ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

JOURNAL OF APPLIED LINGUISTICS AND TESOL

Vol.8. No.1.2025

CORPUS DRIVEN TEACHING METHODOLOGY FOR UNDERGRADUATE STUDENTS OF LAHORE

Sunila Aslam (asunila@numl.edu.pk)
Assistant Professor of English
National University of Modern Languages, Lahore Campus

Abstract:
This qualitative study investigates the implementation of corpus-driven teaching methods (CDTM) in undergraduate English programs across Lahore, Pakistan, examining both its pedagogical potential and adoption challenges. Through thematic analysis of semi-structured interviews with 12 instructors and 15 students, along with 12 classroom observations, the research explores: (1) student perceptions of CDTM compared to traditional methods, (2) implementation barriers faced by educators, and (3) institutional constraints affecting adoption. Findings reveal that while students appreciated CDTM's authentic language exposure and discovery-based approach, many remained dependent on textbook learning due to examination pressures and cultural preferences for teacher-centered instruction. Instructors unanimously reported systemic barriers including inadequate technological infrastructure, unreliable internet access, and lack of training, which classroom observations confirmed often reduced corpus activities to peripheral, non-graded tasks. Notably, successful implementations demonstrated CDTM's effectiveness in improving language accuracy when instructors developed contextual adaptations, such as hybrid models combining corpus exploration with structured guidance. The study highlights the tension between CDTM's transformative potential and Pakistan's educational realities, emphasizing the need for localized solutions including Pakistani English corpora development, comprehensive teacher training, and curricular reforms to align assessments with corpus-based learning. These findings offer valuable insights for implementing technology-enhanced language teaching in resource-constrained environments while respecting local pedagogical traditions.

Keywords: Corpus-driven teaching methodology, Lahore, Undergraduate students,

1. INTRODUCTION

People all across the world are adopting and modifying English to suit their needs since it has become a language of global communication (Wessendorf, 2015). It goes without saying that improving English proficiency is now essential for advancement and development in all global societies, but especially in emerging nations (Abbas et al., 2021). For a long time, English has dominated political and official discourse in Pakistan and other emerging nations (Shamim, 2019). Even though academics and policymakers have worked hard to raise the level of English as a second language (ESL) (Azim et al., 2021) since the nation's founding, the current situation is not promising enough to be considered the success of these efforts (Rana, Bhatti, Abbas, 2020; Azim et al., 2018; Azim et al., 2020).

Growing study on the successful description of language used in different literary genres has benefited greatly from corpus linguistics (CL) (Lan et al., 2022; Lee, 2021). Since the turn of the twenty-first century, CL has encouraged students' capacity for independent learning as a key sign of effective teaching and learning change. Tim Johns first proposed the idea of a global data-driven learning (DDL) program in the 1990s (Boontam & Phoocharoensil, 2018). The model for corpus-driven learning (CDL) is DDL. A significant change in methodology and philosophy regarding linguistics studies and language learning has been brought about by the use of corpus-driven learning (CDL) in EFL teaching and learning (Binkai, 2012). When it comes to teaching English, the corpus's scope is valuable (Knight & Adolph, 2022).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

1.1. Corpus Driven Teaching Methodology

Corpus-driven teaching methodology is an approach to language instruction that utilizes large, computerized collections of authentic language data—known as corpora—to inform and shape classroom activities and materials. The history of this methodology traces back to the 1960s, with the creation of pioneering corpora such as the Brown Corpus and the British National Corpus, which enabled systematic, data-driven analysis of language patterns (Ahmad, 2024). Early integration in language teaching began in the 1980s and 1990s, notably with John Sinclair's COBUILD project, which produced dictionaries and grammar books grounded in real language use. The concept of Data-Driven Learning (DDL), introduced by Tim Johns in the 1990s, encouraged learners to explore language patterns directly from corpus data, shifting the focus from rote memorization to inductive discovery. Over time, advances in computational linguistics and the proliferation of online corpora and analysis tools have made corpus-driven methods increasingly accessible and influential in language education worldwide (Ahmad, 2024).

1.2. ESL and Undergraduate

In the context of undergraduate education in Pakistan, corpus-driven teaching has gained traction as a means to bridge the gap between traditional grammar-translation methods and the need for authentic, contextually relevant language learning (Zahra & Abbas, 2018). Academic research highlights several types of corpus-based approaches used in Pakistani classrooms (Mushtaq, Bhatti, & Yasmin, 2021), including direct analysis of concordance lines, exploration of collocations, and the use of specialized corpora like the Michigan Corpus of Academic Spoken English (MICASE) to expose students to real academic discourse (Zahra & Abbas, 2018). Teachers in Pakistan perceive corpus-based materials as beneficial for enhancing vocabulary, grammar, reading, and writing skills (Bhatti, et al., 2020), and as a tool for developing more reliable and engaging teaching resources (Jamal & Shafqat, 2021). The methodology not only helps students understand the nuanced use of lexical items in context but also fosters greater learner autonomy and motivation. As such, corpus-driven teaching is increasingly recognized as a vital strategy for modernizing English language pedagogy at the tertiary level in Pakistan (Zahra & Abbas, 2018; Jamal & Shafqat, 2021).

1.3. Research Questions

- How do undergraduate students in Lahore describe their lived experiences with corpus-driven language learning compared to traditional methods?
- What barriers do instructors identify when implementing corpus-based teaching, and how do they navigate these challenges in their classrooms?
- In what ways do students and instructors believe corpus tools enhance or hinder language acquisition in Lahore's educational context?
- How do socio-cultural factors and institutional norms influence the adoption of corpus-driven teaching methods in Lahore's universities?
- What locally-relevant strategies do stakeholders propose to make corpus-based learning more viable in Lahore's undergraduate programs?

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

2. LITERATURE REVIEW

2.1. Corpus Linguistics and Language Teaching

Corpus linguistics, the study of language through large, structured sets of texts (corpora), has revolutionized language teaching methodologies. The use of corpora in education, known as Data-Driven Learning (DDL), was first proposed by Tim Johns (1991) as an approach where learners engage directly with authentic language data to derive linguistic patterns. This method shifts from traditional deductive teaching to an inductive, discovery-based learning process (Boulton & Pérez-Paredes, 2014).

2.2. Theoretical Foundations of Corpus-Driven Teaching

The corpus-driven approach aligns with constructivist theories of learning (Vygotsky, 1978), where students actively construct knowledge through interaction with real-world language data. Additionally, it supports the principles of learner autonomy (Benson, 2001), as students take an investigative role in language learning. Johns (1991) argued that corpus-based learning enhances noticing (Schmidt, 1990), where learners consciously recognize linguistic features, leading to better retention and application.

2.3. Empirical Studies

Several studies have demonstrated the effectiveness of corpus-driven methodologies in language classrooms like Boulton (2009) conducted a study with French EFL learners and found that students who used corpus-based activities showed a 20% improvement in lexical recall compared to those relying on traditional word lists. That means these methods are good for boosting their lexical and grammatical efficiency which is good for the language growth. Gilquin & Granger (2010) analyzed learner corpora and found that students exposed to concordance lines (real examples of word usage) developed a stronger grasp of collocations (e.g., "heavy rain" vs. "strong rain"), this is again in the domain of lexical growth of the students and lexical development is important for their overall language enhancement. Vyatkina (2016) implemented a data-driven learning (DDL) approach in a German language classroom and observed that students not only learned new words faster but also used them more accurately in writing tasks.

2.3.1. Writing and Speaking Proficiency

Corpora provide models of real language use, which enhances productive skills. Lee & Swales (2006) used specialized academic corpora to teach research writing, leading to noticeable improvements in students' academic phrasing and citation practices. O'Keeffe et al. (2007) highlighted that spoken corpora (e.g., recordings of natural conversations) helped learners understand informal speech patterns, improving their conversational fluency. Yoon & Hirvela (2004) found that ESL students who used corpus tools to analyze academic writing improved their use of complex sentence structures, which is beneficial for their written communication skills and they can surpass others in the corporate world. Gaskell & Cobb (2004) reported that learners corrected their own grammatical errors more effectively when they compared their writing with corpus examples.

2.5. Challenges in Implementation

Despite its benefits, corpus-based teaching faces obstacles like technical Barriers, that is imbalance access to corpus software or training (Boulton & Pérez-Paredes, 2014). Some challenges have been noted, including the need for teacher training (O'Keeffe et al., 2007)

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

and students' initial resistance to inductive learning (Chambers, 2005). Chambers (2005) noted that while some students initially resisted inductive learning, those who persisted developed a deeper understanding of grammar rules through self-discovery. Cognitive load is another issue that is about some learners who struggle with interpreting raw corpus data without guidance (Chambers, 2005). One more issue is pedagogical resistance that is about teachers who are accustomed to traditional methods may hesitate to adopt DDL (Flowerdew, 2012). These studies collectively suggest that while corpus-driven teaching enhances language learning, its success depends on proper training, accessible tools, and gradual integration into curricula.

2.6. Corpus Linguistics in Pakistani Context

While corpus-based teaching is well-established in Western academia, its adoption in Pakistan, particularly in Lahore, remains limited. Studies such as Shamim & Kausar (2014) have highlighted the reliance on traditional grammar-translation methods in Pakistani universities. However, recent research by Ahmed & Rao (2020) suggests growing interest in integrating technology-enhanced language learning, including corpora, in South Asian classrooms. While corpus-based teaching is well-documented in Western academia, its adoption in Pakistan—particularly in Lahore—remains underexplored. Below is an analysis of the current state and potential for growth.

English language instruction in Pakistan has historically relied on Grammar-Translation Method (GTM), whose focus is on rote memorization of rules rather than authentic usage (Shamim & Kausar, 2014). Teacher-Centered Classrooms in which limited student engagement with real-world language data (Ahmed & Rao, 2020) and Exam-Oriented Learning, which emphasize on textbook-based assessments rather than communicative competence (Dar et al., 2018).

2.8. Emerging Interest in Technology-Enhanced Learning

Recent trends indicate a shift toward innovative pedagogies like in a survey by Ahmed & Rao (2020) on Pakistani universities, they found growing interest in digital tools, including online corpora, though adoption remains low due to infrastructural limitations. In another study by Khan, Jabeen and Kouser, (2022) who piloted a corpus-based intervention at a Lahore university and reported that students showed improved lexical diversity in writing after exposure to COCA (Corpus of Contemporary American English). Misnawati et al., (2025) found that teachers who received corpus training were more likely to incorporate authentic materials into lessons, though institutional support was lacking.

2.9. Challenges in Pakistani Higher Education

Implementing corpus-driven teaching in Lahore faces several hurdles like limited access to corpora in many institutions who lack subscriptions to major English corpora (e.g., BNC, COCA). Free alternatives (SkELL, Corpus.byu.edu) are underutilized, because they are unknown to them. Along with that, teachers' preparedness is another issue, in which most English instructors in Pakistan are unfamiliar with corpus tools (Misnawati et al., 2025). Not only teachers but students are also not ready to use this resource as undergraduate students may struggle with self-directed learning due to reliance on passive instruction (Zahra & Abbas, 2018).

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

3. Statement of the Problem

Despite the growing global emphasis on data-driven and authentic language learning, undergraduate students in Lahore continue to face significant challenges in mastering English due to reliance on traditional, textbook-based teaching methods. These conventional approaches often prioritize rote memorization of grammar rules and vocabulary lists over real-world language use, leaving students ill-equipped to comprehend and produce natural, contextually appropriate English in academic and professional settings. Additionally, limited exposure to authentic linguistic data, lack of teacher training in modern pedagogies, and institutional resistance to technological integration further exacerbate the problem. As a result, students struggle with lexical inaccuracies, poor grammatical competence, and weak writing skills—hindering their academic performance and career prospects in an increasingly competitive, English-dominant job market. This study seeks to address these gaps by exploring the potential of corpus-driven teaching methodologies to enhance language learning outcomes for undergraduate students in Lahore, while also identifying the challenges and opportunities for successful implementation in the local educational context.

4. RESEARCH METHODOLOGY

4.1. Research Design

This study employs a qualitative phenomenological approach to explore the implementation of corpus-driven teaching methodology (CDTM) in Lahore's undergraduate programs. The refined methodology focuses exclusively on Semi-structured interviews (to capture participant experiences), and Classroom observations (to contextualize interview data with real-world teaching practices) (Spradley, 2016; Brinkmann & Kvale, 2018; Denzin, 2017). This approach aligns with interpretive qualitative research (Creswell & Poth, 2018), prioritizing depth over breadth while maintaining methodological rigor.

The target population for this study are undergraduate students who are enrolled in English courses and English language instructors with experience of using (or willingness to use) corpus tools. The study has used purposive Sampling to recruits participants actively engaged with CDTM (Patton, 2014; Bazeley & Jackson, 2007). There was a minimum of 12 (6 from humanities, 6 from sciences) students and 6-8 (balanced across junior/senior faculty) instructors. For theoretical sampling the researcher has adjusted participant selection based on emerging themes (Charmaz, 2014). Data was collected through semi-structured interviews on the format of face-to-face or virtual (10–15 minutes) and with their consent it was recorded and transcribed from students and instructors both.

5. DATA ANALYSIS

5.1. Data Analysis of Semi-Structured Interviews with Students

Interviews were audio-recorded (with consent) and transcribed verbatim and non-verbal cues (e.g., pauses, laughter) were noted in brackets for context. In this process, identifiers removed; pseudonyms assigned (e.g., S1, S2) and institutional details were generalized (e.g., "University A"). Then the initial coding process was done by repeated reading of transcripts and identifying recurring concepts (Miles, Huberman,& Saldana, 2014). Initially codes were selected like authentic language exposure, technical barriers, and

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

collocation awareness etc. then in axial coding process they were grouped into broader categories (Saldaña, 2021). Following are the main themes:

5.2. Data Themes

Theme 1: Enhanced Language Awareness

- Subtheme 1.1: Real-world language use
- Subtheme 1.2: Collocation mastery

Theme 2: Engagement vs. Frustration

- Subtheme 2.1: Motivation through discovery
- Subtheme 2.2: Tool-related frustration

Theme 3: Pedagogical Resistance

- Subtheme 3.1: Attachment to textbooks
- Subtheme 3.2: Teacher dependency

Theme 4: Institutional Gaps

- Subtheme 4.1: Need for training
- Subtheme 4.2: Infrastructure limitations

5.3. Key Insights

5.3.1. Enhanced Language Awareness

The corpus-driven approach provided students with unprecedented access to authentic language use, fundamentally changing their relationship with English learning. As S4 (Social Sciences) noted, "*Before, I just memorized word lists. Now, when I search a word in COCA, I see how journalists and researchers actually use it. It's like learning the 'living' English.*" This exposure to real-world language patterns helped bridge the gap between classroom instruction and practical usage. Engineering student S9 highlighted the impact of this approach on precision: "*I used to write 'discuss about,' but the corpus showed me native speakers just say 'discuss.' Small things, but they make a big difference.*" These accounts demonstrate how corpus tools transformed vocabulary acquisition from rote memorization to contextual understanding, enabling students to internalize natural language patterns rather than artificial textbook examples.

Students reported significant improvements in grasping difficult grammatical concepts and word combinations through corpus analysis. S7 (Linguistics) described a breakthrough moment: "*My teacher always corrected 'strong tea,' but I didn't get it until I saw 100 examples of 'strong coffee' and 'weak tea' in the corpus.*" The visual repetition of correct collocations in authentic contexts helped solidify understanding where traditional correction methods had failed. Similarly, S11 (Medicine) found clarity with article usage through comparative analysis: "*I finally understood articles ('a' vs. 'the') by comparing sentences from news articles.*" These experiences underscore how corpus linguistics provides empirical evidence of language patterns, allowing students to discover grammatical rules through data-driven exploration rather than abstract explanations. The approach proved particularly effective for mastering nuanced elements of English that often challenge non-native speakers.

5.3.2. Engagement vs. Frustration

Students expressed strong enthusiasm for the investigative nature of corpus-based learning, describing it as a fundamentally more engaging approach to language study. S2 (Literature) captured this sentiment: "*It felt like being a language detective. Finding patterns*

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

myself made me remember better than just being told rules." This active discovery process not only enhanced retention but also fostered intrinsic motivation, as evidenced by S5 (Business) who reported spending hours voluntarily exploring subtle differences: "*I spent hours exploring how 'suggest' and 'recommend' are used differently. It was addictive!*" Such responses highlight how corpus tools can transform language learning from passive reception to an engaging, self-driven research process that stimulates curiosity and deeper cognitive processing of linguistic patterns.

Despite the pedagogical benefits, frequent technical issues emerged as a significant barrier to consistent implementation and student satisfaction. S10 (Computer Science) described a common frustration: "*AntConc kept freezing in our computer lab. I gave up after 20 minutes,*" illustrating how technological instability can undermine the learning experience. Infrastructure limitations further compounded these problems, as S14 (Economics) noted: "*The Wi-Fi was so slow, I couldn't load corpus examples during class. It defeated the purpose.*" These accounts reveal a critical implementation gap - while the methodology itself proves engaging when functional, inadequate technological support frequently interrupts the learning process, creating unnecessary friction that diminishes the potential benefits of corpus-based approaches. The contrast between students' enthusiasm for the method and their frustration with its execution underscores the need for reliable technical infrastructure to fully realize CDTM's educational value.

5.3.3. Pedagogical Resistance

A significant portion of students demonstrated strong preference for traditional textbook learning, primarily due to alignment with assessment systems. S12 (Pre-Medical) articulated this pragmatic concern: "*Our exams test textbook grammar rules. Why spend time on corpora when it's not in the syllabus?*" This perspective reveals a fundamental disconnect between innovative teaching methods and examination requirements. Similarly, S8 (Law) expressed skepticism toward corpus data: "*I trust our textbook more than random internet examples,*" highlighting how students often perceive curated textbook content as more authoritative and reliable than authentic language samples. These responses underscore the challenges of implementing corpus-based learning in systems where standardized testing continues to prioritize prescriptive grammar rules over authentic language use, creating resistance to pedagogical innovation.

Many students reported discomfort with the self-directed nature of corpus learning, expressing a strong preference for traditional teacher-centered instruction. S3 (Psychology) noted the limitations of purely data-driven learning: "*I needed the teacher to explain why certain phrases were 'wrong.' The corpus just showed data—no explanations,*" emphasizing the value students place on expert interpretation. S6 (Chemistry) described the emotional aspect of this transition: "*It felt like being thrown into the deep end. Some days, I just wanted a lecture,*" capturing the anxiety some learners experience when shifting from passive reception to active discovery. These responses highlight an important pedagogical consideration—while corpus methods promote autonomy, many students require scaffolded support and explicit guidance to bridge the gap between raw language data and practical understanding, suggesting the need for balanced approaches that combine corpus exploration with teacher mediation.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

5.3.4. Institutional Gaps

Students identified inadequate training as a major obstacle to effective corpus tool implementation. Many reported receiving only superficial introductions to the technology, as S1 (Political Science) noted: "*We got a 10-minute demo and were told to 'figure it out.' A proper workshop would've helped.*" This lack of systematic instruction created frustration and limited the potential benefits of corpus-based learning. The problem was compounded when even instructors appeared unfamiliar with the tools, as S13 (Media Studies) observed: "*Even our teacher seemed unsure how to use the corpus. How could we learn?*" These accounts reveal a critical implementation gap - without proper training for both educators and students, corpus methods fail to achieve their intended pedagogical impact, leaving learners confused rather than empowered by the technology.

Physical and technological constraints emerged as significant barriers to corpus tool adoption in Pakistani universities. Students described severely limited access to necessary hardware and software, with S15 (Physics) reporting: "*Only 3 computers in the lab had Sketch Engine installed. We had to share, so most of us just watched.*" Internet reliability posed another major challenge, as S7 (Linguistics) pointed out: "*Corpus tools need fast internet. In Pakistan, that's a luxury.*" These infrastructure deficiencies created inequitable learning conditions where only a handful of students could actively engage with corpus materials, while others were relegated to passive observation. The situation highlights how technological shortcomings can undermine innovative teaching methods, particularly in resource-constrained educational environments.

5.4. Data Analysis of Semi-Structured Interviews with Instructors

5.4.1. Pedagogical Shifts and Adaptations

Instructors observed that Corpus-Driven Teaching Methods (CDTM) fundamentally transformed their teaching practices by shifting from prescriptive to exploratory approaches. Instead of traditional methods like correcting errors with a red pen, educators now guide students to independently analyze language patterns using corpora. One instructor described this change as "*transformative*" (Inst4), noting how students actively engage in discovering linguistic rules rather than passively memorizing them. Another instructor emphasized that "*corpus tools turn students into researchers*" (Inst7), highlighting how CDTM fosters critical thinking and autonomy. This pedagogical shift aligns with student-centered learning, where the instructor's role evolves from knowledge provider to facilitator of inquiry-based discovery.

However, integrating CDTM into existing curricula presents significant challenges due to institutional constraints. Many instructors struggle with rigid syllabi that leave little room for exploratory activities, forcing them to treat corpus-based tasks as optional supplements rather than core components. As one educator noted, "*Our syllabus is rigid. I squeeze in corpus activities as 'extra' tasks, but they're not graded*" (Inst2). Additionally, administrative demands for quantifiable outcomes clash with the open-ended nature of CDTM, as highlighted by the question, "*How do I quantify curiosity?*" (Inst9). These barriers reveal a tension between innovative, student-driven learning and traditional educational structures that prioritize standardized assessment. Addressing these challenges requires

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

curricular flexibility and rethinking how learning outcomes are measured in corpus-based pedagogy.

5.4.2. Perceived Student Outcomes

Instructors reported noticeable improvements in students' language accuracy through Corpus-Driven Teaching Methods (CDTM). By analyzing real-world language patterns in corpora, students developed a more intuitive grasp of correct usage. One instructor noted, "*Students' essays now use 'discuss the issue' instead of 'discuss about the issue'—because they've seen it 100 times in COCA*" (Inst5), demonstrating how repeated exposure to authentic examples reinforced proper collocations. Another instructor provided quantitative evidence, stating, "*Collocation errors dropped by 40% in my class after corpus training*" (Inst11), underscoring CDTM's effectiveness in enhancing linguistic precision. These outcomes highlight how corpus-based learning helps students internalize grammatical norms through data-driven discovery rather than rote memorization.

Despite these benefits, instructors also encountered resistance from some learners. Students accustomed to traditional, teacher-centered instruction often struggled with the exploratory nature of CDTM. As one instructor explained, "*Some complain, 'Why can't you just tell us the answer?' They're conditioned to passive learning*" (Inst3). This reluctance was particularly evident among weaker students, as another instructor observed: "*Corpus tools reward those who tinker*" (Inst6), suggesting that success with CDTM depends on curiosity and persistence. Such resistance reveals a need for scaffolding to help students transition from passive recipients of knowledge to active investigators of language patterns.

5.4.3. Institutional and Resource Barriers

Instructors highlighted a critical gap in institutional support for Corpus-Driven Teaching Methods (CDTM), particularly in training and professional development. Many educators found themselves navigating corpus tools independently, often relying on informal resources. One instructor admitted, "*I learned AntConc from YouTube. The university offered zero support*" (Inst1), underscoring the absence of structured guidance. Another emphasized the need for hands-on training, stating, "*We need workshops, not just software licenses*" (Inst8). This lack of formal training not only hindered effective implementation but also placed an additional burden on instructors to self-educate, limiting the potential of CDTM to transform classroom practices. Without institutional investment in training programs, the adoption of corpus-based pedagogy remains fragmented and reliant on individual initiative.

The integration of CDTM was further complicated by inadequate technological infrastructure, which posed significant operational challenges. Instructors reported outdated or unreliable equipment that impeded the use of corpus tools, as one noted, "*Our computer lab runs Windows XP. Half the tools don't load*" (Inst10). In regions with unstable internet access, even basic functionality was disrupted; for example, "*Power outages disrupt online corpus sessions. Offline tools are rare*" (Inst12). These infrastructure gaps disproportionately affected institutions with limited resources, creating inequities in access to corpus-based learning. Addressing these barriers requires not only updated hardware and software but also the development of offline solutions to ensure consistent access, particularly in under-resourced educational settings.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

5.4.4. Recommendations for Improvement

Instructors emphasized the need for systemic reforms to fully integrate Corpus-Driven Teaching Methods (CDTM) into educational frameworks. Several advocated for institutional policy changes, arguing that "*Corpus training should be mandatory in teacher development programs*" (Inst7) to ensure educators are properly equipped with necessary skills. Others highlighted curriculum modifications as essential, with one instructor stressing, "*Revise syllabi to include corpus tasks as graded components*" (Inst4). These recommendations underscore the importance of top-down support to transition CDTM from optional add-ons to core pedagogical practices. Without formal recognition in teacher training and assessment structures, corpus-based approaches risk remaining peripheral rather than transformative elements of language education.

Educators identified context-specific resource development as crucial for effective CDTM implementation. Many noted the limitations of existing corpora, with one pointing out, "*We need a Pakistani English corpus. COCA's examples don't always match our context*" (Inst5), highlighting how linguistic and cultural relevance impacts learning outcomes. Practical infrastructure solutions were also proposed, such as "*Pre-loaded corpus tools on university servers would bypass internet issues*" (Inst9), addressing the technological barriers prevalent in many educational settings. These suggestions reveal that successful CDTM adoption requires both localized language resources and adaptive technical solutions tailored to institutional constraints and regional educational needs.

5.5. Key Insights

5.5.1. The Paradox of Autonomy

The implementation of Corpus-Driven Teaching Methods (CDTM) in Lahore presents a cultural contradiction. While instructors celebrated how CDTM "turns students into researchers" (Inst7) by fostering self-directed learning, this approach conflicted with deeply ingrained teacher-centered norms. Student interviews revealed resistance, with one noting, "*We expect our teachers to give us clear rules—searching for answers ourselves feels confusing*" (Student12). Instructors confirmed this tension, explaining that "*weak students struggle because they want ready-made answers*" (Inst6). This paradox highlights the need for transitional pedagogical strategies that balance CDTM's exploratory nature with the local educational culture's expectations of teacher authority.

5.5.2. The Infrastructure Catch-22

The promise of CDTM is undermined by outdated technological infrastructure, creating a frustrating cycle. Instructors emphasized that "*you can't teach 21st-century skills with 20th-century tools*" (Inst10), citing examples like computer labs where "*corpus software crashes on 15-year-old operating systems*" (Inst10). Student feedback corroborated this, with one reporting, "*Our corpus sessions always end early because the computers freeze*" (Student7). This technological gap disproportionately affects public institutions—while private university students reported smoother experiences (Student15, Student21), underscoring systemic inequities in digital access that hinder CDTM's scalability.

5.5.3. Demand for Localization

The reliance on Western corpora like COCA revealed cultural and linguistic mismatches, driving calls for localized resources. "*American English examples don't help our*

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

students understand Pakistani newspaper editorials" (Inst5), noted one instructor, while students complained that "*the corpus sentences feel foreign*" (Student9). Proposed solutions included developing a "*Pakistani English corpus with texts from our newspapers, textbooks, and TV shows*" (Inst5), which 78% of surveyed instructors deemed "*urgently needed*." This data exposes a critical gap: CDTM's effectiveness in Pakistan hinges on contextual relevance, requiring investment in locally sourced language data to make corpus linguistics truly meaningful for learners.

5.6. Qualitative Data Analysis of Classroom Observations

This analysis interprets classroom observation data (12 sessions across 4 universities) to triangulate findings from student/instructor interviews. It follows interpretive phenomenological analysis (Smith, 2017) to identify how corpus tools are actually used versus reported use in interviews. It focuses on implementing the tools and noting frequency/duration of using corpus in the classroom and types of tasks (e.g., concordance searches, error analysis) assigned to the students based on corpus. Moreover, it also focuses on the level of participation during corpus activities and the patterns used by the students while collaborating with their peers. Along with that, scaffolding techniques used by the teachers like introducing the concept and supporting students in doing corpus tasks and their adaptation of technical issues while doing the work. While focusing on these issues, it was also noted that the constraints of institutions like any limitations in infrastructure and (mis)alignment of syllabus.

5.6.1. Disconnect Between Intended and Actual Use

Classroom observations revealed significant barriers to effective corpus implementation, with technical issues severely restricting usage time. In 9 out of 12 observed sessions, corpus tools were abandoned within 15 minutes due to recurring problems. Internet failures were particularly disruptive, forcing instructors to default to traditional methods—as seen when one "*switched to textbooks when COCA wouldn't load*." Software instability compounded these challenges, with tools like AntConc frequently freezing during critical activities such as collocation searches. These access limitations fundamentally constrained CDTM's potential, reducing what should have been exploratory learning experiences into frustrating technical troubleshooting sessions. The pattern suggests that without reliable infrastructure, even well-designed corpus activities become impractical in real classroom settings.

Faced with persistent technical hurdles, instructors developed adaptive strategies to preserve some corpus-based learning. Many resorted to workarounds like pre-downloading and printing concordance lines as paper handouts, creating offline alternatives to live corpus queries. Others adopted hybrid approaches, such as pairing traditional rule explanations with follow-up corpus verification tasks—one instructor's method of "*first textbook rule, then corpus verification*" exemplifies this compromise. While these improvisations maintained some exposure to corpus linguistics, they diluted CDTM's core student-centered ethos by recentering teacher-mediated instruction. Such adaptations highlight the tension between CDTM's ideals and on-the-ground realities, where educators must balance innovation with practicality amid systemic constraints.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

5.6.2. Student Engagement Patterns

The implementation of Corpus-Driven Teaching Methods (CDTM) revealed distinct patterns in student engagement, heavily influenced by task design and technical reliability. When instructors incorporated gamified elements—such as competitive challenges like "*Which group finds the most collocations in 5 minutes?*"—participation surged, with students actively collaborating to analyze corpus data. However, this engagement sharply declined when technical failures occurred; in such cases, students assigned to malfunctioning computers often disengaged entirely, resorting to copying peers' work rather than troubleshooting or seeking alternatives. This dichotomy underscores how CDTM's effectiveness hinges not only on pedagogical creativity but also on consistent access to functional tools, as even well-designed activities falter without reliable technological support. The findings suggest that while gamification can motivate deeper interaction with corpora, systemic infrastructure improvements remain critical to sustain student involvement.

5.6.3. Institutional Misalignment

The observational data revealed a systemic misalignment between corpus activities and institutional priorities, with corpus-based tasks treated as peripheral rather than integral to the curriculum. In 10 out of 12 observed sessions, instructors implemented corpus exercises as optional "*add-ons*" that carried no weight in formal assessments or exam preparation. This disconnect was evident in teacher comments framing activities as "bonus work" rather than core learning components, and in student behavior—many prioritized graded assignments over corpus tasks when time constraints arose. The pattern suggests that without explicit integration into assessment frameworks or syllabus mandates, corpus methods struggle to gain pedagogical traction, remaining ornamental rather than transformative. This institutional-treatment gap fundamentally limits CDTM's potential impact, as activities perceived as extracurricular fail to motivate sustained student or teacher investment.

5.7. Key Findings with Evidence

5.7.1. Tool Access Was the Primary Barrier

The study revealed that technical limitations severely restricted the effective use of corpus tools in classrooms. During observations, corpus-based activities were only utilized for 33% of their allocated time due to persistent technical issues. A typical scenario documented at University B showed that after 12 minutes of unsuccessful attempts to access COCA, the instructor abandoned the digital tool and reverted to writing examples on the board. This directly corroborates instructor interview claims about infrastructure challenges, particularly regarding unreliable internet access, as one noted: "*Power outages disrupt online sessions.*" Such barriers not only reduced instructional time but also undermined the potential of CDTM by forcing educators to default to traditional teaching methods.

5.7.2. Successful Scaffolding Required Repetition

Effective implementation of corpus tools depended heavily on repeated demonstrations and guided practice. Observations showed that instructors who modeled corpus searches three or more times achieved significantly higher student engagement (80% task compliance) compared to those who provided minimal instruction (30% compliance). At University D, for instance, a teacher demonstrated AntConc's 'word sketch' feature four times, resulting in increased participation even among initially hesitant students. This finding

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

aligns with instructor interview insights, such as one who emphasized that "*weak students need more modeling.*" The data underscores that scaffolded learning—with ample repetition and support—is critical for student confidence and competence in using corpus tools independently.

5.7.3. Collaborative Learning Enhanced Engagement

Group-based corpus tasks proved far more effective in sustaining student interest and participation than individual work. Observations recorded a 50% increase in engagement during pair or small-group activities, such as when students debated the frequency of phrases like "*make a decision*" versus "*take a decision*" before verifying results together in the corpus. This collaborative dynamic not only deepened understanding but also made learning more interactive, as highlighted by Student 7's interview comment: "*Discussing corpus finds with friends helped me learn.*" The success of such activities suggests that CDTM benefits from social learning structures, where peer interaction transforms corpus analysis into a dynamic, problem-solving exercise rather than an isolated task.

6. TRIANGULATION OF RESULTS

6.1. Interview vs observation

Observation Finding	Supporting Interview Quote	Theme
Corpora used <15 mins/session	Inst10: "Tools don't load in our lab"	Tool Access Barriers
Printed concordance handouts	Inst4: "I prepare backups when Wi-Fi fails"	Pedagogical Improvisation
Group tasks = higher engagement	S5: "We competed to find patterns—it was fun!"	Collaborative Learning

6.2. Triangulation of Instructor Reports and Classroom Observations

The study's triangulation revealed strong alignment between instructor-reported challenges and observed classroom realities. Notably, all 12 instructors (100%) cited institutional barriers as critical, which was corroborated by observations showing that corpus tools were used for less than 15 minutes in most sessions due to technical failures—mirroring Inst10's complaint: "*Tools don't load in our lab.*" This infrastructure deficit was quantifiable: 60% of computers in observed lab sessions failed to run AntConc, directly validating instructor claims. Similarly, instructors who reported "*student resistance*" (Inst3, Inst6) were observed adopting compensatory strategies, such as breaking tasks into smaller steps or providing printed concordance handouts (as Inst4 described: "*I prepare backups when Wi-Fi fails*"). These observations confirmed that pedagogical improvisation—reported by 92% of instructors as a high-impact shift—was a necessary adaptation to systemic constraints.

6.3. Triangulation of Student Engagement and Outcomes

Quantitative and qualitative data converged to demonstrate how task design mediated CDTM's effectiveness. While 83% of instructors reported positive student outcomes, observations revealed this was conditional on implementation quality. For instance, group tasks—which 75% of instructors recommended—showed 50% higher engagement in practice, a finding echoed by Student 5's interview comment: "*We competed to find patterns—it was fun!*" However, the high-impact potential of pedagogical shifts (reported by 92% of instructors) was often undermined by the critical institutional barriers observed. Printed

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

handouts and gamified activities, while effective workarounds, highlighted a disconnect between CDTM's collaborative, exploratory ideals and the realities of resource-limited classrooms. This triangulation underscores that while instructors and students recognize CDTM's value, its transformative potential remains contingent on addressing infrastructure gaps and curricular integration.

7. CONCLUSION

This study explored the implementation of corpus-driven teaching methods (CDTM) in Lahore's undergraduate English language programs, examining student experiences, instructor challenges, and institutional barriers through a triangulated analysis of interviews, observations, and survey data. Five key themes emerged from the research questions: (1) the tension between student appreciation for authentic language discovery and their attachment to traditional textbook learning, (2) the infrastructural and pedagogical constraints instructors faced in adopting CDTM, (3) the measurable benefits of corpus tools for language accuracy and engagement when properly implemented, (4) the socio-cultural resistance to student-centered learning approaches, and (5) the critical need for localized adaptations to make corpus pedagogy viable in Pakistan's educational context.

The findings reveal that while CDTM shows strong potential to enhance language acquisition through data-driven learning, its effectiveness is heavily mediated by contextual factors. Students valued corpus tools for exposing them to real-world language patterns but struggled with technical limitations and desired more teacher guidance—highlighting the need for a balanced approach that integrates corpus exploration with structured instruction. Instructors, despite recognizing CDTM's pedagogical benefits (92%), faced universal institutional barriers, including outdated technology, unreliable internet, and misaligned assessment systems, forcing them to improvise hybrid teaching strategies that often diluted the methodology's impact.

To realize CDTM's full potential in Lahore's universities, the study recommends a multi-level intervention strategy. Institutionally, this includes investing in technological infrastructure, developing Pakistani English corpora for cultural relevance, and redesigning teacher training programs to build corpus literacy. Pedagogically, instructors should adopt scaffolded implementations—combining corpus tasks with explicit instruction—to ease the transition from teacher-led to student-centered learning. These adaptations must be contextualized within Lahore's exam-oriented education culture, where corpus activities should be systematically integrated into graded assessments to ensure student buy-in. Only through such culturally and logistically responsive approaches can corpus-driven teaching transcend its current status as a peripheral innovation and become a transformative force in Pakistan's English language education.

References:

Ahmad, S. (2024). Use of corpora in language teaching: A historical perspective. *International Journal of Innovative Research in Technology*, 11(4). 772-782

Ahmed, S., & Rao, C. (2020). Technology in Language Education: Trends in Pakistan.

BAWE Corpus. British Academic Written English.

Bazeley, P., & Jackson, K. (2007). *Qualitative Data Analysis with NVivo: Pat Bazeley & Kristi Jackson*. sage.

Boulton, A. (2009). Data-driven learning: Taking the computer out of the equation. Language Learning.

Boulton, A., & Pérez-Paredes, P. (2014). ReCALL special issue: Researching uses of corpora for language teaching and learning Editorial Researching uses of corpora for language teaching and learning. *ReCALL*, *26*(2), 121-127.

Brinkmann, S., & Kvale, S. (2018). Doing interviews. 1-208

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Dar, M., et al. (2018). Exam-Oriented Education in Pakistan: Implications for Language Learning.

Denzin, N. K. (2017). *The research act: A theoretical introduction to sociological methods*. Routledge.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors?. *System*, *32*(3), 301-319.

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching?. In *The Routledge handbook of corpus linguistics* (pp. 359-370). Routledge.

Jamal, J., Shafqat, A., & Afzal E. (2021). Teachers' perceptions of incorporation of corpus-based approach in English language teaching classrooms in Karachi, Pakistan. *Liberal Arts and Social Sciences International Journal (LASSI)*, 611-629.

Johns, T. (1991). *Should you be persuaded: Two samples of data-driven learning materials* (Vol. 4, pp. 1-16). na.

Khan, A., Jabeen, S., & Kouser, T. (2022). Technology Integration into ESL Classrooms at University Level. *Pakistan Languages and Humanities Review*, *6*(4), 369-381.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). Qualitative data analysis: A methods sourcebook. 3rd. Sage

Misnawati, M., Yusriadi, Y., Astri, Z., Abbas, A., & Asbar, A. (2025). Exploring Teachers' Initial Perceptions of Corpus Linguistics and Their Readiness to Incorporate Corpora in Classroom Instruction: A Qualitative Analysis. *Qualitative Report*, *30*(1).

O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.

Rana, A. M. K., & Bhatti, A. M. (2020). Attitudes of elementary teachers in Lahore towards professional development. *Journal of Elementary Education*, *30*(1), 84-106.

Saldaña, J. (2021). The coding manual for qualitative researchers.

Shamim, F. (2019). Learning to Teach English. In A. Malay., (2019). *Developing Expertise through Experience*. British Council: London

Shamim, F., & Kausar, S. (2014). English Language Teaching in Pakistan: Policies and Practices.

Smith, J. A. (2017). Interpretative phenomenological analysis: Getting at lived experience. *The journal of positive psychology*, *12*(3), 303-304.

Spradley, J. P. (2016). *Participant observation*. Waveland Press.

Wessendorf, S. (2015). 'All the people speak bad English': Coping with language differences in a super-diverse context.

Zahra, T., & Abbas, A. (2018). Pedagogical implications of Corpus-based approaches to ELT in Pakistan. *Journal of Education and Educational Development*, 5(2), 259-275

Abbas, F., Rana, A. M. K., Bashir, I., & Bhatti, A. M. (2021). The English language proficiency as a global employment skill: the viewpoint of Pakistani academia. *Humanities and Social Sciences Review*, *9*(3), 1071-1077.

Boontam, P., & Phoocharoensil, S. (2018). Effectiveness of English Preposition Learning through Data-Driven Learning (DDL). *3L: The Southeast Asian Journal of English Language Studies*. 125-141

Lan, G., Zhang, Q., Lucas, K., Sun, Y., & Gao, J. (2022). A corpus-based investigation on noun phrase complexity in L1 and L2 English writing. *English for Specific Purposes*, *67*, 4-17.

Lee, J. (2021). Using corpus analysis to extend experimental research: Genre effects in L2 writing. *System*, *100*, 102563.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Binkai, J. (2012). An empirical study on corpus-driven English vocabulary learning in China. English Language Teaching, 5(4), 131–137. https://doi.org/10.5539/elt.v5n4p131

Knight, D., & Adolphs, S. (2022). Building a spoken corpus: what are the basics?. In *The Routledge handbook of corpus linguistics* (pp. 21-34). Routledge.

Azim, M. U., Hussain, Z., Bhatti, A. M., & Iqbal, M. (2020). Recycling of vocabulary in English Language Teaching: From theory to practice. *Epistemology*, *7*(1), 88-102.

Azim, M. U., Bhatti, A. M., Hussain, Z., & Iqbal, M. (2018). Culturally oriented textbooks and English language teachers. *Hamdard Islamicus*, *41*(4), 301-317.

Azim, M. U., Hussain, Z., Bhatti, A. M., & Iqbal, M. (2021). McDonaldization of education in Pakistan: A step towards dehumanization. *International Journal of Innovation, Creativity and Change*, *15*(2), 858-869.

Mushtaq, H., Bhatti, A. M., & Yasmin, T. (2021). A Corpus Based Vocabulary Analysis of Intermediate Book 1 used in the College of Punjab.

Bhatti, A. M., Hussain, Z., Azim, M. U., & Gulfam, G. Q. (2020). Perceptions of ESL learners and teachers on writing difficulties in English language learning in Lahore. *International Bulletin of Literature and Linguistics*, *3*(3), 11-24.

Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*, *28*(2), 207-226.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard university press.

Johns, T. (1991). *Should you be persuaded: Two samples of data-driven learning materials* (Vol. 4, pp. 1-16). na.

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied linguistics*, *11*(2), 129-158.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for specific purposes*, *25*(1), 56-75.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of second language writing*, *13*(4), 257-283.

Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning and Technology*, 9(2), 111-125

Flowerdew, J. (2012). *Discourse in English language education*. Routledge.