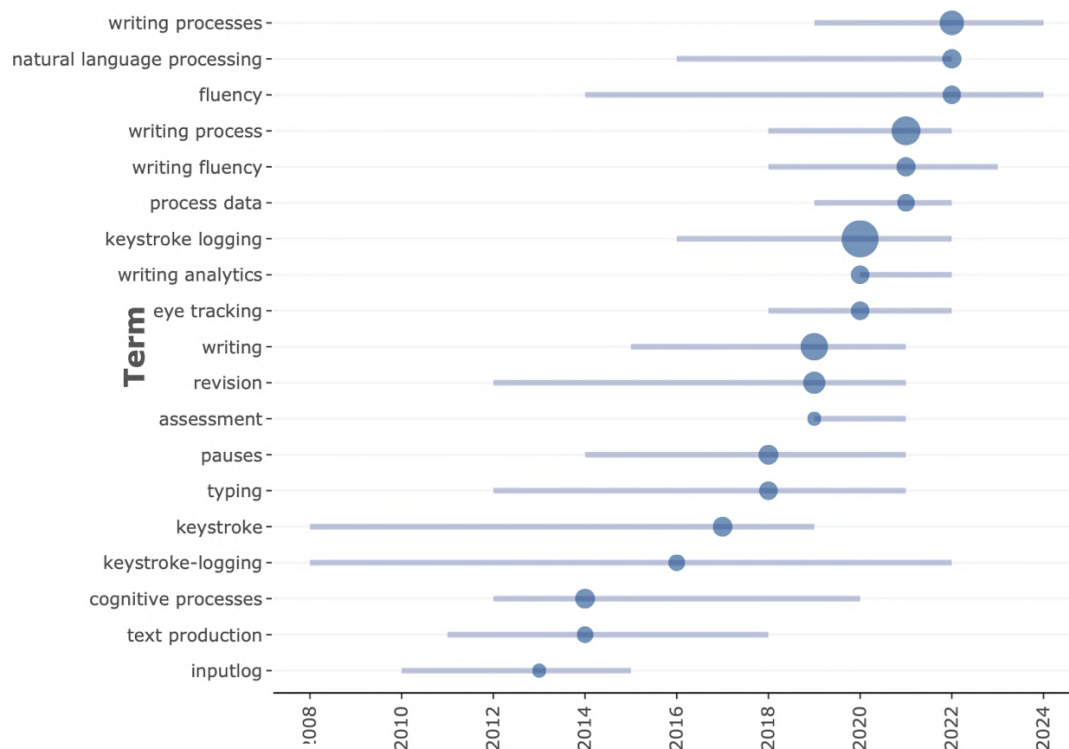
But it is more interesting to perform a co-word analysis and draw clusters of keywords in the form of a thematic map (Figure 17). This type of visualization implemented in bibliometrix is a very intuitive plot: themes are classified according to the quadrant in which they are placed: (1) upper-right quadrant – motor-themes; (2) lower-right quadrant – basic themes; (3) lower-left quadrant – emerging or disappearing themes; (4) upper-left quadrant – very specialized/niche themes (Büyükkıdık, 2022; Derviş, 2019; Rashid, 2023).

The basic themes of our collection are the general problems of the study of writing processes based on keystroke data, analysis of revisions in writing, analysis of writing fluency. Among the basic themes there is a combination of keystroke logging and eye-tracking to study cognitive effort of the writers, the problems of studying writing in a second language including Chinese, problems of plagiarism detection via keystroke analysis. The motor themes are the study of pauses and cognitive processes between them using

Inputlog, the analysis of metacognition in second language writing, problems of linking the textual quality of writing as a product with the characteristics of writing as a process, the use of NLP and data science methods for writing analysis.

The niche themes are the development of the methods for code completion using keystroke data, construction of text history, analysis of pause and revision behavior with connection of second language proficiency, writing intervention as part of writing pedagogy, analysis of keystroke on smartphones, the use of deep learning methods for natural language generation.

The declining themes are the study of workplace writing from multiple sources. The emerging themes are the study of writing tasks complexity based on keystroke data, the use of machine learning for revision analysis, application of mixture models for pause analysis and assessment of typing skills.

The research trends could also be revealed via overlay visualization offered by VOSviewer (Figure 18).

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма…
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using…

110

Figure 17. Thematic map of the collection
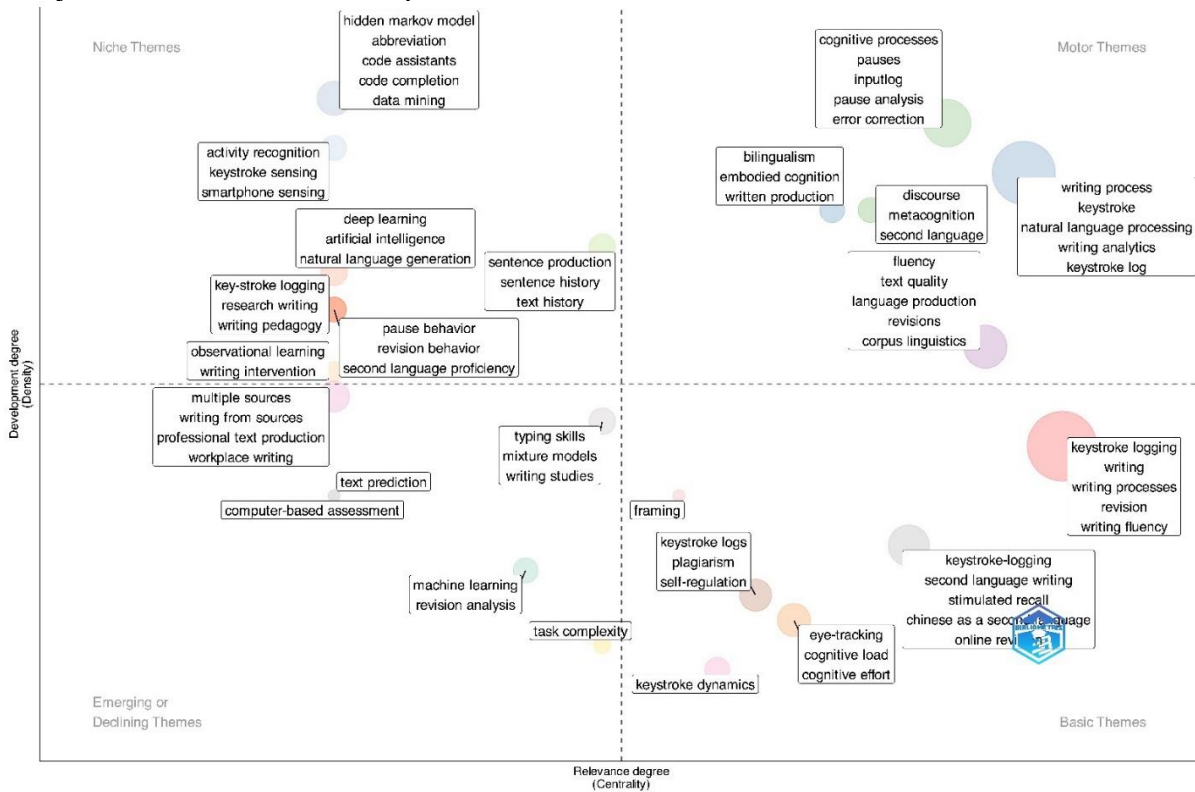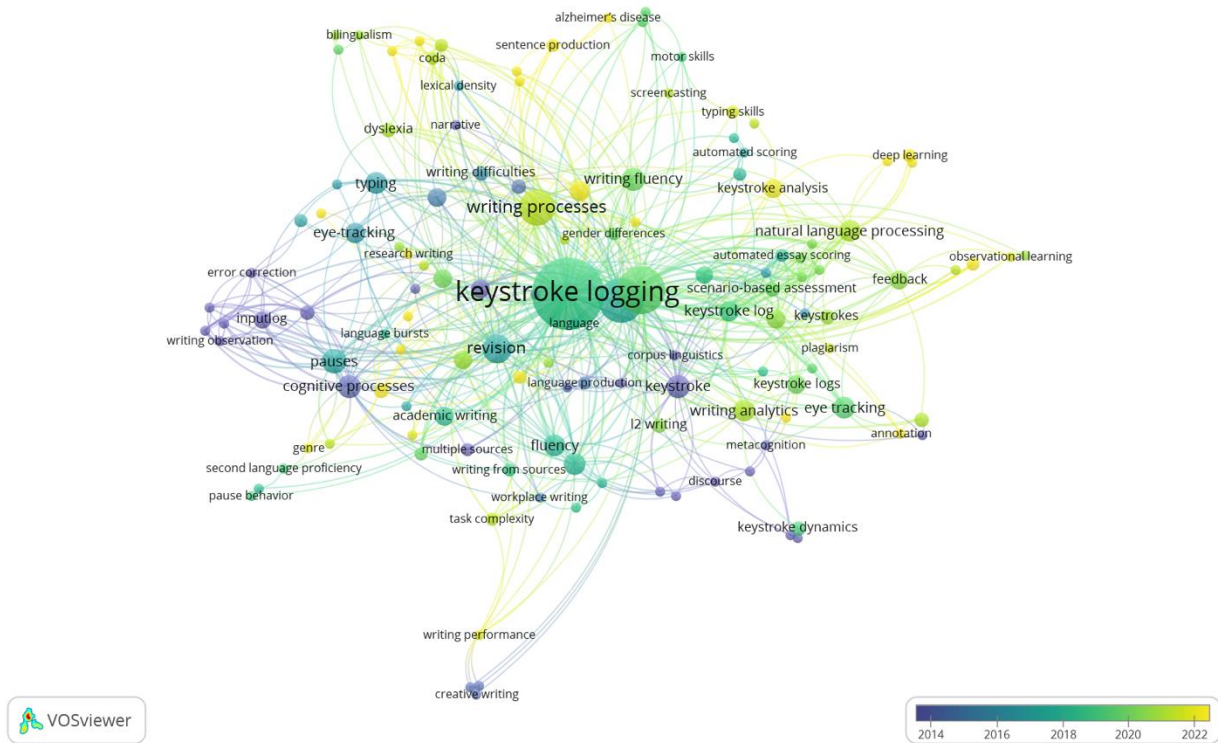**Рисунок 17.** Тематическая карта коллекции

Figure 18. Overlay visualization of domain based on an authors' keywords co-occurrences
Рисунок 18. Оверлейная визуализация совместной встречаемости ключевых слов

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

111

In the overlay visualization, the color of a node indicates a certain property of the node. In Figure 6, the overlay color of each circle corresponds to the average publication year of all the papers that include the corresponding term. In this map, the terms with cold colours (e.g. blue) represent research activities with an older average publication year, and the terms with hot colours (e.g. red) show terms with a more recent average publication year. Overlay visualization of domain was constructed on an authors' keywords co-occurrences (association strength was used for normalization which is a default method, weights = occurrences, min strength=1).

The general conclusions about the topic evolution offered by VOSviewer are in line with those revealed by the thematic map. Analysis of cognitive processes during writing, the study of metacognition, error correction, use of Inputlog to study writing were at the centre of research attention at the first stage of analysis, while in the recent years the problems of task complexity assessment, writing analytics, use of deep learning in writing research, analysis of typing skills, observational learning have become hot topics.

3.2.2. Co-citation analysis

Co-citation analysis uncovers the foundations of knowledge relied on by scientific research in the field, wherein clusters are formed based on the cited documents (references) that frequently appear together.

The clusters of the journals with strong co-citation relations are shown in Fig. 19. Three large clusters and a small one are revealed. The largest cluster – **green** – consists of the journals related to the general problems of writing research and writing assessment. The **red** cluster is made up of psychology and cognitive science journals, the **blue** cluster is made up of language learning journals, and the small **yellow** cluster is made up of journals related to writing by people with disabilities. This cluster solution shows different directions of writing research using keystroke data and methodology.

Figure 19. Co-citation network (cited sources)
Рисунок 19. Сеть совместных цитирований (источники)

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

112

3.2.3. Bibliographic coupling

We also consider bibliographic coupling, which is the opposite of co-citation, as there is an overlap in the reference lists of publications. In bibliographic coupling visualization, clusters are formed based on the referencing similarity of citing documents (articles). The larger the number of references journals have in common is, the stronger the bibliographic coupling relation between them is (Figure 20). Figure 20 shows a very different clustering compared to the co-citation network (Figure 19): one of the two large clusters (**red**) consists of the top-3 most cited and relevant journals of our collection devoted to the study of writing, as well as *Journal of Pragmatics*, – on one side, and proceedings from computer science subject area – *LNCS* and *Journal of Physics: Conference series*, which points to the existence of a multidisciplinary direction of studies in the analyzed domain combining methods from data science and writing research.

Another large cluster – the **green** one – contains journals devoted to different aspects of language learning, especially computer assisted language learning. The **blue** cluster is made up of the journals which consider psychological and physiological aspects of language production.

Figure 20. Bibliographic coupling of the journals
Рисунок 20. Сеть совместной встречаемости ссылок в пристатейных списках
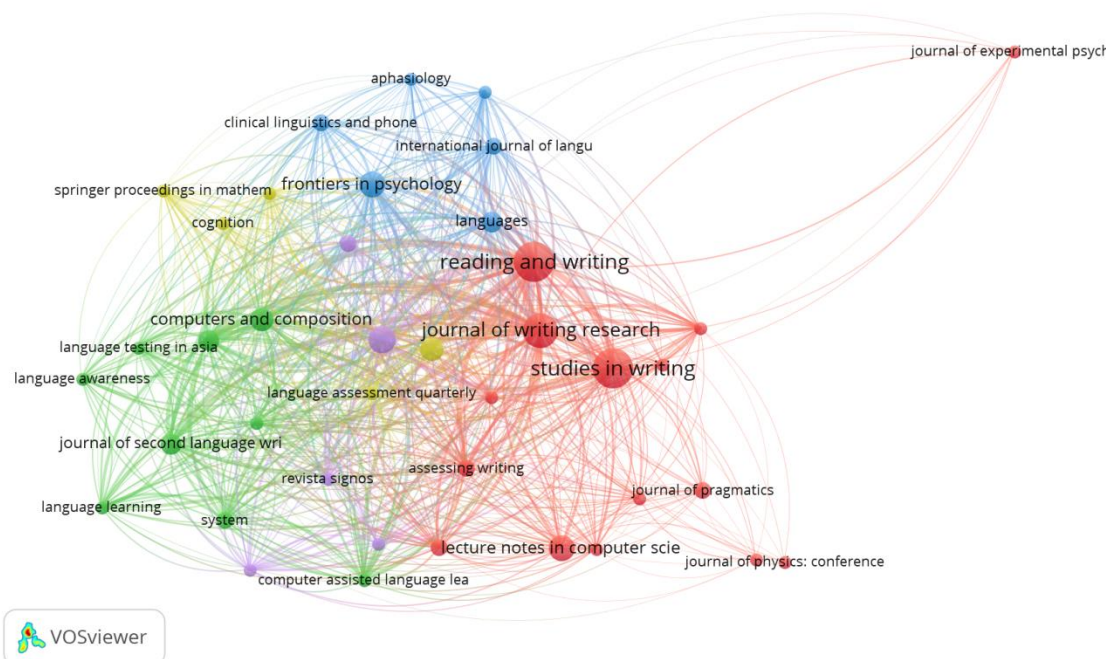
Conclusions

This study utilizing a combination of the bibliometrix R-package and VOSviewer fills the research gap by providing a holistic view of the research area related to the study of writing using keystroke data which can help scholars understand its current conceptual structure and development trends. The analysis highlights the feasibility of using R and the existing libraries for bibliometrics.

As our review shows, there has been an upward trend in the number of papers using keystroke loggers to study the writing process over the past five years. However, only few countries (the USA, Belgium, the Netherlands, Sweden) have the most published and cited papers, with Belgium

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

113

being the leader in terms of the cited papers and the most influential authors. The level of country collaboration is quite low. Thus, only few languages are involved in this important field of study, which obviously hinters further development and refinement of the models of writing. More international collaborations are needed, and more research groups all over the world should be involved in this field.

We believe that the knowledge mapping we present in this paper using a set of approaches will attract researchers who are new to this field to find directions for future research.

Nonetheless, this review acknowledges certain limitations. While this study is an attempt to fill a gap in bibliometric studies of writing research based on keystroke data, it considers only one citation database, which may not capture all the studies devoted to the use of keystroke methodology to understand the writing process. In the future, it would be beneficial to add data from some other databases (i.e., PubMed, Dimensions). Only English-language sources were analysed. Furthermore, the content analysis in this study does not take into account the use of synonyms. With future revisions, we want to address the aforementioned restrictions.

While a bibliometric review allows to analyze hundreds of papers and obtain an objective holistic view of the research field, it is unable to consider specific problems of the field. Therefore, our future plans include performing systematic scoping reviews (an excellent example of systematic scoping reviews is presented in (Tikhonova et al., 2023) to address some specific questions such as the linguistic nature of writing bursts, i.e., minimal units of text production, and ways of their extraction based on raw keystroke log data.

References

Alfalahi, H., Khandoker, A. H., Chowdhury, N., Iakovakis, D., Dias, S. B., Chaudhuri, K. R. and Hadjileontiadis, L. J. (2022). Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis. *Sci Rep* 12, 7690. https://doi.org/10.1038/s41598-022-11865-7

Aria, M. and Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Bixler, R. and D'Mello, S. (2013). Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. In Proceedings of the 2013 International conference on Intelligent user interfaces (IUI '13). Association for Computing Machinery, New York, NY, USA, 225–234. https://doi.org/10.1145/2449396.2449426

Büyükkıdık, S. (2022). A bibliometric analysis: A tutorial for the bibliometrix package in R using IRT literature. *Journal of Measurement and Evaluation in Education and Psychology* 13(3), 164–193. https://doi.org/10.21031/epod.1069307

Cui, Y., Mou, J. and Liu, Y. (2018). Knowledge mapping of social commerce research: A visual analysis using CiteSpace. *Electronic Commerce Research*, 18(4), 837–868.

Derviş, H. (2019). Bibliometric analysis using bibliometrix an R package. *Journal of Scientometric Research* 8(3), 156–160. https://doi.org/10.5530/jscires.8.3.32

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research* 133, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Farrukh, M., Raza, A., Mansoor, A., Khan, M. S. and Lee, J. W. C. (2023). Trends and patterns in pro-environmental behaviour research: a bibliometric review and research agenda. *Benchmarking: an International Journal* 30(3), 681–696. https://doi.org/10.1108/BIJ-10-2020-0521

Lee, S. A. and Park, J. H. (2024). Do Individuals with Mild Cognitive Impairment and Healthy Aging People Have Different Keystroke Dynamics? A Systematic Review. *International Journal of Gerontology* 18(2), 64–69. https://doi.org/10.6890/IJGE.202404_18(2).0001

Leijten, M. and Van Waes, L. (2013). Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. *Written Communication* 30(3), 358–392. https://doi.org/10.1177/0741088313491692

Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма...
Dekhnich O. V., Litvinova T. A. A keystroke analysis to study writing: a bibliometric review using...

114

Leijten, M., Van Waes, L., Schriver, K. and Hayes, J. R. (2013). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research* 5(3), 285–337. https://doi.org/10.17239/jowr-2014.05.03.3

Logan, G. D. and Crump, M. J. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science* 330(6004), 683–686. https://doi.org/10.1126/science.1190483

Maalej, A. and Kallel, I. (2020). Does Keystroke Dynamics tell us about Emotions? A Systematic Literature Review and Dataset Construction. In 2020 16th International Conference on Intelligent Environments (IE), Madrid, Spain, 60–67. https://doi.org/10.1109/IE49459.2020.9155004

Nguyen, T. M., Leow, A. D. and Ajilore, O. (2023). A Review on Smartphone Keystroke Dynamics as a Digital Biomarker for Understanding Neurocognitive Functioning. *Brain Sci.* 13, 959. https://doi.org/10.3390/brainsci13060959

Omotehinwa, T. O. (2022). Examining the developments in scheduling algorithms research: a bibliometric approach. *Heliyon* 8(5), e09510.

Passas, I. (2024). Bibliometric Analysis: The Main Steps. *Encyclopedia* 4, 1014–1025. https://doi.org/10.3390/encyclopedia4020065

Quraishi, S. J. and Bedi, S. S. (2018). Keystroke Dynamics Biometrics, A tool for User Authentication – Review. In 2018 International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 248–254. https://doi.org/10.1109/SYSMART.2018.8746932

Rashid, M. F. A. (2023). How to Conduct a Bibliometric Analysis using R Packages: A Comprehensive Guidelines. *Journal of Tourism, Hospitality & Culinary Arts* 15(1), 24–39.

Raul, N., Shankarmani, R. and Joshi, P. (2020). A Comprehensive Review of Keystroke Dynamics-Based Authentication Mechanism. In: Khanna, A., Gupta, D., Bhattacharyya, S., Snasel, V., Platos, J., Hassanien, A. (eds). International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, vol. 1059. Springer, Singapore. https://doi.org/10.1007/978-981-15-0324-5_13

Rejeb, A., Rejeb, K., Appolloni, A., Kayikci, Y. and Iranmanesh, M. (2023). The landscape of public procurement research: a bibliometric analysis and topic modelling based on Scopus. *Journal of Public Procurement* 23(2), 145–178. https://doi.org/10.1108/JOPP-06-2022-0031

Saini, G. K., Lievens, F. and Srivastava, M. (2022). Employer and internal branding research: a bibliometric analysis of 25 years. *Journal of Product & Brand Management* 31(8), 1196–1221. https://doi.org/10.1108/JPBM-06-2021-3526

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *J Am Soc Inf Sci* 24, 265–269. https://doi.org/10.1002/asi.4630240406

Stevenson, M., Schoonen, R. and Glopper, K. de. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing* 15(3), 201–233. https://doi.org/10.1016/j.jslw.2006.06.002

Tikhonova E., Kosycheva M., Kasatkin P. (2023). Exploring Academic Culture: Unpacking its Definition and Structure (a Systematic Scoping Review). *Journal of Language and Education* 9(36), 151-168. https://doi.org/10.17323/jle.2023.18491

Van Eck, N. J. and Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau and D. Wolfram (Eds.), Measuring scholarly impact: Methods and practice (pp. 285–320). Springer. http://dx.doi.org/10.1007/978-3-319-10377-8_13

Wang, S. and Gu, Z. (2022). Mapping the Field of Value Chain: A Bibliometric and Visualization Analysis. *Sustainability* 14, 7063. https://doi.org/10.3390/su14127063

Wengelin, Å. (2006). Examining Pauses in Writing: Theory, Methods and Empirical Data. In Computer Keystroke Logging and Writing: Methods and Applications. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9780080460932_008

Wengelin, Å. and Johansson, V. (2023). Investigating Writing Processes with Keystroke Logging. In: Kruse, O. et al. Digital Writing Technologies in Higher Education. Springer, Cham. https://doi.org/10.1007/978-3-031-36033-6_25

Yang, L. and Qin, S.-F. (2021). A Review of Emotion Recognition Methods From Keystroke, Mouse, and Touchscreen Dynamics. *IEEE Access* 9, 162197–162213. https://doi.org/10.1109/ACCESS.2021.3132233

Научный результат. Вопросы теоретической и прикладной лингвистики. Т. 10, №3. 2024
Research result. Theoretical and Applied Linguistics, 10 (3). 2024

115

All authors have read and approved the final manuscript.

Все авторы прочитали и одобрили окончательный вариант рукописи.

Conflicts of interests: the authors have no conflicts of interest to declare.

Конфликты интересов: у авторов нет конфликтов интересов для декларации.

Olga V. Dekhnich, Candidate of Philological Sciences (Cand. in Philology), Deputy Director for Science and Research of the Institute of Cross-Cultural Communication and International Relations, Associate Professor of the Department of English Philology and Cross-Cultural Communication, Belgorod State National Research University, Belgorod, Russia.

Дехнич Ольга Витальевна, кандидат филологических наук, доцент, заместитель директора Института межкультурной коммуникации и международных отношений по научной деятельности, доцент кафедры английской филологии и межкультурной коммуникации, Белгородский государственный национальный исследовательский университет, Белгород, Россия.

Tatiana A. Litvinova, Doctor of Philological Sciences (D.Sc. in Philology), Head of the Textual Psycholinguistic Modelling Research Laboratory, Professor of the Department of Russian Language, Modern Russian and Foreign Literature, Voronezh State Pedagogical University, Voronezh, Russia.

Литвинова Татьяна Александровна, доктор филологических наук, заведующий научно-исследовательской лабораторией текстового психолингвистического моделирования, профессор кафедры русского языка, современной русской и зарубежной литературы, Воронежский государственный педагогический университет, Воронеж, Россия.