P-ISSN 2708-6453	Pakistan Languages and Humanities Review	October-Dec 2024, Vol. 8, No. 4
O-ISSN 2708-6461	https://doi.org/10.47205/plhr.2024(8-IV)44	[484-493]

Pakistan Languages and Humanities Review
www.plhr.org.pk

RESEARCH PAPER

Corpus-Based Analysis of Lexical Variation in Pakistani Medical Blogs: Exploring Trends in Digital Health Communication

[1]Waiza Azhar, [2]Dr. Fakhar Hussain Malik and [3]Dr. Adnan Tahir*

1. Research Scholar, Department of English, Emerson University, Multan, Punjab, Pakistan
2. Lecturer, Department of Language & Translation Studies, College of Humanities & Social Sciences, Arar, Northern Border University, KSA
3. Associate Professor, Department of English, Emerson University, Multan, Punjab, Pakistan

***Corresponding Author** | adnan.tahir@eum.edu.pk

ABSTRACT

This study investigates lexical variation in Pakistani medical blogs over a period of ten years (2015–2024) using a corpus-driven approach. Lexical variation is automatically measured using the web-based Lexical Complexity Analyzer developed by Lu and Ai (2010). In addition to indices like Type-Token Ratio (TTR), Mean Segmental TTR (MSTTR), and more advanced metrics like CTTR, RTTR, and the Uber Index, the lexical variation across the decade was further evaluated using verb-based variation indices (VV1, SVV1, CVV1) and the remaining five (LV, VV2, NV, ADJV, ADVV, MODV) estimate the lexical word variation within the examined corpus. The results show a moderate level of lexical variation, with some increase in vocabulary richness, particularly reflected in the Uber Index, which showed a gradual rise over time. Although TTR values remained fairly stable, the diversity of verb-based indices and other word categories remained limited, suggesting that while medical blogs evolve in content, they prioritize clarity and consistency in communicating healthcare information.

KEYWORDS Healthcare Discourse, Lexical Complexity, Lexical Indices, Lexical Variation, Pakistani Medical Blogs, Verb Diversity

Introduction

Language is a dynamic system that changes itself by context and according to the communicative needs of particular domains. In healthcare communication, where clarity and accessibility are at par, word choice is significant in conveying, communicating, and understanding medical information. A key element of this is lexical variation. Lexical variation denotes the range of words in written or spoken discourse (McCarthy & Jarvis, 2010; Zhang et al., 2021). Lexical diversity is often used as an equivalent to lexical richness (Daller, van Hout & Treffers-Daller 2003).

Medical blogs are an important medium of health communication whose contents are readily accessible and user-friendly. Medical blogs simplify the complex, raise health awareness, and serve as a medium for healthcare professionals to share their experiences and knowledge (Díaz-Martín et al., 2020). In Pakistan, medical blogs have been acting as a bridge between doctors and the general public because traditional healthcare resources are quite inaccessible. They not only educate readers about medical issues but also act as collaborative spaces for healthcare professionals to share research findings and experiences, contributing to the broader discourse on healthcare practices (Taylor & Humphrey, 2021)

Though lexical variation is gaining much interest across several disciplines, most previous research has been on academic or formal writing. At the same time, medical discourse is often neglected, especially in the Pakistani context. These studies contribute to our understanding of language use in formal settings; relatively little attention has been paid to how lexical variation manifests itself in less formal but important contexts, such as medical blogs. To fill this gap, the present study examines the patterns of lexical variation through the use of computational indices.

This research examines lexical variation in Pakistani medical blogs over ten years (2015–2024). Utilizing a corpus of over one million words, the study adopts a diachronic approach to analyze variation patterns in these blogs. By employing computational tools such as the Lexical Complexity Analyzer (Lu & Ai, 2010), various indices—including Type-Token Ratio (TTR), Mean Segmental TTR (MSTTR), and Uber Index etc—are used to explore patterns of lexical variation throughout the corpus. The analysis focuses on how lexical variation presents across different word classes-for instance, nouns, adjectives, and verbs-across the blogs. The findings of this study aim to provide insights into linguistic variation in medical blogs, a platform that can be useful for further research in healthcare linguistics.

Literature Review

Medical blogs are now a vital medium through which health information is passed. They reflect the linguistic dynamics of professional and layperson communication. Healthcare professionals use medical blogs to share knowledge, experiences, and research results, thus facilitating collaborative learning (Gurtskoy, 2022). Lexical variation, also known as lexical diversity (Zheng, 2016; Treffers-Daller et al., 2018), refers to "the range of different words used in a text" (McCarthy and Jarvis, 2010). Lexical Variation is systematically measured using computational tools such as the Lexical Complexity Analyzer (LCA), developed by Lu (2012) and further extended in a web-based version by Ai and Lu (2010). The LCA calculates 25 measures of lexical richness using pre-processed, part-of-speech-tagged, and lemmatized input files. The current study employs a range of these indices to quantify lexical variation and change in medical blogs over time, focusing on type-token ratios (TTR) and verb-based indices.

The Type-Token Ratio (TTR), a foundational measure, assesses lexical diversity by dividing the number of unique words (types) by the total number of words (tokens) (Templin, 1957). However, because TTR is sensitive to text length, other refined measures are included, such as the Mean Segmental TTR (MSTTR), which divides texts into equal 50-word segments and calculates an average TTR for stability (Johnson, 1944; McCarthy & Jarvis, 2010). Additional refined measures include the Corrected TTR (CTTR), which adjusts TTR by dividing the total tokens by the square root of two times the total words (Guiraud, 1960), the Root TTR (RTTR) that corrects TTR by dividing tokens by the square root of total words (Carroll, 1964), and the Logarithmic TTR (LogTTR), which applies a logarithmic transformation to account for text length variations (Herdan, 1964). The Uber Index is more complex; it combines TTR with a logarithmic transformation to facilitate comparability across text samples (Arnaud, 1992; Lockwood et al., 2018).

In addition to TTR-based measures, this study systematically analyzes lexical variation indices targeting verb diversity and broader lexical categories. For verb-based indices, VV1 calculates the ratio of unique verb types to total verbs (Harley & King, 1989), and sample size effects are addressed by squaring or correcting verb ratios in measures such as its refinements, SVV1 (Squared VV1) and CVV1 (Corrected VV1), Chaudron & Parker, 1990; Wolfe-Quintero et al., 1998. The VV2 (Verb Variation II) index calculates verb

tokens as a ratio to the total lexical words (McClure, 1991). Similar approaches are also used in determining noun and adjective variation by employing indices, like NV for Noun Variation, which indicates the quotient of noun types by lexical words. Other similar indices for evaluating diversity are AdjV or Adjective Variation, AdvV or Adverb Variation, and ModV for Variation of Modifier (McClure, 1991). The indices, therefore, result in an organized statistical interpretation of the lexical diversity over time. Though research provides insights into lexical variation in many contexts, there is still a lack of studies on lexical variation in Pakistani medical blogs. This study fills this gap by examining these computational indices in medical blogs from 2015 to 2024.

Material and Methods

Research Design

This study used a mixed-method, corpus-driven approach to analyze lexical variation in Pakistani medical blogs between 2015 and 2024. The analysis was mostly quantitative, employing the Lexical Complexity Analyzer to calculate lexical diversity with different indices. Along with the quantitative approach, qualitative research was carried out to establish context around the results, allowing interpretation from the perspective of the healthcare discourse trends. In this regard, by integrating the above two approaches, the study aims to offer a balanced outlook over nearly ten years regarding the development of lexical variation within medical blogs.

Data Collection

The data were extracted from various Pakistani medical blogs, which fall under several categories, including Dental Health, Diseases and Disorders, Drugs and Medications, Eye Care, and many others. Two methods were used to extract the blog data. First, the manual collection method was applied, whereby blogs written in English with substantial content concerning healthcare were accessed manually, reviewed, and classified based on their relevance to the study's objectives. Web scraping tools were also used to maximize efficiency and increase data volume. These tools streamlined collecting, filtering, and systematizing a more general data set for further analysis.

Sampling

For collecting data, purposive sampling was used to choose relevant medical blogs from categories such as mental health, diseases and disorders, drugs and medications, and many more health-related topics. These were accessed manually and reviewed to ensure they contained enough healthcare-related content in English. Different samples were randomly selected from each year between 2015 and 2024 to analyze the data. This approach, which involved purposive sampling for data collection and random sampling for analysis, allowed for a thorough and objective examination of lexical variation in Pakistani medical blogs.

Data Pre-processing and Corpus Compilation

After collecting all the data, AntFileConverter converted the text into plain text format to work with corpus analysis tools. The data was cleaned in pre-processing by removing unwanted elements such as HTML tags, embedded links, ads, and irrelevant content. In this way, the relevant medical content from the blogs was focused on. The data set was organized in folder-by-year style to facilitate the diachronic (yearly) analysis. The final corpus was made using the following two main tools: (1) for Sketch Engine the corpus size was adjusted to 999,344 words while maintaining a balance across the years. Also,

yearly subcorpora were created within a unified corpus; (2) for AntConc (4.3.1), separate yearly corpora were created to facilitate diachronic analysis of lexical variation. A complete corpus containing about 1,251,452 words from about 1,500 blog files was prepared for a holistic analysis.

Data Analysis Framework

To analyze lexical variation, the study used the web-based Lexical Complexity Analyzer developed by Lu and Ai to calculate Type-Token Ratio (TTR) and its variants, including Root TTR (RTTR), Corrected TTR (CTTR), Mean Segmental TTR (MSTTR), Uber Index, as well as verb-based indices such as VV1, CVV1, SVV1, VV2, and other lexical measures like Lexical Word Variation (LVW), noun variation (NV), adjective variation (ADJV), adverb variation (ADVV), and modifier variation (MODV), thus ensuring a comprehensive and reliable analysis. The mean values and standard deviations were used to provide the reliability and authenticity of the results. The information regarding the indices used for measuring lexical variation is presented in the table below:

Table 1
Lexical Variation Indices (Adapted from Lu, 2012)

Measure	Code	Formula
Type-Token Ratio	TTR	TTR= Types/Tokens
Mean Segmental TTR (50)	MSTTR	Mean TTR of all 50-word segments
Corrected TTR	CTTR	$T/\sqrt{2N}$
Root TTR	RTTR	$T/\sqrt{N}$
Bilogarithmic TTR	LogTTR	LogT/LogN
Uber Index	Uber	$Log^2N/Log(N/T)$
Lexical Word Variation (LV)	LWV	Tlex / Nlex
Verb Variation-I (VVI)	VV1	Tverb / Nverb
Squared VVI (SVVI)	SVV1	T^2 verb /Nverb
Corrected VVI (CVVI)	CVV1	Tverb /$\sqrt{2}$Nverb
Verb Variation-II (VVII)	VV2	Tverb / Nlex
Noun Variation (NV)	NV	Tnoun / Nlex
Adjective Variation (AdjV)	AdjV	Tadj / Nlex
Adverb Variation (AdvV)	AdvV	Tadv / Nlex
Modifier Variation (ModV)	ModV	(Tadj+ Tadv) / Nlex

Ethical Considerations

This study adhered strictly to ethical guidelines. All the blog data featured in this research was already in the public domain, and no private information features were included. The data used had only academic purposes, and the integrity of the content was fully respected regarding the authors' work. Care was taken regarding source confidentiality and privacy.

Data Analysis

The analysis of lexical variation in Pakistani medical blogs from 2015 to 2024 was carried out using various indices to measure the complexity and diversity of the text. All these measures were computed using the web-based Lexical Complexity Analyzer. Due to their sensitivity to text length, appropriate sample sizes have been selected for analysis. The cumulative table clearly shows mean values of lexical variation indices in Pakistani medical blogs during the ten years, from 2015 to 2024, with standard deviation (in parentheses) indicating the consistency of the results.

Table 2

Lexical Variation in Pakistani Medical Blogs (2015–2024)

Index	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
TTR	0.02 (0)	0.02 (0.006)	0.02 (0)	0.02 (0)	0.02 (0.006)	0.03 (0)	0.037 (0.006)	0.04 (0)	0.04 (0)	0.04 (0)
MSTTR	0.357 (0.006)	0.36 (0.01)	0.36 (0.006)	0.37 (0)	0.37 (0)	0.37 (0.01)	0.37 (0.006)	0.38 (0.006)	0.37 (0)	0.373 (0.006)
CTTR	0.473 (0.006)	0.47 (0.02)	0.51 (0.07)	0.47 (0.03)	0.49 (0.11)	0.53 (0.04)	0.53 (0.06)	0.58 (0.02)	0.56 (0.02)	0.59 (0.04)
RTTR	0.673 (0.006)	0.67 (0.02)	0.72 (0.11)	0.66 (0.04)	0.75 (0.09)	0.75 (0.07)	0.73 (0.09)	0.83 (0.06)	0.75 (0.09)	0.85 (0.06)
Log TTR	0.45 (0)	0.44 (0.006)	0.45 (0.02)	0.44 (0.006)	0.46 (0.02)	0.46 (0.01)	0.46 (0.02)	0.47 (0.006)	0.47 (0.01)	0.48 (0.01)
Uber	5.79 (0.006)	5.79 (0.05)	5.91 (0.19)	5.77 (0.05)	6 (0.19)	5.94 (0.15)	5.92 (0.23)	6.14 (0.12)	6 (0.07)	6.19 (0.19)
VV1	0.013 (0.006)	0.02 (0.006)	0.01 (0.006)	0.02 (0.01)	0.02 (0.006)	0.02 (0.01)	0.027 (0.006)	0.02 (0)	0.02 (0)	0.02 (0)
SVV1	0.013 (0.006)	0.02 (0.006)	0.01 (0.006)	0.02 (0.01)	0.02 (0.006)	0.02 (0.01)	0.027 (0.006)	0.02 (0)	0.02 (0)	0.02 (0)
CVV1	0.083 (0.006)	0.63 (0.46)	0.1 (0.01)	0.10 (0.01)	0.1 (0.01)	0.10 (0.02)	0.097 (0.012)	0.10 (0.01)	0.09 (0.006)	0.10 (0.01)
LWV	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)
VV2	0.00 (0)	0.00 (0)	0.00 (0)	0.0 0 (0)	0 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
NV	0.00 (0)	0.007 (0.006)	0.007 (0.006)	0.007 (0.006)	0.01 (0)	0.003 (0.006)	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)
AdjV	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
AdvV	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
ModV	0.00 (0)	0.01 (0)	0.007 (0.006)	0.007 (0.006)	0 (0)	0.007 (0.006)	0.01 (0)	0.01 (0)	0.01 (0)	0.01 (0)

Multiple random samples from each year were taken and analyzed to ensure reliability and mean values were calculated, which gives the overall lexical variation value for each year. So, to add authenticity, standard deviations were also calculated while summarizing the results, indicating consistent results over samples. The combined table reveals notable trends and patterns in lexical variation indices over the ten years (2015–2024) in Pakistani medical blogs. Here is an overview of key observations:

TTR (Type-Token Ratio)

TTR improves slightly with time and ranges from 0.02 in 2015 to 0.04 in 2024. Although the trend is minimal, it does convey a gradual improvement in lexical diversity, especially in recent years.

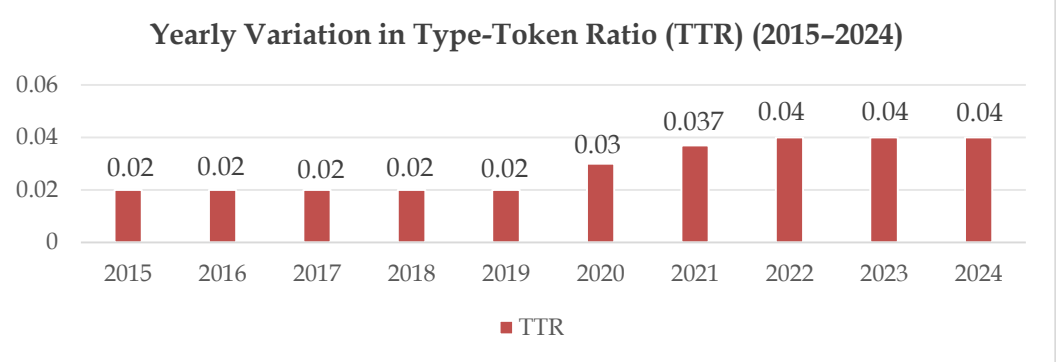

Figure 1: Trends in Lexical Variation using TTR

MSTTR (Mean Segmental TTR)

MSTTR scores are relatively constant, from 0.357 in 2015 to 0.373 in 2024. This stability indicates that lexical diversity is constant at the segmental level throughout the corpus.

CTTR (Corrected TTR)

CTTR shows a clear upward trend, from 0.473 in 2015 to 0.59 in 2024. This increase shows that vocabulary richness and lexical complexity have improved in later years.

RTTR (Root TTR)

RTTR also indicates a development trend through the years, expanding from 0.673 in 2015 to 0.85 in 2024. This increase reflects better lexical consistency and complexity across the blogs under evaluation. Here is the graph of comparison between RTTR and CTTR since both of the indicators have trends in them over the duration:

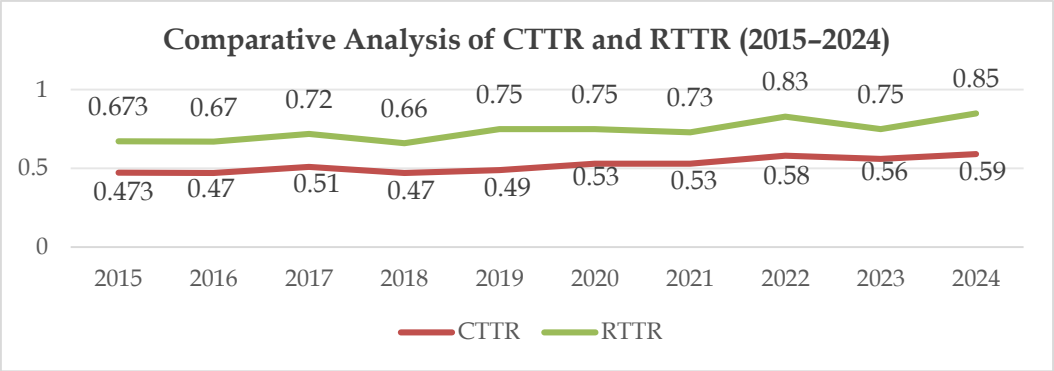

Figure 2: Comparative analysis of CTTR and RTTR

LogTTR:

LogTTR shows a smooth, steady increase from 0.45 in 2015 to 0.48 in 2024, thus further confirming the observation of gradual improvement in lexical diversity over time.

Uber Index

The Uber Index showed the most significant growth, increasing from 5. 79 in 2015 to 6. 19 in 2024. Such a big difference shows the emergence of lexical diversity and richness, especially in the last years of the data.

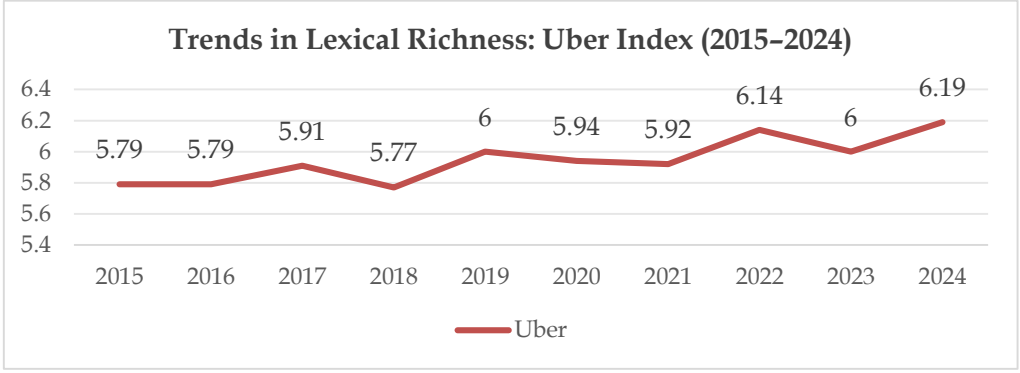

Figure 3: Lexical Variation using Uber Index (2015-2024)

Verb-Based Indices (VV1, SVV1, CVV1)

The verb-based indices (VV1, SVV1, CVV1) show minimal variation, maintaining low values throughout the years. For example, VV1 is around 0. 02 without any rising trends. Such stability of these values suggests limited diversity in verb usage across the blogs.

Lexical Word Variation (LWV)

Lexical Word Variation stays at 0.01 for all years, indicating no significant change in general lexical word variation.

Noun, Adjective, Adverb, and Modifier Variation (NV, AdjV, AdvV, ModV)

Noun Variation (NV) has risen marginally from 0.00 in 2015 to 0.01 in recent years. ModV shows little variation, rising from 0.007 to 0.01 in the later years. AdjV and AdvV are at 0.00 throughout the period; thus, these categories have little diversity.

Diachronic analysis of variation in the Pakistani medical blogs gives steady yet persistent trends within the decade from 2015 to 2024. For this purpose, indexes like CTTR and RTTR, along with the Uber Index, present a gradual ascendency in the trend, depicting uniform improvement in lexical variety and richness. For instance, the RTTR grew from 0.673 in 2015 to 0.85 in 2024, but the Uber Index showed great growth during the same period, from 5.79 to 6.19. Such changes may imply greater shifts in medical vocabulary, determined by global healthcare developments and local linguistic styles. In contrast, indices such as MSTTR are relatively flat, so lexical diversity does not change much between segments. Interestingly, the verb-based indices (VV1, SVV1, CVV1) and adjective variation (AdjV) are constant, signifying lesser stylistic variation in using verbs and descriptive lexis. This suggests that even though lexical diversity is increasing, the technical and specialized nature of the medical discourse limits the scope for variation. The data presents a dynamic yet controlled evolution of lexical patterns of Pakistani medical blogs.

Discussions

The lexical diversity trends featured in Pakistani medical blogs from 2015 to 2024 thus systematically answer research questions. In response to the first research question, there is moderate lexical variation in the whole corpus as presented by its indices, such as the CTTR, RTTR, and the Uber Index. All these exhibit annual changes but are stable in the period examined. CTTR ranges between 0.473 and 0.59, while RTTR is relatively high at 0.673 to 0.85. Indices of this kind indicate a form of balanced lexical diversity and, therefore, signal technicality in medical blogs in which accuracy and homogeneity of linguistic use are noted. The Mean Segmental Type-Token Ratio (MSTTR) also holds constant at 0.357 and 0.373. This would suggest, though there is a degree of lexical diversity within certain contexts, that overall, the discourse is set firmly enough within conventional and standardized medical vocabulary to limit uncontrolled divergence. Other measures, namely noun usage (NV) and adjectives (ModV), change little and tend to reinforce the controlled nature of the linguistic configuration.

The second question of interest shows general trends and rises in individual indices for the year. However, CTTR and RTTR offer impressive growth that continues to evolve in diversification and lexical density during 2015-2024. Uber Index grew from 5.79 in 2015 to 6.19 in 2024, pointing out general growth in vocabulary due to themes and trends of

healthcare. These new trend emergences could probably be elements of medical adaptation cutting across more health-related or general medicine as a broad area influenced, perhaps by the processes of globalization, the evolution of technologies, and problems in societal health. However, adjective variation (AdjV) and verb variation (VV1, SVV1, CVV1) consistently standardize creative and action-oriented wording in medical writings. The Pakistani medical blogs reflect average lexical variation across most indices but display increased diversity by CTTR, RTTR, and the Uber Index, showing this discourse's ability to cope with changing topics and audiences. This variation and consistency balance would ensure that the discourse was accessible and professional enough for the communicative needs of the healthcare context.

Conclusion

This study analyzed the lexical variation in Pakistani medical blogs over ten years, from 2015–2024, with substantial patterns in healthcare discourse. The analysis of lexical variation, with measures such as Type-Token Ratio (TTR), Mean Segmental Type-Token Ratio (MSTTR), Cumulative Type-Token Ratio (CTTR), Running Type-Token Ratio (RTTR), and the Uber Index, etc., depicted moderate diversity in the vocabulary used over the years. These metrics were calculated using Lexical Complexity Analyzer's automated tool, which shows continually rising lexical variation, highly visibly marked over time. While the core medical vocabulary remained relatively stable, there was a distinct enhancement in lexical richness, particularly after 2020. This increase in variation reflects a broader trend of expanding lexical resources within the medical discourse of Pakistani blogs. The study highlights the dynamic nature of lexical variation, providing valuable insights into the evolving language use patterns in Pakistan's online healthcare communication.

References

Ai, H., & Lu, X. (2010). *A web-based system for automatic measurement of lexical complexity* [software]. Paper presented at the 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10), Amherst, MA, June 8–12.

Arnaud, P. J. L. (1992). Objective lexical and grammatical characteristics of L2 written compositions and the validity of separate-component tests. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 133–145). MacMillan.

Carroll, J. B. (1964). *Language and thought.* Prentice-Hall.

Chaudron, C., & Parker, K. (1990). Discourse markedness and structural markedness: The acquisition of English noun phrases. *Studies in Second Language Acquisition, 12*(1), 43–64.

Daller, H., van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics, 24*(2), 197–222. https://doi.org/10.1093/applin/24.2.197

Díaz-Martín, A. M., Schmitz, A., & Yagüe Guillén, M. J. (2020). Are health e-mavens the new patient influencers? *Frontiers in Psychology, 11*, 779. https://doi.org/10.3389/fpsyg.2020.00779

Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique* [Problems and methods of statistical linguistics]. D. Reidel.

Gurtskoy, L. D., Zudin, A. B., & Vvedenskiy, A. I. (2022). The health control of employees in digitization conditions. *Problems of Social Hygiene, Public Health and History of Medicine*.

Harley, B., & King, M. L. (1989). Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition, 11*(4), 415–440.

Herdan, G. (1964). *Quantitative linguistics.* Butterworths.

Johnson, W. (1944). Studies in language behavior: I. A program of research. *Psychological Monographs, 56*(2), 1–15.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

McClure, E. (1991). A comparison of lexical strategies in L1 and L2 written English narratives. *Pragmatics and Language Learning, 2*, 141–154.

Taylor, K. A., & Humphrey, W. F. Jr. (2021). Impact of medical blog reading and information presentation on readers' preventative health intentions: Mixed methods, multistudy investigation. *Journal of Medical Internet Research, 23*(12), e23210. https://doi.org/10.2196/23210

Templin, M. C. (1957). *Certain language skills in children: Their development and interrelationships.* University of Minnesota Press.

Treffers-Daller, J., & Williams, S. (2016). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics, 39*, amw009. https://doi.org/10.1093/applin/amw009

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity* (Report No. 17). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

Zhang, H., Chen, M., & Li, X. (2021). Developmental features of lexical richness in English writings by Chinese beginner learners. *Frontiers in Psychology, 12*, 665988. https://doi.org/10.3389/fpsyg.2021.665988

Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System, 56*, 40–53. https://doi.org/10.1016/j.system.2015.12.004