CONTENT ANALYSIS OF CORPUS RESEARCH ACROSS FIVE YEARS: STUDY OF JOURNAL ARTICLES (CORPORUM 2020-2024)

Fatima Tuz Zahra

PhD Scholar, Air University Islamabad, Lecturer, Minhaj University, Lahore
Email: fatematuzzehra@gmail.com

Dr. Tehseen Zahra

Associate Professor, Bahria University, Islamabad
Email: tehsin.azhar@yahoo.com

Memoona Saleh

Email: bintesaleh1599@gmail.com

Abstract

Corpus linguistics have provided ample evidence of linguistic features in the form of corpora. Most of the linguistic research published in the journals have used corpus as a method or a tool but none have reviewed and identified popular genres of discourse and how Pakistani English corpora are evaluated. Current research analyses the published articles in the leading journal (CORPORUM) and reports on the utilization of genres for research purposes. It employs quantitative content analysis of articles published between 2020-2024. Articles are classified into four dimensions covering topical content and discourse genre. Finally, corpus of Pakistani English, PakLocCorpus is evaluated for the presence and absence of different genres of discourse. Findings indicate that media discourse has become the most popular genre in corpus research. Also, it was observed that PakLocCorpus is rich in all genres of Pakistani English discourse.

Keywords: *Corpus linguistics, discourse genre, corpora, Pakistani English*

1. Introduction

Corpus linguistics is widely recognized as a transformative approach to linguistic inquiry, enabling systematic exploration of language through large, machine-readable text collections (Hunston 2002; Sinclair 2004). Over time, the field has diversified in both theme and technique, spanning lexical studies, discourse analysis, academic writing, translation, computational linguistics, and beyond. In Pakistan, Corporum: Journal of Corpus Linguistics (biannual, peer-reviewed) has emerged as a key platform for corpus-based research. *Corporum* aims to promote corpus research nationally and internationally, offering visibility especially to studies on Pakistani English, Urdu, bilingual phenomena, discourse in local media, and academic writing within regional contexts. A meta-analysis of corpus research published in *Corporum* is significant for several reasons. First, it allows mapping of local research priorities—revealing which languages, genres, and phenomena are foregrounded. Second, it sheds light on methodological trends (e.g., adoption of annotation schemes, statistical/ computational tools) and potential constraints faced by Pakistani researchers. Third, by exposing underexplored areas, such an analysis can guide future research directions (for example, the relative scarcity of spoken corpora or computational deep learning methods in *Corporum*'s published work). For instance, Afifa, Ghulam Fatima, and Sajjad Rasool (2025) observed lexical bundles in legal discourse and Mumtaz, Zahra and Bhatti (2025) focused on political discourse in print media. Lastly, it contributes to the reflexivity of the discipline by viewing *Corporum* not merely as a venue but as an evolving corpus-linguistic community.

Many general corpus-linguistics journals focus heavily on English or widely studied languages and advanced computational methods. *Corporum*'s specialization in local or regional corpora

(e.g. Pakistani English, Urdu, bilingual discourses) allows us to see how corpus methods are adapted for non-dominant languages and regional contexts. Understanding what *Corporum* emphasizes (themes, language pairings, genres) helps situate Pakistani corpus scholarship within global trends. By investigating which corpus tools, annotation schemes, and statistical / computational techniques are used over time, one can observe how accessible or cutting-edge methods diffuse among local scholars. It may also reveal resource constraints (e.g. lack of large corpora, less use of deep learning) or methodological conservatism. A meta-study of *Corporum* can thus expose to what extent the authors adopt global innovations or remain anchored in more classical corpus practice.

Current study is a systematic review, and it highlights genres that are represented and underrepresented (for instance, spoken corpora, cross-lingual corpora, multimodality) in *Corporum*. This in turn can guide researchers toward fruitful new areas. Moreover, such a review strengthens the standing of *Corporum* by showing its cumulative contributions and potential growth areas. Meta-analyses of discipline-specific journals contribute to reflexivity that is, the field thinking about itself. Just as general linguistic disciplines periodically examine their own publishing trends, a focused look at *Corporum* encourages Pakistani corpus researchers to reflect on their collective profile, strengths, and blind spots. Findings can be used to evaluate existing corpora or plan future research.

Research Questions

Current study seeks the answers of following research questions:

1: How do articles published in *Corporum* reflect the use of Corpus approaches across diverse genres of Pakistani English discourse?

2: What range of discourse genres and registers are represented in *Corporum* publications, and how do these correspond to the genre-based structure of *PakLocCorpus* as analysed through Biber's register theory?

3: In what ways does *PakLocCorpus*, within the framework of World Englishes, serve as a comprehensive and purpose-built resource for documenting and analysing the evolving discourse of Pakistani English?

2. Literature Review

Qualitative research contributes to an understanding of the human condition in different contexts and of a perceived situation. However, there is no perfectly designed study, and unexpected events will always appear. The main issue is how much financial resources, time and effort the researchers in a study team are able to invest in trying to understand the phenomena under study (Patton, 2002). Nevertheless, a researcher has to create the best study design possible, through accurate and considerate planning based on existing circumstances by identifying available resources. Firstly, there is the mapping of external resources, such as economics, time and potential informants, because some methods for data collection and data analysis are cost and time-consuming, and the choices of methods must be adapted to such restrictions. Second, internal resources of the study team, such as knowledge and ability, have to be known. The researcher's self-reflection is an essential part of qualitative research whatever chosen qualitative method (Burnard, 1995). The researcher must take into consideration his or her "pre-understanding", both in the planning process as well as during the analyzing process, in order to minimize any bias of his/her own influence (Elo et al., 2014, Long and Johnson, 2000). To have preconceived knowledge of the subject and to be familiar with the context can be an advantage as long as it does not affect the informants or the interpretation of the results. The researcher needs to understand both the context

and circumstances in order to detect and take into account misrepresentations that may crop up in the data

Corpus research has expanded over the past few years and it has covered several major domains of research. Most of these studies have applied corpus building research strategies, and only a few have used content analysis. Current study used qualitative inquiry and aimed at identifying the most popular genres of Pakistani English discourse as reported in the prestigious journal of corpus linguistics (CORPORUM). It has published research articles on the related domain, since 2018. The current research aimed to identify the most popular genres of written discourse of Pakistani English. Qualitative research contributes to an understanding of the human condition in different contexts and of a perceived situation (Bengtsson, 2016). However, there is no perfect designed study, and unexpected events will always appear. The main issue is how much financial resources, time and effort the researchers in a study team are able to invest in trying to understand the phenomena under study (Patton, 2002). For this purpose, research articles published between 2019-24 were downloaded and analyzed using summative qualitative framework (Rapport, 2010). Out of the 21 articles published in three latest issues of the journal, the articles focused on spoken discourse were excluded. Over the period of time many remarkable corpora are built and reported in the journal. As it creates a space for such content and has a significant in the mark of corpus research, it can be used to identify the most popular genres of written discourse, as utilized by the local scholars.

As corpus linguistics have been extensively utilized, both in terms of methodology or tools, evaluation of the research articles will ensure the quality of the scholarly contributions. Rapport (2010) presented the technique of summative analysis that is qualitative in description and used in several investigations across disciplines. As a collaborative, analytical procedure, it focused on consensus-building activities. It highlighted trauma suffering of Holocaust survivors and how these testimonies are presented. Majhi et al. (2016) observed journal articles on Wiki and presented that most of the publications exploited survey procedures. Preferred method of inquiry used in majority of the studies is descriptive data analysis. Mohammed and Raewf (2018) offered a clear picture of research methodology, sample size, participants' profile, analysis technique and data collection instruments implemented in the publications investigated for Language Learning Strategies (LLS).

Järvelin and Vakkari (2022) illuminate the status of the studies as published in Library and Information Science (LIS). How theses studies are evolved in the past years and what methodologies were used to report findings. This groundbreaking work highlight that the scientific and professional communication is now more in focus with the discussion of information storage and retrieval. Recent work of Alyateemei (2024) shows how technological development is joining hands with corpus linguistics (CL). The findings reveal a rapid development of the field in terms of practical and methodological perspectives, specifically regarding the investigations of language uses in different contexts. Thus, this paper indicates a significantly strong correlation between CL and technological development, such as natural language processing (NLP), and how this approach could fill the research gap of utilising CL in other areas of linguistics. The correlation between both is significant for natural language processing (NLP), and how this approach could fill the research gap of utilizing CL in other areas of linguistics. As field of CL is expanding and it has continued to provide an insight on the linguistic landscape of local varieties, there is a need to identify the most popular genres of discourse and how these are analyzed in the research over the last five years. By analyzing large collections of data, corpus linguists are able to uncover linguistic

findings that were previously inaccessible through traditional methods. A major strength of applying CL is providing empirical evidence for linguistic theories through real-life examples. Similar studies have focused attention towards reporting the content of the scientific observations and how these reports are evolved over the past few years. Though content analysis is concentrated on the description of the findings, none have investigated the published research on the area of corpus linguistics.

Corpus Linguistics have provided a new dimension of research to the discourse analysts and the scholars have built personal and specialized corpora to reflect the unique features of Pakistani English discourse. CORPORUM is the only journal in Pakistan that is focused on publishing articles that are using corpus from one way to another. Journal articles reflect that the editors have deliberately published research based on both spoken and written corpora. Also, it has published papers reporting on the corpus of other local varieties or the corpus of Urdu. This wide range motivate scholars from different universities to contribute significantly in the field of corpus linguistics. Corpus tools and methodologies are extensively utilized to identify, compare or analyze unique linguistic features of language data sets. In addition, some scholars have attempted to design purpose-built corpora and tools for the analysis. Over the period of time many remarkable corpora are built and reported in the journal. As it creates a space for such content and has a significant in the mark of corpus research, it can be used to identify the most popular genres of written discourse, as utilized by the local scholars. Though there have been several studies on content analysis, identifying the popular genres of journal articles is yet to be explored. Zahra et al. (2023a) discussed neoliberal cooption of English using Pakistani English discourse. It utilized PakLocCorpus for the analysis (Zahra et al., 2022).

Current investigation exploited summative content analysis (Rapport, 2010) for the investigation. Content analysis is a widely used qualitative research technique. Rather than being a single method, current applications of content analysis show three distinct approaches: conventional, directed, or summative. By analyzing data sets, corpus linguists provide evidence-based insightful information about the unique linguistic features of language varieties, and these features cover a specific area of the discourse. Zahra et al. (2023b) observed collocation patterns in literary discourse. Discourse can be written or spoken, and the data can be of any language or even translated text, depending on the research objectives. Reviewing the journal articles not only provide insightful information about the evolution of the publications but also present the perspectives of the scholars. The popular tools of analysis, theoretical frameworks and developmental challenges are also reported in the content analysis.

3. Methodology

Current study uses a mixed-methods, systematic review design combining quantitative content analysis and qualitative discourse analysis. The goal is to identify the most frequently published genres in *Corporum*, describe analytic focus across genres. This study is grounded in the theoretical framework of Corpus-Assisted Discourse Studies (CADS), drawing on Biber's register theory (1988) to categorize genres and supported by the World Englishes model (Kachru, 1985) to contextualize Pakistani English as a distinct linguistic variety. The integration of these frameworks enables a systematic exploration of discourse genres as represented in *Corporum* and captured in *PakLocCorpus*, highlighting their role in advancing corpus-based research on Pakistani English. All peer-reviewed research articles published in five years, *Corporum: Journal of Corpus Linguistics* between 2020 and 2024. Original empirical articles using corpus methods, either fully corpus-based or corpus-assisted.

ISSN E: 2709-8273
ISSN P:2709-8265

JOURNAL OF APPLIED LINGUISTICS AND TESOL

JOURNAL OF APPLIED
LINGUISTICS AND
TESOL

JALT

Vol.8. No.4.2025

This study follows a qualitative corpus review design combining content analysis and comparative evaluation.

- First, a systematic review of articles published in *Corporum* between 2020 and 2025 was conducted to identify recurrent genres, research themes, and analytical methods.
- Each article was coded according to genre focus (e.g., political, academic, digital, legal, or literary) and corpus methodology used.
- Next, the genre coverage of *PakLocCorpus* was reviewed using available metadata and documentation, focusing on the representation of written and spoken registers, institutional domains, and discourse types.
- The findings were interpreted through Biber's register theory to determine genre diversity and through CADS principles to identify how discourse is operationalized in empirical studies.
- Finally, both datasets were evaluated within the World Englishes framework to assess their contribution to the description and legitimization of Pakistani English.

For close reading, search and retrieval online archive of the journal was accessed and all issues of the selected years were downloaded along with metadata (volume, number, year). Download full texts or PDFs of all articles that meet inclusion criteria. Studies using corpus tools or methodologies were selected. Data log was also maintained (spreadsheet) with bibliographic metadata: author(s), year, title, volume/issue, pages, DOI/URL, abstract, keywords.

4. Findings and discussion

Corporum has significantly advanced the adoption of CADS methodologies in Pakistan. Studies frequently apply corpus tools to explore ideology in media discourse, stance in academic writing, and identity in political communication. This methodological consistency underscores an emerging scholarly culture that values data-driven discourse inquiry. The combination of corpus linguistics and discourse analysis reflects the global CADS movement's emphasis on linking linguistic patterning with social meaning. As such, *Corporum* has established itself as a key channel for promoting methodological literacy and critical engagement among Pakistani corpus researchers.

Table 1: Dispersion of published research articles across five years

Sr no	Genre	2020	2021	2022	2023	2024
1	Academic Discourse	3	2	3	1	2
2	Media Discourse	1	1	3	2	3
3	Literary Discourse	1	1	1	2	1
4	Others	0	1	0	0	2
	Total	5	5	7	5	8

Table above shows that across the five years and corresponding first issues of *Corporum: Journal of Corpus Linguistics*, a total of 30 research articles were published, covering four broad discourse categories: academic (11), media (10), literary (6), and others (3). The data reveal a consistent pattern of genre inclusivity, as each issue publishes at least one article from each discourse category, reflecting the journal's commitment to representing multiple domains of language use within corpus-based research. Academic discourse remains the most dominant genre, accounting for approximately 37% of the total publications. This trend aligns with the journal's central focus on advancing corpus methodologies and their pedagogical and scholarly applications. Studies in this category commonly explore research writing, citation practices, and meta discourse, themes

that mirror *Corporum*'s alignment with Biber's (1988) Register Theory, emphasizing structured, formal, and research-oriented registers. This dominance also demonstrates how *Corporum* serves as a scholarly bridge between applied linguistics and corpus pedagogy in Pakistan. The media discourse category**,** with 10 publications (33%)**,** highlights growing interest in language use within public and institutional communication, such as newspapers, political discourse, and online journalism. These studies often employ Corpus-Assisted Discourse Studies (CADS) approaches to uncover ideological patterns, stance markers, and evaluative language in Pakistani media. The literary discourse category**,** comprising 6 articles (20%)**,** illustrates the versatility of corpus methods in stylistics and literary linguistics. Through keyword, collocation, and lexical pattern analysis, these studies reveal stylistic distinctiveness in fiction, poetry, and translation, broadening the scope of corpus linguistics beyond traditional language teaching or media studies. The 'others' category**,** with only 3 publications (10%)**,** includes specialized or cross-disciplinary studies such as computational corpus design, learner corpora, or multimodal text analysis. Though numerically small, this category reflects innovation and experimentation within the field, suggesting emerging research directions that extend corpus linguistics toward digital humanities and technology-based inquiry. In summary, the findings suggest that *Corporum* maintains a balanced representation of discourse genres, though academic discourse remains the core of its publication strategy. The consistent inclusion of media, literary, and specialized studies across all issues reveals a deliberate editorial approach to inclusivity. This genre diversity, coupled with cross-corpus engagement and local authorship, positions *Corporum* as a vital scholarly forum for documenting and analyzing Pakistani English within the broader World Englishes paradigm**.**

Genre Representation and Register Diversity

Using Biber's Register Theory (Biber, 1988) as a lens, the review shows that *Corporum*'s research output is strongest in academic and journalistic registers, with limited studies on bureaucratic, literary, and spoken genres. In contrast, *PakLocCorpus* provides a systematic and inclusive representation of Pakistani English genres, encompassing news, academic texts, literature, legal discourse, bureaucratic writing, and online communication. The corpus's balanced design positions it as an essential empirical foundation for future linguistic, pedagogical, and sociolinguistic analysis. Its comprehensive register coverage also enables comparative cross-genre studies that were previously constrained by data availability.

The comparison reveals strong genre alignment between *Corporum*'s published research and the textual data available in the PakLocCorpus. Both focus heavily on academic, journalistic, and institutional discourse, reflecting the formal nature of Pakistani English (Mukherjee, 2020). It provides a systematic analysis of past and present corpus-based research into various Asian Englishes, moving from the perspective of individual varieties and regional clusters of Asian Englishes to a more comparative and global view of Asian Englishes and world Englishes. However, significant gaps remain in representing spoken and digital communication, areas increasingly shaping linguistic behavior in Pakistan.

While *Corporum* serves as an interpretive and analytical platform, offering discourse-analytic and corpus-based perspectives, PakLoc remains a data-oriented repository. For instance, recent studies in *Corporum*, such as Mumtaz, Zahra, and Bhatti (2025) and Zahra et al. (2024) explored the globalization of narratives through corpus assisted discourse analysisImportantly, *Corporum* demonstrates a growing academic interest in policy, institutional, and gender discourse (e.g. *A Corpus Analysis of Gender Harassment Policies*, *Constructing Gender through Metadiscourse*, Azher, Jahangir, & Mahmood, 2023), signaling a shift from purely linguistic description to socially

engaged corpus linguistics."In contrast, PakLocCorpus' current composition underrepresents dynamic communicative forms such as blogs, advertisements, and spoken interactions, which are vital for understanding how English functions in everyday Pakistani contexts. The PakLocCorpus is one of the most significant linguistic resources representing Pakistani English. It was designed to document the lexical, grammatical, and discourse-level features of Pakistani English through a balanced collection of written and spoken texts. As noted in Zahra et al. (2024) and Zahra et al. (n.d). The corpus includes academic writing, journalistic discourse, institutional documents, and a smaller sample of creative texts. The resource aims to provide a systematic foundation for corpus-based linguistic research in Pakistan and to support comparative studies with other regional varieties of English. According to its description (PakLocCorpus, 2023), the corpus comprises three main categories and several genres:

- Academic and professional writing (research papers, theses, reports, student essays)
- Media texts (newspapers, editorials, features)
- Literary and stylistic texts (fiction, poetry, short stories)

Table 2: Comparison of Genre Type Representation

Genre Type	Representation in *Corporum*	Representation in PakLocCorpus	Remarks
Academic Discourse	Highly represented	Well represented	Both emphasize scholarly writing, though *Corporum* adds interpretive analysis.
Media Discourse	Highly represented	Strongly represented	Common in both; often used for CDA and discourse framing studies.
Literary Discourse	Moderate	Well represented	*Corporum* analyses literary style; PakLocCorpus stores sufficient samples.
Legal Discourse	Moderate	Absent	PakLocCorpus does not include policy and legal texts; *Corporum* often analyses them critically.
Social Media Discourse	Limited	Absent	Both underrepresent digital genres.
Specialized Genres	Limited	Absent	*Corporum* includes newer institutional genres not yet captured in PakLocCorpus.
Spoken Discourse	Limited	Absent	A major gap in both resources.

Table 2 shows comparison of *Corporum* is fulfilling an important role in the corpus linguistics landscape of Pakistan: it reflects both academic rigor and growing responsiveness to social, institutional, and policy discourse. As the field moves forward, expanding genre coverage and method diversity will be crucial to capturing the full complexity of language in society. This review provides a baseline against which future progress can be measured.

Pakistani English within the World Englishes Paradigm

From the perspective of the World Englishes framework, there is sufficient discussion available on the discourse features of local corpus. Publication in *Corporum* provides theoretical and analytical discussions of how English functions within Pakistan's multilingual context. Studies provide an in-depth analysis of localized variety and unique linguistic features of Pakistani English. *PakLocCorpus* supplies the empirical data necessary to describe its structural, lexical, and pragmatic features. Together, they illustrate the localized adaptation of English and highlight how corpus research can promote linguistic equity by validating non-native Englishes within global scholarship.

Future Implications

Future studies should leverage *PakLocCorpus* for advanced analyses of genre variation, linguistic identity, and discourse strategies in Pakistani English. Expanding corpus development to include spoken and multimodal data will enhance the representativeness of national English usage. The integration of digital discourse (e.g., social media, online journalism) and AI-based annotation tools is also recommended to capture evolving linguistic trends. Collaborative research between *Corporum* contributors and corpus developers can foster methodological innovation and ensure that corpus-based insights inform curriculum design, teacher education, and language policy. Sustained investment in corpus infrastructure and cross-institutional partnerships will further consolidate Pakistan's position within the global corpus linguistics community. Future research should build upon the foundations established by *Corporum* and *PakLocCorpus* to further expand the scope and depth of corpus-based discourse inquiry in Pakistan. One promising direction is the development of multimodal and spoken corpora that capture the richness of Pakistani English in everyday, media, and educational settings. Integrating digital discourse data from social media and online platforms could also illuminate new patterns of language change, hybridity, and identity expression among younger speakers. Researchers are encouraged to apply cross-corpus comparative methodologies, using *PakLocCorpus* alongside international corpora such as ICE or GloWbE to highlight distinct linguistic features of Pakistani English. Additionally, the incorporation of AI-driven annotation tools and computational linguistics techniques would enable more refined analyses of lexical, syntactic, and pragmatic variation. Finally, collaboration among universities, publishers, and linguistic research centres should be strengthened to ensure that corpus-based approaches continue to inform curriculum design, teacher education, and national language policy.

Conclusion

This study concludes that *Corporum* has played a transformative role in popularizing corpus-based discourse research in Pakistan, offering a platform for empirical and critical engagement with language use across genres. *PakLocCorpus*, on the other hand, stands as a landmark achievement, a purpose-built national corpus that captures the diversity of Pakistani English discourse across written, spoken, and digital registers. The complementarity between *Corporum* and *PakLocCorpus* signifies a maturing stage in Pakistani corpus linguistics: one that combines rigorous theoretical analysis with robust empirical documentation. Together, these initiatives strengthen Pakistan's contribution to the study of World Englishes and reaffirm the legitimacy of Pakistani English as a dynamic and evolving linguistic variety. The review demonstrates that *Corporum* has played a central role in advancing corpus-based discourse research in Pakistan by encouraging scholars to apply CADS methodologies to socially relevant issues. At the same time, *PakLocCorpus* has emerged as a comprehensive, purpose-built corpus that complements and extends these academic

efforts. By systematically including major genres and registers of Pakistani English discourse, *PakLocCorpus* enables data-driven inquiry that was previously constrained by limited linguistic resources.

Together, *Corporum* and *PakLocCorpus* embody the progressive development of corpus linguistics in Pakistan: one offering a platform for theoretical and critical reflection, and the other providing a foundation for empirical, comparative, and pedagogical applications. Within the World Englishes paradigm, their combined contribution reaffirms Pakistani English as a vibrant and legitimate variety, one that deserves continued scholarly exploration across genres, registers, and communicative domains.

References

Altameemi, Y. M. (2024). State-of-the-Art Review of the Corpus Linguistics Field From the Beginning Until the Development of ChatGPT. *Theory and Practice in Language Studies*, *14*(2), 423-431.

Azher, M., Jahangir, H., & Mahmood, R. (2023). *Constructing Gender through Metadiscourse: A Corpus-Based Inter-Disciplinary Study of Research Dissertations of Pakistani M. Phil Graduates. Corporum: Journal of Corpus Linguistics,* 6(2), 81–101.

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus open*, *2*, 8-14.

Biber, D. (1988). *Variation across speech and writing*. Cambridge university press.

Burnard, P. (1995). Interpreting text: an alternative to some current forms of textual analysis in qualitative research. *Social Sciences in Health*, *1*(4), 236-245.

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative content analysis: A focus on trustworthiness. *SAGE open*, *4*(1).

Afifa, Ghulam Fatima, & Sajjad Rasool. (2025). Analyzing Lexical Bundles in Legal Discourse: A Corpus-Based Analysis of Court Judgments. *CORPORUM: Journal of Corpus Linguistics*, *7*(2), 145-163.

Järvelin, K., & Vakkari, P. (2022). LIS research across 50 years: content analysis of journal articles. *Journal of Documentation*, *78*(7), 65-88.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. *English in the world: Teaching and learning the language and literatures/Cambridge UP*.

Long, T., & Johnson, M. (2000). Rigour, reliability and validity in qualitative research. *Clinical effectiveness in nursing*, *4*(1), 30-37.

Majhi, S., Jal, C., & Maharana, B. (2016). Content analysis of journal articles on wiki in science direct database. *Library Philosophy and Practice*, *1331*, 1-15.

Mohammed, P. T., & Raewf, M. B. (2018). The content analysis of the research papers on language learning strategies. *Qalaai Zanist Journal*, *3*(4), 904-911.

Mumtaz, I., Zahra, F. T., & Bhatti, T. R. (2025). *Corpus-based critical discourse analysis of political themes in Pakistani English print media. Corporum: Journal of Corpus Linguistics,* 7(2), 14–29.

Mukherjee, J., & Bernaisch, T. (2020). Corpus linguistics and Asian Englishes. In K. Bolton, W. Botha, & A. Kirkpatrick (Eds.), *The Handbook of Asian Englishes* (pp. 741–762). Wiley-Blackwell.

Patton, M. Q. (2002). Qualitative, research & evaluation methods. Thousand Oaks, California: Sage publications Inc.

Rapport, F. (2010). Summative analysis: A qualitative method for social science and health research. *International Journal of Qualitative Methods*, *9*(3), 270-290.

Sinclair, J. & Karter, R. (Eds.). (2004). *Trust the text: Language, corpus, and discourse*. (1st Ed.). Routledge. https://doi.org/10.4324/9780203594070

Zahra, F., Zahra, T., & Abbas, A. (2022, June 15). PakLocCorp. Retrieved July 8, 2022, from https://pakloccorp.com/

Zahra, F., Zahra, T., & Abbas, A. (2023a). Neoliberal Co-option of English: A Study of Urduized Terms in Pakistani Magazine Articles through Corpus Approaches. *Rivista Italiana di Paleontologia e Stratigrafia. RIFL, 17*(2), 174-188

Zahra, F., Zahra, T., & Abbas, A. (2023 b). Corpus assisted discourse analysis of Urduized words and selected collocation patterns: a case study of trespassing by Uzma Aslam Khan. *Research Journal of Language and Literary Studies, 3(2), 125–144.*

Zahra, F., Zahra, T., & Abbas, A. (2024). Globalization of Narratives for Clothing and Appearance: Corpus-Assisted Analysis of Intercultural Communication. Migration Letters, 21(13), 532–551.

Zahra, F. T., Zahra, T., & Abbas, A. (in press/accepted, expected 2026). Developing an online specialized corpus of Pakistani English. Corpora, 21(3). (Forthcoming)

Appendix A
Articles Published in Year 2020

Sr no	Title	Writer 1st	Medium	Language	Genre /sub-genre
1	Metadiscursive Study of Kashmir Issue through Attitude Markers in Pakistani English Newspapers	Female	written	English	Academic
2	A Corpus-based Analysis of Conceptual Metaphors in "The Forty Rules of Love" by Elif Shafak	Male	written	English	Academic
3	A Cross-linguistic Study of Metadiscourse in English and Urdu Newspaper Editorials	Male	written	English	Media
4	A Corpus-based Study of James Joyce's A Portrait of the Artist as a Young Man	Male	written	English	Literary
5	Exploring Authorial Voice in Verb-Control Citations: An Inter-disciplinary Study of Ph.D. Theses	Male	written	English	Academic

Articles Published in Year 2021 issue 1

Sr no	Title	Writer 1st	Medium	Language	Genre /sub-genre

1	Linguistic Variation Among Pakistani Internet Blogs: A Multidimensional Analysis	Female	written	English	Others/social media
2	Urdu Conjunct Predicates (N+V) Inventory from Urdu Universal Dependency Corpus	Female	written	English	Academic
3	Foregrounding through Lexical Deviation: A Corpus-Based Analysis of Yousafi's Aab-e-Gum	Male	written	English	Literary Discourse
4	Neuro-Linguistic Programming: A Corpus-Based Critical Analysis of Motivational Speeches	Male	speech	English	Academic
5	A Corpus-Based Study of Austen's Novel Pride & Prejudice using Cirrus Tool	Male	written	English	Literary Discourse

Articles Published in Year 2022 issue 1

Sr no	Title	Writer 1st	Medium	Language	Genre /sub-genre
1	Semantic Frames of the Urdu Conjunct Verb lagnā: A Corpus-based Studys	Female	written	Urdu	Academic
2	The Behaviour of Public and Press towards Burqa: A Corpus-Assisted Discourse Analysis	Male	written	English	Media Discourse
3	Corpus-Based Analysis of Career Services Language in the UK and Pakistani University Websites	Female	written	English	Media discourse
4	A Comparative Corpus-based Analysis of Collocational Patterns in Self and Other-translators	Female	written	English	Literary Discourse

5	Use of Reporting Verbs in Research Articles by Pakistani Authors: A Corpus-Based Study	Female	written	English	Academic discourse
6	A Corpus Based Study of Hedges and Boosters in the Speeches of Benazir Bhutto	Female	spoken	English	Media Discourse
7	Citation Practices: A Corpus-Based Analysis of Literature Reviews Sections of Ph.D.Theses in Biological Sciences	Male	written	English	Academic

Articles Published in Year 2023 issue 1

Sr no	Title	Writer 1st	Medium	Language	Genre /sub-genre
1	Delicate Patterns of Colligations: A Corpus-based Study of Prepositions	female	written	English	Native corpus/academic Student essays
2	A Corpus-based Study of Newspaper Articles on Lockdown Issue in Indian Occupied Kashmir	male	written	English	Media discourse
3	A Corpus-Stylistic Analysis of Color Symbolism in Sidhwa's Water	female	written	English	Literary Discourse
4	The Representation of Afghan Refugees in Pakistani Urdu and English Blogs: A Corpus-Assisted Discourse Analysis	Male	written	English	Media discourse
5	Tracing intersectionality, individual and familial struggle in Morrison's novels	male	written	English	Literary Discourse

Articles Published in Year 2024 issue 1

Sr no	Title	Writer 1st	Medium	Language	Genre /sub-genre
1	Implicit Messages in Pakistani ESL Course Books: A Corpus-Based Study of Gender Representation	Male	written	English	Academic
2	A Corpus-Based Analysis of the Use of Singular Nouns by Pakistani Students in Higher Education	Male	written	English	Academic
3	A Computational Approach to Understanding Agglutinative Structures in Urdu	M	spoken	Urdu /English	Translated corpus
	Eurocentric Beauty Standards: A Corpus-Assisted Discourse Analysis of Pakistani Skincare Blogs	F	written	English	Media/health care
5	Exploring Passive Structures in Political and Religious Text: A Corpus Based Study of Pakistani Writing	M	written	English	Policy documents/religious essays
6	Adjectives and Modal Verbs Used in Horoscopes in Dawn Newspapers and their Impact on Readers: A Corpus-Assisted Study	M	Written	English	Media
7	A Corpus-Based Analysis of Interactive Markers in English Editorials of The News and Dawn News	M	written	English	Media
8	A Corpus Based Analysis of Sympathy in Language: Constructed Victimhood in Fiction	F	written	English	Literary Discourse