INDONESIAN JOURNAL OF MULTIDISCIPLINARY SCIENCE

P-ISSN: 2808-5957
E-ISSN: 2808-6724

Exploring Lexical Innovation in World Englishes: A Global Bibliometric Study

Mahvish Nawaz Mokal*
University Utara Malaysia, Malaysia
Email: mokalmahvish@gmail.com

Keywords:
Lexical variation, World Englishes, Bibliometric analysis, Corpus linguistics, dialectology, Linguistic identity, AI-driven learning

ABSTRACT

Lexical variation plays a pivotal role in linguistics by facilitating the codification of new lexical items, the emergence of New Englishes, and the global recognition of World Englishes. This study presents a comprehensive bibliometric analysis of Scopus-indexed research on lexical variation to identify key trends, thematic focuses, and emerging directions within the field. Using R-based bibliometric tools, the dataset of international publications was systematically analyzed to reveal influential authors, dominant themes, and leading journals. Dialectology emerged as the most extensively studied theme, Dirk Geeraerts as a prolific author, and the International Journal of Corpus Linguistics as a leading outlet. The United States leads in publication output and citation impact, while the Federal University of Bahia is among the most influential affiliated institutions. Co-citation analysis demonstrates the intellectual structure and interdisciplinary connections in the field. The findings highlight growing scholarly interest in lexical variation and underscore the role of sociocultural factors in shaping linguistic identity, innovation, and variation. Future research is encouraged to explore the social dimensions of lexical diversity and examine how technological platforms, including AI- and IoT-driven learning environments, influence language use, emphasizing the need for linguistically inclusive systems.

This is an open access article under the CC BY-SA license.

INTRODUCTION

The concept of lexical variation in world Englishes has gained increasing prominence as global societies acknowledge its influence on linguistic and social dynamics. Within the framework of applied linguistics, the study of world Englishes highlights how sociocultural, educational, political, and historical factors shape diverse forms of English across regions (Melchers & Shaw, 2019; Schneider, 2023). This diversity is reflected in variations across phonetics, lexemes, grammar, semantics, and pragmatics, illustrating the adaptability of English within multicultural contexts. As multilingualism and global mobility expand, examining lexical variation becomes essential for understanding language evolution and identity construction (Gulnazir & Salehuddin, 2023). Furthermore, recent linguistic scholarship emphasizes the integration of quantitative and corpus-based methodologies to complement traditional qualitative approaches, allowing a more systematic analysis of language variation (Mukherjee, 2022). Studies in intercultural communication, linguistic competence, English standardization, and discourse in context underscore the importance of codifying new vocabulary and tracking lexical innovation as dynamic indicators of linguistic change (Jenkins, 2013; Bolton, 2021). This paradigm shift encourages scholars to move beyond descriptive analysis toward data-driven and rational dimensions of language variation, bridging sociolinguistic theory with empirical linguistic evidence (Blanco-Mesa et al., 2017).

Lexical variation is a fundamental structural phenomenon inherent in all spoken and written languages, with English serving as a prominent example (Lipka, 2010). The historical trajectory of English, from its origins as a native language to its status as a global lingua franca, has been marked by continuous evolution and diversification (Crystal, 2025).

The concept of World Englishes (WEs) encapsulates this phenomenon, highlighting the emergence of distinct English varieties shaped by local linguistic practices and cultural

identities. Schneider's (2025) Dynamic Model and Kachru's (1985) Three Circles Model offer frameworks for understanding these developments, emphasizing the fluidity and adaptability of English in different sociolinguistic contexts. These models challenge traditional notions of Standard English, advocating for a more inclusive perspective that recognizes the legitimacy of various English varieties.

Recent studies have further explored the implications of lexical variation in English. For instance, Sandow and Braber (2025) examine how lexical choices reflect social identities and group affiliations, illustrating the role of language in constructing and negotiating meaning within communities. Similarly, Botha (2025) investigates the connections between World Englishes and sociolinguistic variation, providing insights into how English adapts to and is shaped by local contexts.

Understanding lexical variation in English is crucial for educators, linguists, and policymakers, as it informs language teaching practices, language policy development, and the promotion of linguistic diversity. By acknowledging and embracing the multiplicity of English varieties, we can foster a more inclusive and equitable approach to language use and education in a globalized world.

Khan (2013) explored many examples of loan words from Urdu and Pakistan's native languages that have become mainstream use. These borrowed words from local languages and the English grammatical scheme used by print media. Similarly, Uzair, Mahmood & Raja (2012) used the words such as "goondas" and "jirgas," which are taken from local languages with the use of the suffix Englishes. It also explored the development of another kind of term. This is accomplished by using the prefix de- in phrases such as "de-notify," "de-load," and so on. The study also mentioned the use of suffixes -ee, -er, and-ism in the newspapers several times. For instance, "notorious history-sheeter," where the newspaper exploited an unusual way of using a purely English word. Equally, in Pakistan, English suffixes are combined with Urdu suffixes. i.e., "ziaism," "shariatism," etc., there are many hybridized words in Pakistani English newspapers. For example, “lathi-charged," "double-roti," "paan- Shop," and many more.

Uzair, Mahmood & Khan (2012) explored another feature of word-level Pakistani English is its use of outdated vocabulary. There are some examples of such words cited as "botheration," "conveyance," and "thrice," etc. In addition, some English words evolved to meet the needs of local communities. Some instances of this include the words “baton-charge”, “brickbat”, “charge sheet”, “eveninger”, “history-sheeter”, “time-barred”, and “wheel-jam strike.”

As a pioneer in this area, Trudgill and Watts (2002) researched diversity in the English language and experimented with various English dialects. He stressed that some issues shifted from the phonological, lexical, and historical levels to a syntactic level where all dialectologists took less care. It justified that the English language changes and provides different dialects when traveling from one place to another (Trudgill,1999) It is opined that the amount of non-native English speakers has been growing faster than the number of native English speakers over the years because they provide variation in their English borrowing dialects from the local community.

According to Yamanaka (2006), the Inner Circle consists of 320–380 million native English speakers, whereas the Outer Circle consists of 150–300 million, and the Expanding Circle consists of 100–1000 million. The same phenomenon has been observed by many other linguists as well. According to Kachru (1996), the native speaker cannot match the same lexemes as the non-native speakers. Besides, four non-native speakers must be there to communicate with each native speaker. These linguistic and cultural variations influence the growth and extension of linguistic culture in wide English varieties worldwide (Mokal &Halim, 2023).

Kachru and Smith (2008) explained the term "World Englishes" in detail. They advocated the true idea of pluricentric behind the term that has spread with formal variations all over the

globe. Ahmad (2018) also claimed that English symbolizes functional and formal linguistic plurality and systematic acculturation. Galloway (2017) observes that it will be important to see the future of English worldwide if these Modern English will retain its status in the world or if they are incorporated into the other native variants. The paradigm shift will proceed, and if it is anticipated that this transition will take place, consumers of New English should likewise adhere to the rights that are asserted by speakers of their mother tongue (Davies, 2019). In such a classroom community of conversation, students would feel safe interpreting background knowledge, questioning assumptions, engaging with other cultural patterns, and clarifying moral perspectives beyond ideological, political, and cultural borders (Cummins, 2021; García & Kleyn, 2016).

Furthermore, Pastor (2009) indicated that linguistic variability exists regardless of using English as a combination of various languages. Cultural diversity enriches universal language as it is used by foreign authors and often provides various ways of conveying one idea. He also argued that language variability is induced by mother tongue interference when non-native writers and speakers convey their language patterns while writing and communicating in the target language. Similarly, Newman (1996) defined that the subject will be transferred between agents while using the ditransitive verb. Future studies are expected to adopt mixed-methods approaches to explore learners' collocational knowledge from multiple perspectives, as such approaches align with broader calls in World Englishes research to recognize linguistic diversity and plural norms (Mokal & Ahmad, 2024)

Conversely, since editorial guidelines do not enable researchers to monitor language variation, non-native writers need help writing research papers. They update their writing-up many times so that their work will follow the appropriate requirements in the international format of American and British committees-led publications (Balouch & Panhar, 2019). Therefore, the local language is transferred from non-native to native because of the environment change and socio-cultural context, making new Englishes. Thus, these local non-native linguistic terms and cultural variations with the local impact influence the growth and extension of linguistic culture in wide English varieties worldwide.

This research aims to illuminate the understudied common ground, enhance understanding of lexical variations, and identify characteristics and linguistic acceptability to close these gaps. The authors conducted a bibliometric study to determine how the new lexical variations that have emerged over time in literary and scientific production have affected the disciplines' receptivity to contributions from other linguistic regions. The essay is organised as follows. The design of the review methodology and the data search procedure are emphasised in the following section, which describes the methodological components of the systematic literature review. The third section provides a summary of the significant findings from the descriptive and keyword analyses, as well as the primary themes of the selected publications. The fourth section is a summary of the major findings regarding scientific productivity. In Section 5, recommendations and gaps requiring additional research are presented.

This study offers several significant benefits. Theoretically, it enriches the understanding of lexical variation within the framework of World Englishes by providing a systematic bibliometric overview of the field's evolution, dominant themes, and intellectual structure. Methodologically, it establishes a clear map of research trends, influential authors, collaboration networks, and emerging directions, serving as a valuable reference for future scholarly inquiry. Practically, the findings offer actionable insights for language educators, linguistic policymakers, and developers of language technologies, aiding in the design of more inclusive approaches that acknowledge and incorporate global English diversity into teaching, policy-making, and digital language systems.

METHODOLOGY

The bibliometric analysis technique has been employed to carry out this study project. This methodology's primary goal is to discover, arrange, and analyse the critical elements within a particular study subject (Vallero & Parajes, 2021). The usefulness of this review approach rests in its capacity to synthesise the current evidence within a certain topic or field of knowledge, to boost the codification and analysis of literature results, and, as a result, to identify gaps that need to be filled by subsequent studies.

A total of 396 publications about lexical variations in world English have been examined utilizing mathematical, statistical, and mapping with R-tool to find any existing links. The primary databases that are directly related to the field of study have been consulted, and the research from Scopus has been chosen because it has many publications, authors, and journals that match the standards for scientific excellence set by peer review (Wess & Priemer, 2023). In addition, this repository has the most information that is currently available regarding the total number of authors, institutions, and nations that contributed to the information that is stored here. The search for these study papers was conducted in April 2023, whereas the statistical analysis concentrated on the most recent 20 years, or the period ranging from 2003 to 2022. The review relied on the information that is detailed in the accompanying table, which you may access by clicking here.

Table 1. ***Description of Data Collection***

Main Information About Data	
Timespan	2003:2023
Documents	396
Keywords Plus (ID)	927
Authors	838

Source: Author's compilation based on Scopus database analysis (2003–2023)

Figure 1. illustrates the logical process used to get the data needed for the study. In this manner, The R-tool was used to analysing the keywords included in this study, as well as the publication year, the journal, the subject area, the author and co-authors, the institution, and the nation. For each variable, this software creates network maps that enable the grouping and processing of words. Using the R-tool, we may also generate, and display co-occurrence networks of significant phrases taken from the scientific literature (Owolabi & Sajjad, 2023).

Finally, the number of articles that were published during each of the four decades that make up the most recent twenty years is used as a basis for analyzing the progression of scientific production. This analysis considers the productivity of authors, countries, and institutions by counting the number of papers presented in each field, as well as the number of citations, the H-index, and the SJR impact factor of the significant articles.

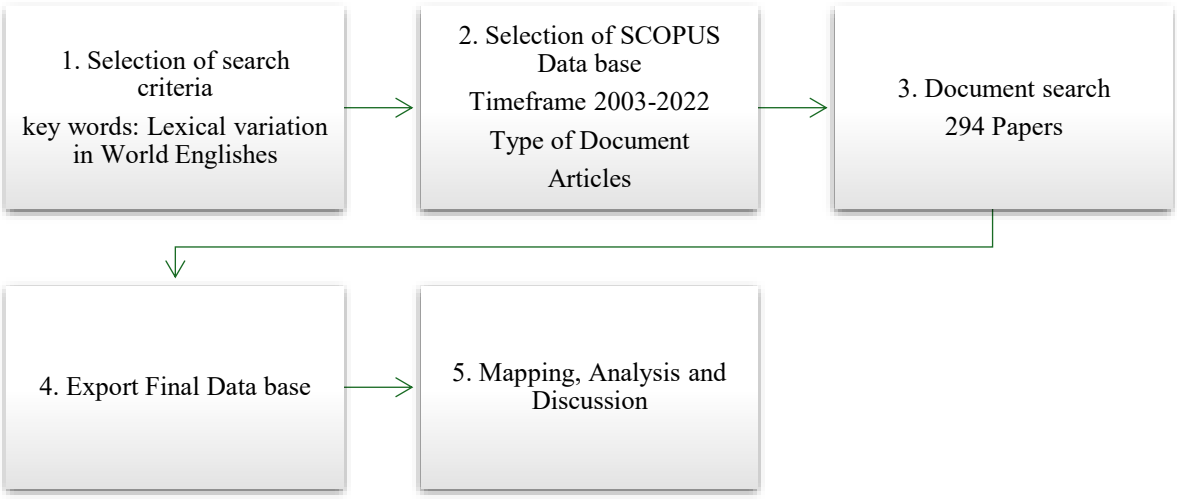


Figure 1. **Flow diagram of Methodology used.**
Source: Author's design for systematic bibliometric review

RESULT AND DISCUSSION

Evolution of Scientific Production

This section presents the key characteristics of scientific production, an estimate of the number of articles published, and the total number of journals where articles are published with the research topic under study.

The evolution of the critical traits of the articles published on Lexical Variation in World Englishes between 2013 and 2022 is thus shown in Figure 2. To make the analysis more manageable, the time range, which spans twenty years, has been broken into intervals of 5 years. Since the scientific output of each five years nearly triple that of the prior five-year period, the scientific production for the keywords studied during this time has increased exponentially.

Table 2. ***Characteristics of annual Scientific production from 2003 to 2022***

Year	Articles	Authors	Countries	Citations	TC/A	Journals
2003-2007	23	63	9	64	2.78	13
2008-2012	63	149	16	204	3.23	29
2013-2017	133	286	37	467	3.51	78
2018-2022	171	437	59	1139	6.66	132

Source: own elaboration. (TC/A): an average of citations per article.

This means that if only 23 research articles are published in the first five years of the analysis (2003-2007), 171 articles are published in the last five years (2018-2022), representing a 743.48% increase. The number of publications throughout the five years (2018-2022) is significant because it increased the growth of publications in the studied field of knowledge.

Even though this line of research emerged in 2020 when Lehtonen affirmed that the significance of words is fluid, which is often examined in words that can never alter their basis. He explained Sebastian's (2002) definition of "Semantics." He identified that the meanings of language bodies, including terms, sentences, grammar shapes, and phrases, are concerned only with semantics. Linguistic expressions are also a question of semantics. An international standard that sets, voluntarily guiguidesr all types of organisations in any beneficial sector to follow to conduct their business in a socially responsible manner in a society that is becoming

increasingly demanding. Figure 2 illustrates quite clearly this large growth in the number of citations as well as the authors. Lexical variation 2003-2022.

The number of authors who made contributions to the research topic is 935. Similarly, there are more articles now than five years ago; from 2017 to 2022, there are also more authors overall (6.97% of all authors). For their part, 221 countries have shown interest in the subject of the analysis. Like the preceding indicators, there has also been a tremendous increase in this one. From 09 between 2003-2007 to 16 between 2008-2012, 37 countries from 2013-2017, and 59 from 2018-2022, it has 6 times multiplied by twenty years.

Throughout the course of the period, there was a consistent increase in the number of citations, which went from 64 during the five years 2003–2007 to 1139 citations during the period 2018–2020, which resulted in a total of 1874 citations for the entire period that was examined. The study looked at a period that went from 2003 through 2020. Because of this, the median number of citations per article has increased to 17.79, which is indicative of a considerable boost in interest in research.

Ultimately, the number of journals has skyrocketed, moving from 13 in the first five years analysed to 132 in the period 2018–2022, which represents an increase of 10.15% over the previous figure. In this way, all the analysed variables reveal a rise in international interest and scientific production, which makes it evident that this path of research was a burgeoning one at the beginning of the 21st century.

Analysis of Scientific Production by Areas: Journals, Authors, and Countries

There were many publications over the period studied (2003-2022), but we chose the ten most relevant in which the 390 research articles analysed regarding lexical variance in world Englishes were framed. Figure 3 depicts the ten major periodicals where Scopus matches these articles. As a result, the Dialectologia journal got the most research papers over the period studied, with a total of 8, accounting for 22.86% of the total. They are followed in importance by Spanish in context (4, 11.42%), Asian EFL Journal (3, 8.57%), International Journal of Corpus Linguistics (3, 8.57%), Language related research (3, 8.57%), Language resources and evaluation (3, 8.57%), and Pluricentricity: Language variation and Sociocognitive dimensions (3, 8.57%), Proceedings of the 10th international conference on Language Resources and Evaluation, Energy (595, 8.78%), LRE (3, 8.57%), Sign Language Studies (3, 8.57%), and American Speech (3, 8.57%). Furthermore, all these thematic parts have had articles published over the analysed period (2003-2022) by addressing social and environmental issues and a generic instrument that promotes linguistic development, indicating the cross-disciplinary nature of the study path. Table 2 displays the significant characteristics of the research articles published in the ten highly beneficial scientific journals between 2003 and 2022.

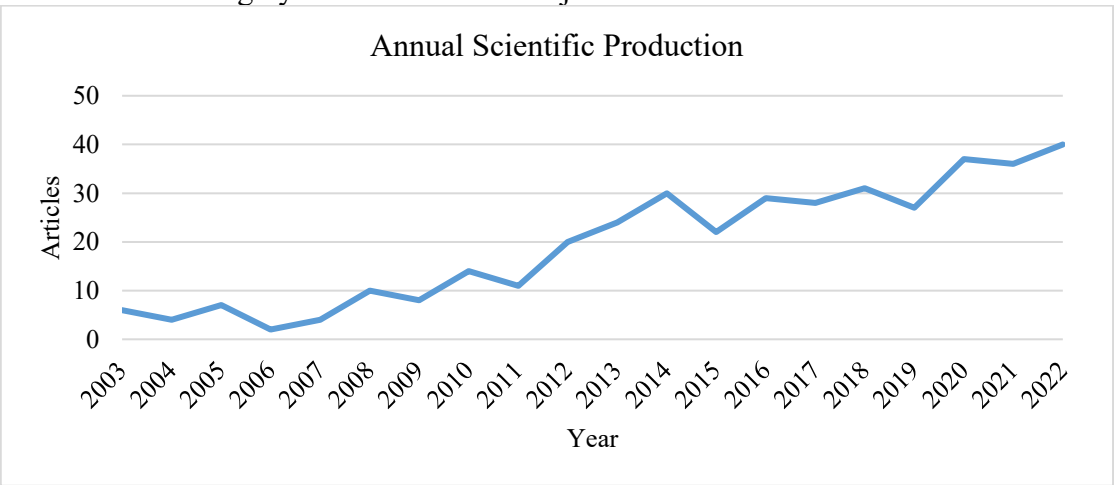


Figure 2. Annual Scientific Production
Source: Bibliometric analysis of Scopus data (2003–2022)

Distribution of Publications by Subject Area and Journal

If during the first five years under consideration (2003–2007), a total of 23 articles were published, then during the most recent five years (2018–2022), the number of publications will be 171, which indicates an increase of 743.48% over the previous period. Because it accounts for 62% of all articles about knowledge that have been analysed, the quantity of publications that have been produced in the past five years is particularly noteworthy.

Most Relevant Journals

It is determined that the highest impact index is the one presented by the International Journal of Corpus Linguistics (42), followed by Pluricentricity: Language Variation and Sociocognitive Dimensions. Regarding the number of research papers received, lexical Variation has been the most productive (16), and the Journal of Dialectologia (10). On the other hand, Gema Online Journal of Language Studies, Literary and Linguistic Computing, Sign Language Studies, Spanish In Context, 15th Conference of The European Chapter Of The Association For Computational Linguistics, each 2017 - Proceedings Of Conference are the 5 journals that have published 74 scientific articles during the two decades analyzed, while Language Resources And Evaluation is the only one that has received 55 research papers on this line of research in the last ten years. This is a resounding indication that the research issue has been capturing the interest of the academic community and, as a result, of scientific journals.

The journal that has received the greatest number of articles over the past five years is the Proceedings of the 10th International Conference on Language Resources and Evaluation. This journal also has the highest H-index in the study area, which is 3. However, it is important to note that 80% of the most prolific scientific journals are published in nations that are a part of the United States, while the remaining 20% are published in countries that are a part of the European Union. Furthermore, there are no journals published on any of the other continents that are included in the top 20 most productive journals list.

Table 3. Evolution of Most Relevant Journals

Sr.no	Sources	Articles
1	Diabetologia	8
2	Spanish in context	4
3	Asian EFL Journal	3
4	International Journal of Corpus Linguistics	3
5	Language related research	3
6	Language Resources and Evaluation	3
7	Plucentricity: Language variation and socio-cognitive dimensions	3
8	Proceeding of the 10th international conference on Language Resources and Evaluation, LREC 2016	3
9	Sign Language studies	3
10	American Speech	2

ource: Scopus-indexed publications on lexical variation in World Englishes (2003–2022)

Table 4. Ranking 10 journals with the highest scientific production.

Sr no.	Journals/J	h_index	g_index	m_index	TC	NP	PY_start
International Journal of Corpus Linguistics		3	3	0.333	42	3	2015

Pluricentric: Language Variation and Sociocognitive Dimensions	3	3	0.273	16	3	2013
Proceedings Of The 10th International Conference on Language Resources and Evaluation	3	3	0.375	17 5	3	2016
Dialectologia	2	3	0.182	10	8	2013
Gema Online Journal of Language Studies	2	2	0.25	9	2	2016
Language Resources and Evaluation	2	3	0.182	55	3	2013
Literary And Linguistic Computing	2	2	0.182	20	2	2013
Sign Language Studies	2	3	0.222	24	3	2015
Spanish In Context	2	2	0.2	6	4	2014
15th Conference of The European Chapter of The Association for Computational Linguistics, Each 2017 - Proceedings of Conference	1	1	0.143	15	1	2017

Source: Bibliometric analysis via R-tool (Scopus, 2003–2022)

(J): Journals; (H): H-index; (TC): Citation per year; (NP): number of Publication

Author Productivity

This section's goal is to highlight the authors who have produced the most work, as well as the authors' collaborative efforts, as measured by the number of works on which they are co-credited. As a result, Table 3 outlines the primary variables investigated by the ten scholars who have contributed the most to the field of lexical variation in World Englishes between the years 2003 and 2022. This information may be found in the table. It is noteworthy that all the authors who demonstrate higher output in the field of research are of European descent; more specifically, sixty percent of them are of Spanish descent, while the remaining forty percent come from a variety of various nations.

Table 5. **Most Productive Authors**

Most Relevant Authors		Author's Production Over Time			Average Citation per Article		
Authors	**A**	**Year**	**Tc**	**Tcpy**	**Country of Author**	**AC**	**TC**
Geeraerts	9	2004	51	2.55	USA	26.1	1625
Speelman	9	2005	37	1.947	Canada	17.8	235
Boberg	4	2015	4	0.444	United Kingdom	29.4	232
Da Silva	3	2016	1	0.125	Netherlands	17.7	206
De Boer	3	2010	28	2	Belgium	7.6	195
De Vos	3	2013	5	0.455	China	13.7	107
Eisenstein	3	2015	4	0.444	Mexico	6.3	41
Kamiran	3	2020	9	2.25	Italy	7.4	38
Karim	3	2021	1	0.333	Hong Kong	9.2	37
Kuiken	3	2022	1	0.5	Portugal	65	37

Source: Scopus data processed with bibliometric R-package (2003–2022)

(A): number of articles; (TC): number of citations; (Tc): an average of citations per article; (Tcpy): An average of citations per year; (TC): Total Article citation ; (AC): Average Citation per article

This section's goal is to highlight the authors who have produced the most work, as well as the authors' collaborative efforts, as measured by the number of works on which they are co-

credited. As a result, Table 4 illustrates the primary variables investigated by the ten scholars who have contributed the most to the field of lexical variations between the years 2003 and 2022. These researchers are listed in order of their level of productivity. Notably, all the authors who demonstrate higher productivity in the field of research are of European ancestry. Of these, fifty percent are of Spanish heritage, while the other fifty percent are from a variety of various countries.

Geeraerts D., who has a total of 9 publications published over the period analysed (2003–2022), is the most productive author. Speelman D., who also has a total of 9 research articles, is the next most productive author. Both authors have the same quality indication in the field of research, even though the first author has received a greater total number of citations (1625). For his part, Boberg C., who is based in the UK, is the author of the work that was published four times and received a maximum of 232 citations. Finally, out of the ten authors that have produced the most work, only Karim A. and Kuiken F. have only written three research articles in their respective study fields, and those works have received a total of 37 citations.

Figure 3 depicts, finally, the co-authorship analysis-based collaboration map among the primary authors who have written about lexical variation in world Englishes. The extent of the circle varies based on the number of articles each author has contributed, and the colours represent the working groups. Only Gonzales Saez and Filippo Goeuriot appear to have a consistent international network of work among the most prolific authors, indicating that the network's extensive dispersion may assist the study field in expanding rapidly. Regarding the remaining most productive authors, it appears that a global collaboration network on lexical variation in world Englishes is developing.

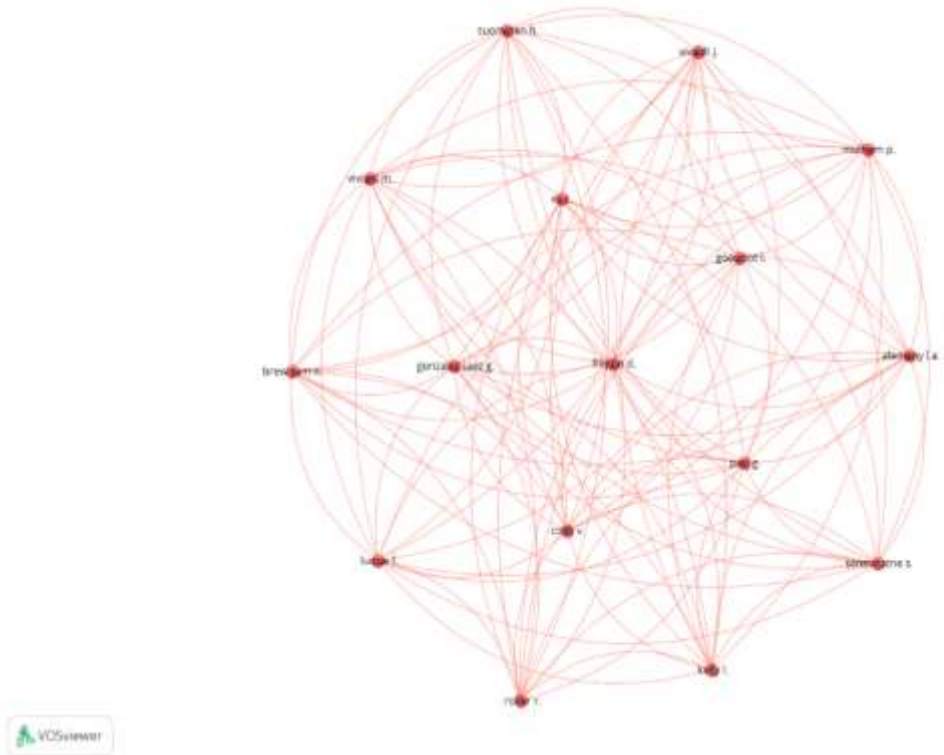


Figure 3 **The network of cooperation, based on the co-authorship of the main author.**
Source: VOSviewer visualization of co-authorship data from Scopus.

Scientific Productivity of Countries and Institutes

This section's objective is to highlight the most fruitful international collaborations, as well as the highest productive international institutes and nations, as well as the rates of

collaboration, and the cooperation network that is based on co-authorship. Figure 4 presents a breakdown of the scientific output of countries that have contributed to the study of lexical differences among World Englishes. The network displays a high degree of dispersion, which has the potential to encourage the rapid expansion of the study field given that it includes some of the most prolific authors.

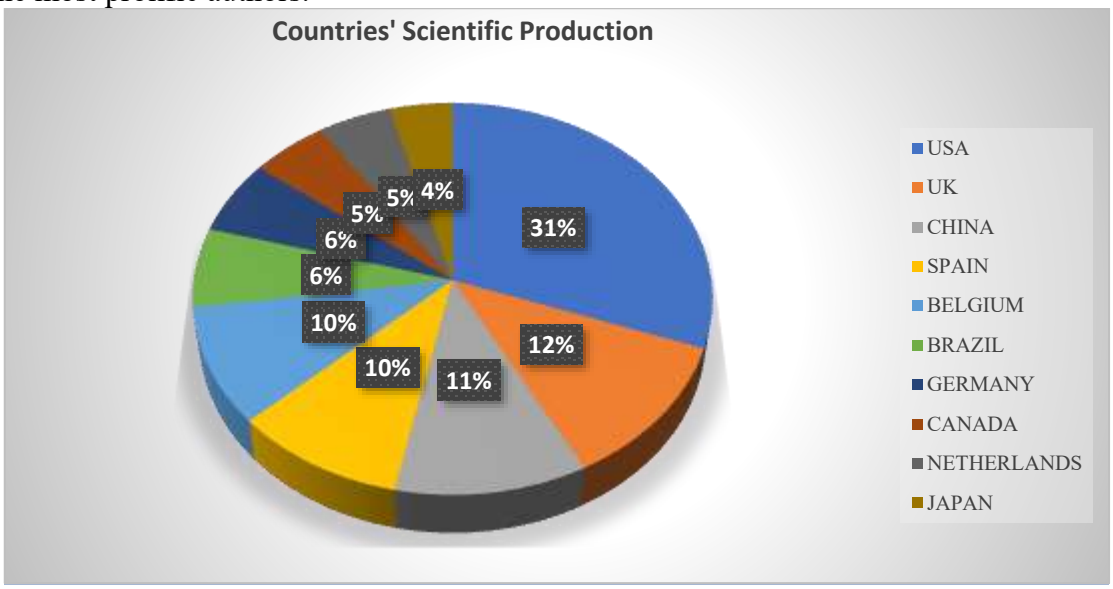


Figure 4 **Scientific Production of Countries**

Source: Geographic distribution of publications based on Scopus metadata (2003–2022).

The networks of international cooperation and key evolution indicators for the most productive nations within the scope of the study and during the time under consideration are presented in Table 6. The United States, with a collaboration rate of 31% and the most international collaborators, is the country with the highest level of international cooperation. The United Kingdom (12%) is the second-most cooperative country, after China (11%). Less than fifty percent of the remaining, most productive nations collaborate internationally. In global networks, the average number of citations is also higher in Spain, Belgium, Brazil, and Germany.

Asian nations have a strong propensity for forming international alliances and attaining high-quality indices in terms of their institutions. As a concluding illustration, Figure 5 depicts the collaboration map based on co-authorship between the main nations. International working groups are represented by colours, with the size of the circle determined by the number of objects. Consequently, the red cluster contains the most nations and is led by the United States (127). It is followed by the blue cluster, which consists of 54 countries, including the United Kingdom, Spain, and the Russian Federation, and the green cluster, which is led by Belgium, China, and Brazil. Australia is in the yellow cluster, which contains 17 countries, including Malaysia. Italy is in the cluster with the fewest nations, which is coloured purple.

Table 6 Scientific Production of Countries and Institutes

Sr	Scientific Productivity of Countries		Scientific Productivity of Institutions	
	Country	**Freq**	**Affiliation**	**Articles**
1	USA	127	Federal University of Bahia	12
2	UK	54	University of Leuven	10
3	Spain	46	Columbia University	7

4	Belgium	43	Mayo Clinic	7
5	China	38	Nanjing University of science and Technology	6
6	Brazil	30	Radbound University	6
7	Germany	28	University of Michigan	6
8	Netherlands	19	Vrije Universiteit Brussel	6
9	Canada	18	Faculty of Arts	5
10	France	18	Northeast Normal University	5
11	Japan	18	State University of New York	5
12	Australia	17	University of Southampton	5
13	Italy	16	Aguascalientes	4
14	Pakistan	16	Centre de Recherche information de montreal (CRIM) QC	4
15	Iran	15	Institute di scienze e technologie della cognizione (ISTC-CNR)	4
16	India	14	Lahore Garrison University	4
17	Mexico	13	Not reported	4
18	Portugal	13	Ritsumeinkan University	4
19	South Korea	11	Universitat de Barcelona	4
20	Indonesia	9	University of Cambridge	4

Source: Author's analysis based on Scopus affiliation data (2003–2022)

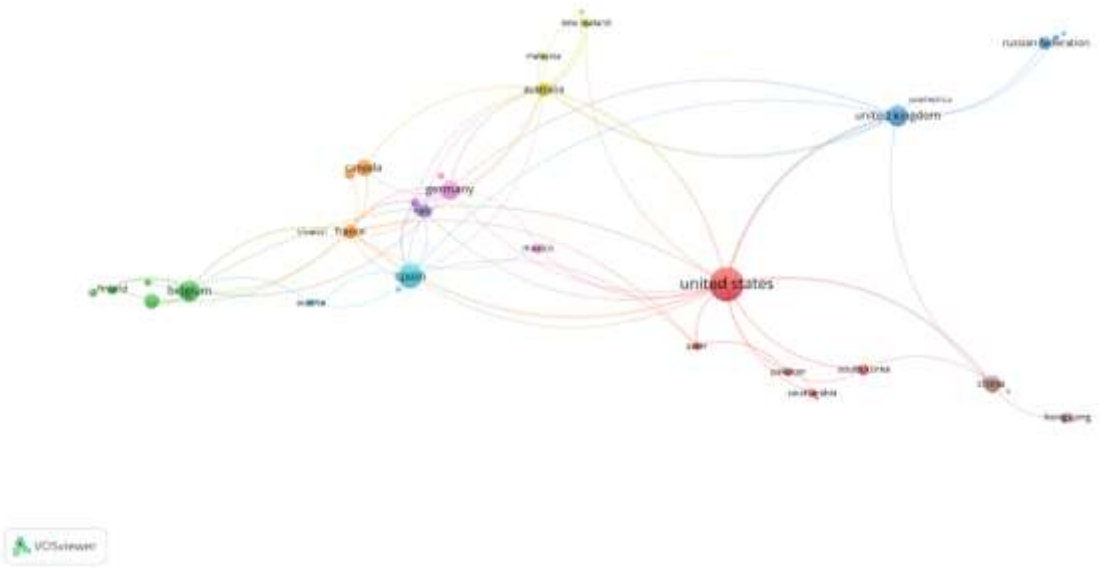


Figure 5 **Evolution of Collaboration Map of Countries**
Source: VOSviewer network analysis of international co-authorship patterns

This section's objective is to highlight the most fruitful international collaborations, as well as the highest productive international institutes and nations, as well as the rates of cooperation, and the cooperation network that is based on co-authorship. Figure 5 presents this information in such a way as to highlight the top 10 institutions in terms of the lexical variance that can be found in world Englishes between the years 2003 and 2022. The classification makes

it quite evident that the United States of America is home to 90 percent of the world's most productive institutions, while Asia accounts for the smallest percentage.

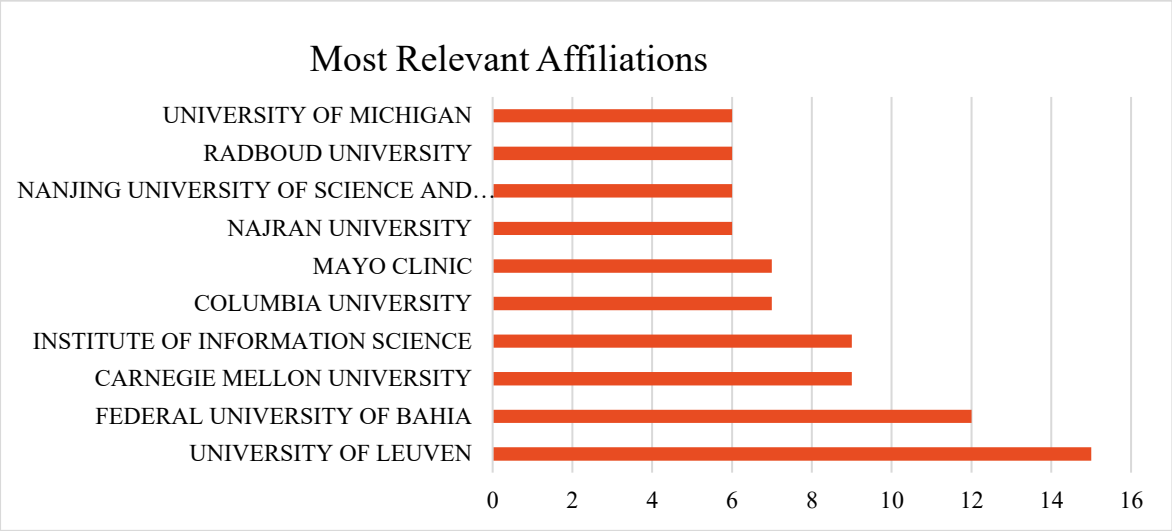


Figure 6 Most relevant Affiliation of Institutes
Source: Scopus institutional affiliation data (2003–2022)

The University of Bahia is the most productive institution, with the highest number of publications published throughout the period investigated; this is followed by the Federal University of Leuven, which has the maximum number of articles published; nevertheless, both universities have distinct H-index.

Keyword Analysis

The total of 396 publications on lexical variation in world English that were analysed between 2003 and 2022 yielded a total of 222 keywords. Table 7 displays the 19 variables that the authors consider to be the most significant because they permit us to categorise the interests generated in the study line based on the contextualization's cited characteristics over the course of the examined period. Notably, the frequency of the term "semantics" in research articles increased over time, reaching 21 out of the total number of research articles published between 2003 and 2022. Figure 6 depicts the relationship between each keyword in the entire sample.

Analysis of the simultaneous occurrence of keywords in the sample articles by the VOS viewer application led to the co-occurrence method, which forms the basis of this grouping. Consequently, the nodes' colours are used to distinguish the various clusters based on the number of co-occurrences, while the nodes' sizes vary based on the number of repetitions. Finally, 222 keywords for the selection procedure were discovered by identifying those with at least 19 interactions. Thus, there are six clusters of different colours, which may correspond to various research areas in the subject being investigated, while it is evident that this is a highly interdisciplinary line of inquiry based on the relationships, we identified between every keyword we identified:

The green cluster is comprised of a total of 222 items and refers to the significance of "lexical variation". Second, the blue cluster contains 21 items related to "semantics" from a comprehensive perspective. The red cluster, which alludes to the "natural language processing system", consists of 18 items in total. Twenty-three items are linked to "human" via the purple cluster. In contrast, the yellow cluster contains 15 items related to information retrieval, 8 of which are associated with "social media" The eleven-item light blue cluster has a direct relationship to the "sentiment analysis".

Table 7. ***Most Frequent Keywords***

Words	Occurrences
semantics	21
natural language processing systems	18
computational linguistics	16
human	15
linguistics	14
natural language processing	14
social networking (online)	14
data mining	13
female	12
male	12
article	10
language	19
Classification (of information)	9
human	8
Social media	8
adult	7
Information retrieval	7
Sentimental analysis	7
Text processing	7

Source: Keyword extraction from 396 Scopus publications using VOSviewer (2003–2022)

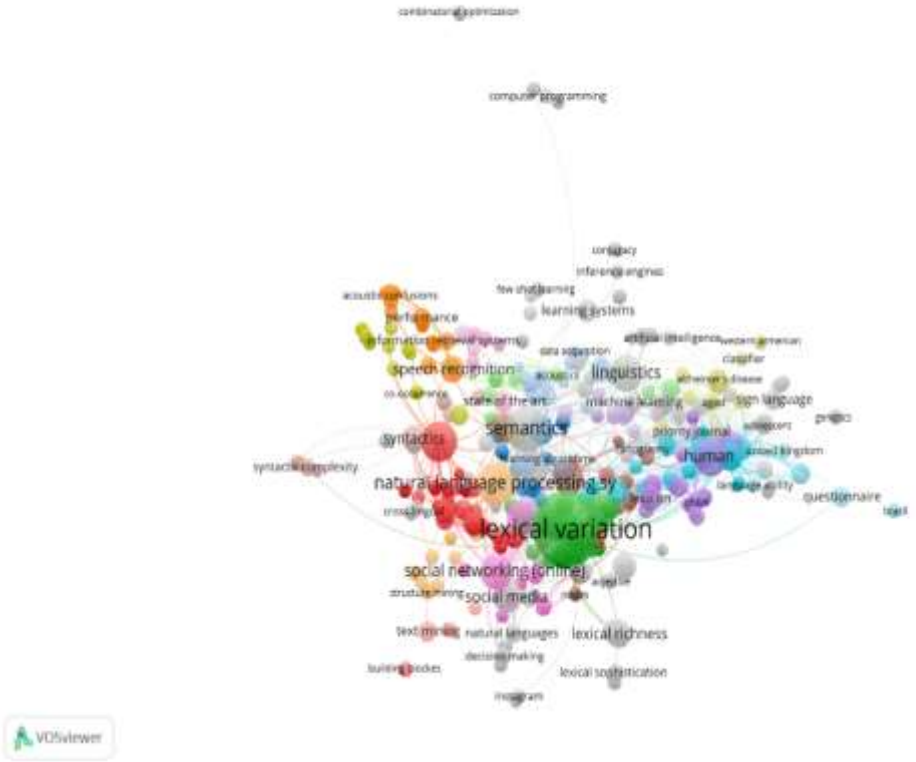


Figure 7 Evolution of keyword clusters based on co-occurrence.
Source: VOSviewer keyword co-occurrence analysis of 396 publications.

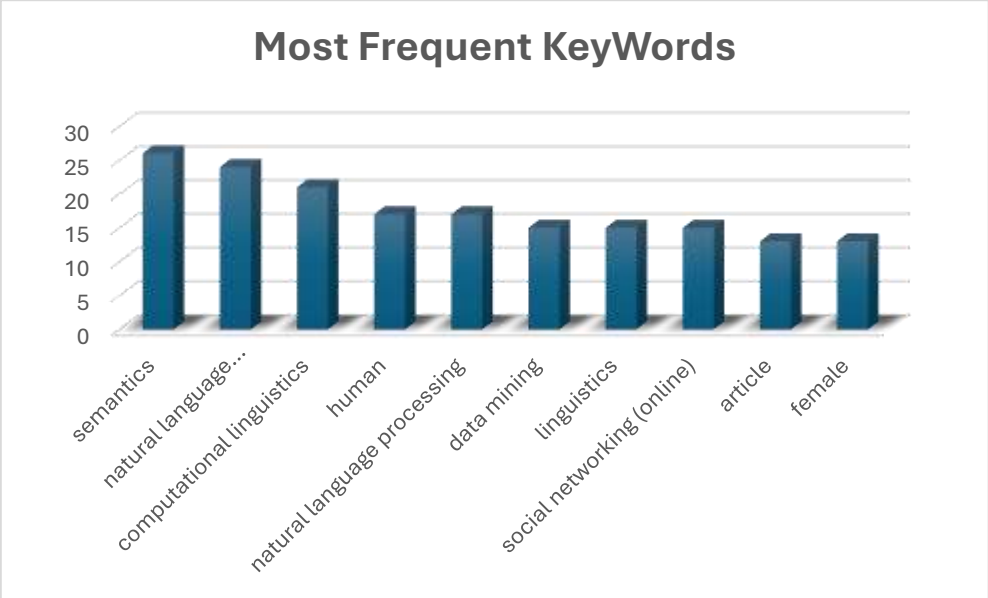


Figure 8 **Most frequent keywords**
Source: Frequency analysis of author keywords from Scopus dataset.



Figure 9 Word cloud of most frequent keywords
Source: Generated from keyword frequency data using R-tool

CONCLUSION

This study investigates the evolution of scientific research on lexical variation in World Englishes from 2003 to 2022, analyzing 396 Scopus-indexed documents. The findings reveal exponential growth in publications, authors, and citations, with a notable surge in scholarly journals, countries, and institutions over the past five years. Dirk Geeraerts emerged as the most prolific author, and American researchers dominated the top ten contributors. Thematic analysis identified "semantics" (24.76%) as the most frequent topic, followed by "natural language processing" (8.10%) and "computational linguistics" (7.20%). The *International Journal of Corpus Linguistics* holds the highest H-index, while the *Proceedings of the 10th International Conference on Language Resources and Evaluation* leads in average citations per article (175). Institutionally, the Federal University of Bahia produced the most research, and despite

increasing national diversity, European and American institutions remain dominant, with collaboration rates generally below 50%. Key research clusters span semantics, human linguistics, and natural language processing, reflecting interdisciplinary engagement. The growing scholarly interest underscores the role of social and technological factors—including AI and IoT-driven learning environments—in shaping lexical variation and linguistic identity, highlighting the need for inclusive language systems and suggesting future directions in linguistic structure analysis and paradigm development.

Future comprehension and evaluation of the lexical diversity of linguists who mobilize, and exchange information and expertise will be facilitated by a bibliometric study that provides specific indicators and generates links between clusters and significant lines. Future research could be informed by the limitations of the current investigation. Because this analysis is quantitative and does not consider any qualitative elements, it is subject to these constraints. These limitations, when paired with the use of alternative computer programs for data analysis, could result in slightly different conclusions. Similarly, the analysis did not consider anything besides the publication of scientific articles between the years 2003 and 2022. These include participation in congresses, book publications, and book chapters. Additionally, the analysis did not consider any other time frames. You may get different results if you use this approach to incorporate a wider variety of research sources.

REFERENCES

Bolton, K. (2021). World Englishes and linguistic innovation: Global spread and local adaptation. World Englishes, 40(2), 175–190. https://doi.org/10.1111/weng.12531

Crystal, D. (2025). English as a Global Language (2nd ed.). Cambridge University Press.

Gulnazir, H., & Salehuddin, K. (2023). Lexical variation and sociocultural influence in Malaysian English: A corpus-based approach. Asian Englishes, 25(3), 284–299. https://doi.org/10.1080/13488678.2023.2184951-

Jenkins, J. (2013). English as a lingua franca in the international university: The politics of academic English language policy. Routledge.

Melchers, G., & Shaw, P. (2019). World Englishes: Second edition. Routledge.

Mukherjee, J. (2022). Corpus linguistics and world Englishes: Methodological perspectives and empirical findings. English World-Wide, 43(1), 1–22. https://doi.org/10.1075/eww.00001.muk

Schneider, E. W. (2023). Dynamic models of World Englishes: Reassessing the evolution of postcolonial varieties. English Today, 39(1), 1–9. https://doi.org/10.1017/S0266078422000023

Schneider, E. W. (2025). Postcolonial English: Varieties Around the World. Cambridge University Press.

Ahmad, Z. (2018). A study of cohesion as a text-forming resource in the academic writing of Saudi undergraduate students of English as a foreign language (EFL) (Doctoral dissertation, De Montfort University). http://hdl.handle.net/2086/17481

Ataboyev, I. M., & Turgunova, F. R. (2022). The concept of semantic field in linguistics. ACADEMICIA: An International Multidisciplinary Research Journal, 12(3), 319–324.

Balouch, A., & Panhar, F. (2019). Ditransitive verbs: An English print media–based comparison of Pakistani English and British English. Asian Journal of Language, Literature, and Culture Studies, 1–9.

Blanco-Mesa, F., Merigó, J. M., & Gil-Lafuente, A. M. (2017). Fuzzy decision making: A bibliometric-based review. Journal of Intelligent & Fuzzy Systems, 32(3), 2033–2050.

Bolton, K. (2021). World Englishes and linguistic innovation: Global spread and local adaptation. World Englishes, 40(2), 175–190. https://doi.org/10.1111/weng.12531

Crystal, D. (2025). English as a global language (2nd ed.). Cambridge University Press.

Cummins, J. (2021). Rethinking the education of multilingual learners: A critical analysis of theoretical concepts. Multilingual Matters.

Davies, A. R. (2019). Urban food sharing: Rules, tools, and networks. Policy Press.

Durán-Sánchez, A., Álvarez-García, J., del Río-Rama, M. C., & Cruz, M. (2014). Active tourism research: A literature review. ROTUR: Revista de Ocio y Turismo, 8, 62–76.

Elendu, I. E. (2017). Lexico-semantic and syntactic variations in outer-circle Englishes: A study of Nigerian English. International Journal of English Language and Communication Studies, 3(1), 33–43.

Galloway, N. (2017). Global Englishes and change in English language teaching. Routledge.

]Gallego-Valero, L., Moral-Parajes, E., & Román-Sánchez, I. M. (2021). Wastewater treatment costs: A research overview through bibliometric analysis. Sustainability, 13(9), 5066.

Gokhale, A., Mulay, P., Pramod, D., & Kulkarni, R. (2020). A bibliometric analysis of digital image forensics. Science & Technology Libraries, 39(1), 1–18.

Gulnazir, H., & Salehuddin, K. (2023). Lexical variation and sociocultural influence in Malaysian English: A corpus-based approach. Asian Englishes, 25(3), 284–299. https://doi.org/10.1080/13488678.2023.2184951

Hijazi, S., Mokal, M. N., Halder, A., Victor, V., & John, G. (2025). AI- and IoT-Driven Mobile Learning Environments: A Framework for Educational Transformation, Workforce Development, and Digital Inclusion. International Journal of Interactive Mobile Technologies (iJIM), 19(24), pp. 4–17. https://doi.org/10.3991/ijim.v19i24.58959

Huang, Y., Huang, Q., Ali, S., Zhai, X., Bi, X., & Liu, R. (2016). Rehabilitation using virtual reality technology: A bibliometric analysis (1996–2015). Scientometrics, 109(3), 1547–1559.

Jayasuriya, U. (2018). Bilingual craftsmanship: The fusion of Sinhala interjections aney and aiyo in Sri Lankan English speech. In Understanding culture through language and literature (p. 81).

Jenkins, J. (2013). English as a lingua franca in the international university. Routledge.

Khan, A. (2012). The evolution of Pakistani English as a legitimate variety of English. International Journal of Applied Linguistics & English Literature, 1(5), 90–94.

Khan, H. I. (2013). An investigation of postgraduate students' and teachers' perceptions of English-medium instruction in Pakistani universities (Unpublished master's thesis). University of Manchester.

Lehtonen, T. (2020). Serres and foundations. Theory, Culture & Society, 37(3), 3–22.

Lipka, L. (2010). An outline of English lexicology (Vol. 3). De Gruyter.

Milian, E. Z., Spinola, M. D. M., & Carvalho, M. M. (2019). Fintechs: A literature review and research agenda. Electronic Commerce Research and Applications, 34, 100833.

Mokal, M. N., & Abd Halim, H. (2023). An analysis of lexico-semantic variations in Pakistani English newspaper corpus. World Journal of English Language, 13(6).

Mokal, M. N., & Ahmad, Z. (2024). The paradigm of World Englishes toward linguistic diversity. Social Sciences, Education & Languages, 6, 34.