

Submission date: 14/01/2023 Acceptance date: 01/02/2023 Publication date: 15/06/2023

CORPUS-BASED PRONUNCIATION RESEARCH: A BIBLIOMETRIC ANALYSIS

^{1,2,*} Bojie Qian & ¹*Farhana Diana Deris

¹Language Academy, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, Johor, Malaysia

²School of Arts and Humanities, Wuhan University of Communication, Wuhan, China

*Corresponding authors. E-mail: cynthiaqbj@163.com; diana@utm.my

ABSTRACT

Since the 2000s, there have been more and more empirical studies on pronunciation research, which result in review papers on pronunciation instruction and computer-aided pronunciation training (CAPT). The topic of pronunciation research, particularly corpus-based pronunciation research, is frequently neglected by academics. This study uses the bibliometric method to provide an overview of corpus-based pronunciation research between 1981 and 2022 and identify a major field trend. Based on the dataset of 271 publications retrieved from the Web of Science and Scopus, this study conducts a bibliometric analysis following the bibliometric research method suggested by Donthu et al. The findings reveal that Martine Adda-Decker is the most prolific author and that the United States is the most productive nation. Speech recognition and perception are the subjects that continue to exist from the past to the present, considering the keyword co-occurrence. Cocitation analysis indicates that "Journal of Phonetics" and "Journal of the Acoustical Society of America" have the most citations. Finally, future research directions and challenges are discussed, which means pronunciation teaching and teacher training will be researchers' concerns in the following years. This study provides some insight into a corpus-based pronunciation research field that is not reviewed by academia.

Keywords: pronunciation research, corpus, bibliometric analysis, collaboration network, keyword cooccurrence

Cite as: Bojie Qian, & Deris, F. D. (2023). Corpus-Based Pronunciation Research: A Bibliometric Analysis. *Law, Policy, and Social Science,* 2(1), 78–89. https://doi.org/10.55265/lpssjournal.v2i1.33

© The Author(s) (2023). Published by Intelligentia Resources. This is an Open Access article distributed under the term of the Creative Commons Attribution Non-Commercial License (<u>http://creativecommons.org/licenses/by-nc/4.0/</u>), whicl permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact <u>intelligentia.resources@gmail.com</u>.



Introduction

As a skill that contributes significantly to learners' second language (L2) competence, pronunciation plays a crucial role. Being able to communicate with others effectively depends on having good pronunciation (Derwing & Munro, 2015; Wahid & Sulong, 2013). While pronunciation is an essential skill for L2 learners, research and academia have not paid much attention to it twenty years ago. It should be noted, however, that this has changed a lot over the past decade. According to Derwing (2019), research studies on L2 pronunciation have exhibited "monumental growth" (p. 27). A few review articles primarily focused on the field of pronunciation instruction, such as the effectiveness of L2 pronunciation instruction (Gilakjani, 2016; Thomson & Derwing, 2015) before 2018. With the rapid advancement of technology in the 21st century, there have been numerous studies that have examined the prominent role of computers in pronunciation training, leading to review papers in this field (e.g., Agarwal & Chakraborty, 2019; Bliss, Abel, & Gick, 2018). Based on these studies, it can be concluded that although computer-aided or computer-assisted pronunciation training systems may appear to be a small niche within the context of L2 research, they prove to be very effective in both teaching and providing visual feedback on articulation. While the number of corpus-related pronunciation research articles has increased since 2006, no review articles have been published in this area until now.

This research provides a summary of the current state of corpus-based pronunciation research throughout the world, including the number of scholarly works that have been published, the most significant researchers and nations, and the degree of academic cooperation between researchers and countries. In order to study a field's intellectual structure, uncover hot research areas, and trace historical trends, scientific mapping—which includes word grouping and cooccurrence analysis—was also conducted. The conclusions of this paper might be useful to the academic community in identifying knowledge gaps and recommending future research directions.

Methodology

Scholars can conduct many different types of reviews to analyse the literature in different fields. Researchers seeking to investigate broad and rich areas of business research can benefit from bibliometric analysis (Donthu, Kumar, Mukherjee, Pandey, & Lim, 2021). There is no doubt that bibliometric analysis can also be used as a method of reviewing research in other areas of social science. This research has been conducted following a bibliometric analysis procedure suggested by Donthu et al. (2021) since it is an applicable method which can be used to analyse the research development in the social science field.



Data Source

This work will utilise data from two of the most esteemed and influential databases in the world, Web of Science (WoS) and Scopus, to examine and synthesise the results of this study.

Research Design

In accordance with Donthu et al. (2021), this research will follow the steps illustrated in Figure **1**. To begin with, the aims and scope of the paper are defined in the introduction part, which is corpus-based pronunciation research. A number of factors must be considered when compiling bibliometric data, including database selection, search query, and refinement criteria. The bibliometric analysis has been carried out using some techniques. For instance, publication-related metrics, co-authorship analysis and clustering, etc. The next step is to visualise the results using the VOSviewer tool. Lastly, the results of the study have been analysed and interpreted in order to reach a conclusion.



Figure 1. Research procedure



Compilation of the Bibliometric Data

Research data were retrieved from two authoritative databases commonly used by researchers: WoS and Scopus. In the top-level subject fields: life sciences, social sciences, physical sciences, and health sciences, Scopus covers nearly 36,377 titles from approximately 11,678 publishers. WoS is one of the best-known and most commonly utilised library resources in the world, with more than 7,000 subscribers, more than twice as much as the nearest rival Scopus (3000+) (Zhang, 2020). Due to this, both of them are considered to be the most reliable sources of bibliographic information for the majority of studies and research assessments. The data has been compiled in the way illustrated in Figure 2. For searching for the data, the keywords used were "corpus", "corpora", and "pronunciation" connected by a Boolean operator such as "AND" or "OR". Based on the keywords searched in the topic field in WoS Core Collection and in the title-abstract-key field in Scopus, there were 767 records in WoS and 1201 in Scopus respectively. By limiting the data from the past to 2022, selecting the document type as "article", categorizing the documents as "Linguistics", "Language Linguistics", "Education Educational Research," "Psychology Education" and "Education Special," on WoS, and selecting the Scopus subject areas as "Social Sciences" and "Arts and Humanities", there were 123 articles identified on the WoS platform and 228 on the Scopus platform. This search has been done on 19th February, 2023. Furthermore, the data were merged and cleaned by using the R-tool (Aria & Cuccurullo, 2017) and a total of 283 articles were found. Having deduplicated the data manually, there remained only 271 records.



Figure 2. Data compilation flow diagram



Research Tools

For conducting bibliometric analysis, there are several prominent research tools available, such as CRExplorer, Publish or Perish, CiteSpace, and VOSviewer (Moral-Muñoz, Herrera-Viedma, Santisteban-Espejo, & Cobo, 2020). VOSviewer is a software tool that facilitates the construction and visualisation of bibliometric networks based on co-citations, or co-authorship relations between journals, researchers, and individual publications as actors in the system (van Eck & Waltman, 2010). As a result, VOSviewer will be used when conducting the bibliometric analysis of the articles.

Results and Discussions

Main Information

DESCRIPTION		RESULTS
Main information about data	Timespan	1981-2022
	Sources	170
	Documents	271
Author	Authors	535
	Authors of single-authored docs	98

Table 1. Main information regarding corpus-based pronunciation research

The key characteristics of the data downloaded from the two databases are summarised in . This topic was first brought to public attention in 1981 when an article was published on the subject. A total of 271 articles have been published from 170 different sources, written by 535 authors, including 98 single authors.

Most Prolific Authors

As shown in Figure 3, the study conducted for this research identified Martine Adda-Decker as the author with the greatest number of publications, with contributions to nine publications, followed by Mirjam Ernestus who made seven publications. There are two scholars who have published four articles and five researchers who have published three articles in this field.





Figure 3. Author's publications

Most Prolific Countries

In terms of the number of articles published by each country, the United States has produced 36 papers, while France has produced 25 papers, as shown in Figure 4. However, despite being the last country on the list, Italy continues to have 8 publications in this field. Another interesting finding is the fact that most of the prolific countries are European countries such as France, Spain, and the United Kingdom. The rapid development of corpora in Europe may explain this phenomenon.

Author Collaboration Network

It is evident from Figure 5 that the most prolific author, Martine Adda-Decker, also maintains a complex network of relationships with other authors. In contrast, Mirjam Ernestus, the second prolific author, cannot be found in the network. This indicates that she worked on her own or rarely teamed up with the scholars in the network.





Figure 4. Countries' Contributions



Figure 5. Author collaboration network



Country Collaboration Network

An overview of the countries affiliated with corpus-based pronunciation researchers can be found in Figure 6. According to this figure, contributors affiliated with the United States and the United Kingdom play a central role in the network, while Canada and a number of European countries, including Spain, France, Germany and The Netherlands, are also highly influential. Most Asian nations, with the exception of China and Japan, can be found on the peripheral edge of the cluster. Nigeria is the only African country on the network's borders that appears on the map. Consequently, the majority of the authors on the network are from the United States and the United Kingdom, though authors from other European nations also play a vital role in the network.



Figure 6. Collaboration work of countries

Keyword Co-Occurrence

There are 30 clusters of keywords that appear in Figure 7, with some of the clusters being particularly salient, such as "speech recognition", "speech analysis", "pronunciation", "perception", etc. A few of them include the terms "Speech Recognition", which has a close relationship with "Speech Analysis", as well as terms such as "data reduction", "pronunciation variation", "speech corpora", and "error detection". In addition, there is a correlation between "pronunciation" and "perception" in terms of "intelligibility", "comprehensibility", "lexical frequency", and "predictability". A number of the dots with lighter colours represent the more recently adopted keywords, such as "pronunciation teaching", "teacher training", "oral corpus", "attitude", "Arabic learner", "experimental", "language acquisition", and "classroom-based". It is



found that scholars began to use corpora to facilitate pronunciation teaching and help teachers to implement the tool into their classroom teaching.



Figure 7. Keyword co-occurrence (Overlay visualization)

Co-Citation Analysis

Co-citation analysis is shown in Figure 8, where the source of the citation is taken as the unit of analysis. There are coloured circles and lines that represent the different sources and how they are linked with one another. It is also worth noting that the size of the circle represents the weight of the citations. Network visualization in Figure 8, both "Journal of Phonetics" and "Journal of the Acoustical Society of America" in blue, have the most citations to their credit and form a blue cluster with co-cited sources like "Phonetica" and "Journal of Memory and Language". Meanwhile, most of the journals in the blue cluster are phonetic journals or cognitive journals, which differ from the traditional linguistic journals that are gathered in the green cluster. In green, "Language Learning" links with "Studies in Second Language Acquisition" and "System".





Figure 8. Co-citation analysis

Further Development Trends and Challenges

An overlay visualisation was applied in order to have a comprehensive understanding of the corpus-based pronunciation research, in order to analyse the trends in the research. Clusters with darker colors indicate that a research topic has been studied earlier. In the meantime, the lighter the cluster, the more recent the research topic. There has been a shift in the focus from the analysis of the corpus materials to the use of those materials as a tool to facilitate teaching and training, as shown in Figure 7 with the keywords switching from "speech analysis" or "speech recognition" to "pronunciation teaching" and "teacher training" as time went on. The scholars became very interested in corpus-based pronunciation instruction as a result of the growing interest in these topics. There is also a growing interest in training pre-service and in-service teachers in corpus-based pronunciation research, which will allow them to integrate the corpus into classroom instruction more effectively.

As a matter of fact, corpus pedagogy, as a technology, is not an easy one for conservative teachers to learn and acquire, especially in the early stages. The teacher educators are going to spend more time and effort trying to change the attitude of the teacher-learners as well as helping them learn it.



Conclusion

The most frequently discussed topics in the articles over the past 41 years and the changing trends of the most popular topics, as well as the most prolific authors, countries, and the collaboration between them, have all been identified through bibliometric analysis of the articles published in the WoS and Scopus databases. According to the findings, even though speech analysis and recognition are among the most popular subjects, and with a lot of papers focusing on continuous speech analysis and error detection before 2013, the bulk of older studies have mainly been focusing on listening to learning participants' speech. After 2018, there are a number of new themes emerging, such as teacher training and teaching pronunciation, that will continue to develop. Additionally, the results also show that the contribution of countries other than English speaking nations, such as France, Germany and Japan. It should be pointed out that the discussion in this study has some limitations, due to the fact that only one bibliometric tool was used for the bibliometric analysis, which made the analysis seem one-sided.

Acknowledgement

This research has been supported by the school-level teaching reform research project of Wuhan University of Communication (No. XJ2023140).

References

- Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies*, 24(6), 3731-3743. doi:10.1007/s10639-019-09955-7
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4), 959-975. doi:10.1016/j.joi.2017.08.007
- Bliss, H., Abel, J., & Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction: A review. *Journal of Second Language Pronunciation*, 4(1), 129-153. doi:10.1075/jslp.00006.bli
- Derwing, T. M. (2019). *Utopian goals for pronunciation research revisited*. Paper presented at the Proceedings of the 10th Pronunciation in Second Language Learning and Teaching conference, Ames, IA.
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam/Philadelphi: John Benjamins Publishing Company.
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296. doi:10.1016/j.jbusres.2021.04.070
- Gilakjani, A. P. (2016). English pronunciation instruction: A literature review. *International Journal of Research in English Education*, 1(1), 1-6.
- Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de la Información*, 29(1), 1-20. doi:10.3145/epi.2020.ene.03
- Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied linguistics*, *36*(3), 326-344. doi:10.1093/applin/amu076

88 | https://lprjournal.com/ojs | E-ISSN: 2948-3964 © 2023 | Published by Intelligentia Resources



- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538. doi:10.1007/s11192-009-0146-3
- Wahid, R., & Sulong, S. (2013). The gap between research and practice in the teaching of English pronunciation: Insights from teachers' beliefs and practices. World Applied Sciences Journal, 21(2013), 133-142.
- Zhang, X. (2020). A bibliometric analysis of second language acquisition between 1997and 2018. *Studies in Second Language Acquisition*, 42(1), 199-222. doi:10.1017/S0272263119000573

*Disclaimer: Facts and opinions in all articles published on LPSS Journal are solely the personal statements of respective authors. Authors are responsible for all contents in their article(s) including accuracy of the facts, statements, citing resources, and so on. LPSS Journal disclaims any liability of violations of other parties' rights, or any damage incurred as a consequence to use or apply any of the contents of this journal.