


Integrating Corpus Linguistics in Vietnamese English Education: A Qualitative Inquiry Into Scholarly Perspectives and Practices


Thi Thuy Tran, PhD

Nam Can Tho University, Can Tho City, Vietnam

 <https://orcid.org/0009-0000-9920-7025>


Trut Thuy Pham, MA

Nam Can Tho University, Can Tho City, Vietnam

 <https://orcid.org/0009-0005-6596-2686>

Thanh Thao Le, MA

Can Tho University, Can Tho City, Vietnam

 <https://orcid.org/0000-0002-8471-8305>

Contact: thuybate@gmail.com

Abstract

Objective: Our qualitative study investigated the integration of corpus linguistics in English education research among Vietnamese scholars, focusing on the experiences and perceptions across different career stages.

Methods: Employing semi-structured interviews, we engaged nine linguistics scholars, including novices as well as those at mid-career and near-end career stages, from two prominent Vietnamese institutions.

Results: Thematic analysis revealed three key themes: (1) challenges in incorporating corpus linguistics into teaching practices, (2) perceived benefits in enhancing language proficiency, and (3) the importance of institutional support for corpus linguistics research. The findings indicated that despite recognizing the pedagogical benefits of corpus linguistics, scholars across all career stages face significant challenges in its practical application, mainly due to a mismatch between traditional teaching methods and innovative corpus approaches. Our study highlighted the universal acknowledgment of the effectiveness of corpus linguistics in fostering students' understanding of real-world language usage, suggesting potential for curriculum enhancement. However, the effectiveness of its integration was contingent on substantial institutional support, especially for novice researchers who require more resources and training.

Conclusions and Implications: Our research contributed to the understanding of how corpus linguistics is being adopted in non-native English-speaking contexts, particularly in Vietnam. It underscored the need for professional development and institutional support to bridge the gap between theoretical knowledge and practical application. Our study's implications extended to similar educational settings, offering guidance for integrating innovative methodologies into language education.

Keywords: *qualitative inquiry, corpus linguistics, scholarly perspectives, scholarly practices, Vietnamese English education*

Date Submitted: March 17, 2024 | **Date Accepted:** October 1, 2024 | **Date Published:** December 6, 2024

Recommended Citation

Tran, T. T., Pham, T. T., & Le, T. T. (2024). Integrating corpus linguistics in Vietnamese English education: A qualitative inquiry into scholarly perspectives and practices. *Higher Learning Research Communications, 14*(2), 64–77. <https://doi.org/10.18870/hlrc.v14i2.1557>

Introduction

The landscape of English language education research is continually evolving, reflecting the dynamic interplay of global trends, technological advancements, and pedagogical theories. Amid these developments, one notable trend has emerged: corpus linguistics. In recent years, corpus linguistics has emerged as a significant trend within this field, especially in non-native English-speaking countries like Vietnam (Ngoc and Barrot, 2023). Our study aimed to explore and elucidate the growing influence of corpus linguistics in English education research among Vietnamese scholars.

Corpus linguistics, the study of language as expressed in corpora (samples of real-world text), offers a unique perspective to language teaching and learning (Meyer, 2023). By analyzing large collections of authentic language data, corpus linguistics provides empirical insights that can significantly enhance language education. Its empirical approach to language analysis empowers educators and researchers to derive insights from authentic language usage patterns. This is particularly pertinent in the context of English education in Vietnam, where the language is predominantly taught as a second or foreign language.

The surge of interest in corpus linguistics among Vietnamese scholars can be attributed to several factors. Foremost among these is the increasing accessibility of digital text corpora and sophisticated analytical tools. First, the increasing accessibility to large digital text corpora and advanced analytical tools has democratized the field, allowing researchers from diverse backgrounds to engage with this methodology (Van Nguyen et al., 2022). Second, the shift toward communicative and empirical methods in language education globally has spurred interest in data-driven approaches like corpus linguistics (Vu, 2019). Finally, Vietnam's integration into the global community, and the corresponding emphasis on English proficiency, has heightened the need for innovative and effective English teaching methodologies (Chau and Truong, 2019).

Despite its growing popularity, there remains a dearth of comprehensive studies examining the adoption and impact of corpus linguistics within the Vietnamese context. Our study sought to address this gap by providing an in-depth qualitative analysis of the current state of corpus linguistics in Vietnam. This gap in the literature signaled a need for a qualitative exploration into how Vietnamese scholars were integrating corpus linguistics into English education research. By employing a qualitative methodology, our study sought to provide an in-depth understanding of the experiences, perceptions, and challenges faced by these scholars.

Literature Review

The purpose of our literature review was to provide a comprehensive overview of corpus linguistics and its application in language teaching. By outlining significant studies and frameworks in this area, we aimed to contextualize our present study within the broader field and highlight its relevance and novelty.

Corpus Linguistics in Language Education

Corpus linguistics, a field that involves the systematic study of language through large collections of real-world text samples, has evolved as a crucial methodological tool in linguistics (McEnery & Hardie, 2011). This discipline focuses on identifying patterns and usage trends within these texts to provide empirical insights into language use. The term “corpus linguistics” refers to this approach of analyzing extensive datasets, or corpora, to draw conclusions about language structure and function.

Corpus linguistics, which initially emerged as a methodological tool in linguistics (McEnery & Hardie, 2011), has significantly influenced language education over the past few decades. This field involves the systematic analysis of large collections of real-world text samples to identify patterns and usage trends. The analysis can include examining the frequency of words, phrase structures, and even syntactic patterns, which are then used to inform various linguistic theories and applications. The foundational principles of corpus linguistics revolve around the systematic analysis of large text samples to uncover patterns and usage trends. Meyer (2023) highlighted its role in providing empirical data about language use, which has proven invaluable in informing language teaching and curriculum design. Recent studies (e.g., Yilmaz & Koc, 2020) have demonstrated the effectiveness of corpus-based approaches in enhancing learners’ understanding of pragmatic aspects of language, such as collocation and idiomatic expressions, which are often challenging for non-native speakers. In summary, corpus linguistics offers a data-driven approach to language analysis that can greatly benefit language education by providing authentic examples and patterns of real-world language use.

Significant Studies in Corpus Linguistics and Language Teaching

Several key studies have demonstrated the impact of corpus linguistics on language education. For instance, Biber et al. (1998) conducted seminal work on using corpora to inform grammar instruction, illustrating how authentic language data can enhance learners’ grammatical competence. Similarly, Conrad (2000) emphasized the importance of corpus-based materials in teaching register and genre, helping students to understand language variation across different contexts. These studies have underscored the utility of corpus linguistics in providing empirical evidence that can refine and improve language teaching methodologies.

Adoption and Impact in Non-Native English-Speaking Countries

The global spread of English has led to a surge in adopting corpus linguistics methods in countries where English is taught as a second or foreign language. This trend has highlighted the adaptability and effectiveness of corpus linguistics across various educational contexts as well as the versatility and effectiveness of corpus linguistics in diverse educational settings. Researchers like Kirkpatrick (2010) and Sarvandy and Ekstam (2018) have documented its adoption in various Asian and European contexts, noting improvements in both teaching methodologies and learning outcomes. Particularly, studies focusing on Asian contexts (e.g., Ma & Mei, 2021) have shed light on how corpus linguistics tools and resources have been adapted to local educational settings, addressing specific linguistic challenges faced by learners in these regions. Additionally, Lang and Jing-Schmidt (2022) have highlighted the significance of building collocational knowledge between lexicon and grammar, demonstrating the role of corpus linguistics in enhancing pedagogical grammar in Chinese as a second language, which parallels its application in other non-native English-speaking contexts.

Developments and Challenges in the Vietnamese Context

In Vietnam, the integration of corpus linguistics into English education is a relatively recent phenomenon. This emerging interest was driven by both educational reforms and the increasing availability of digital corpora. The Vietnamese academic community has shown a growing interest in this field, propelled by both external and internal educational reforms. Nguyen (2018) noted that the increasing interest among Vietnamese scholars in corpus-based research was driven by the growing availability of localized corpora and

the government's emphasis on English proficiency. However, Nghi and Phuc (2022) pointed out challenges, such as limited access to robust language corpora, a lack of training in corpus analysis tools, and resistance to changing traditional, textbook-centric teaching methods.

Studies like those of Nguyen (2018) have highlighted successful corpus-based interventions in Vietnamese higher education settings, showing improvements in students' lexical richness and awareness of language nuances. Conversely, Hashimoto and Nguyen (2018) underscored the need for more comprehensive teacher training programs to facilitate the effective integration of corpus linguistics in classrooms. These findings highlighted both the potential benefits and the practical hurdles of implementing corpus linguistics in Vietnam.

In summary, the literature has suggested that while corpus linguistics has a growing influence on English education in non-native contexts, including in Vietnam, its adoption is not without challenges. Our study aimed to build on existing research by providing a qualitative exploration of these challenges and identifying strategies for effectively integrating corpus linguistics into Vietnamese English education. Our study built upon existing literature to explore these challenges in greater depth, offering insights into how Vietnamese scholars are leveraging corpus linguistics to enhance English language education. Our review set the stage for our present study, which aimed to provide a deeper, qualitative understanding of how Vietnamese scholars are navigating these challenges and harnessing the potential of corpus linguistics in English language education.

Methods

To comprehensively explore the integration of corpus linguistics into English education among Vietnamese scholars, our study employed a qualitative research design, ensuring a nuanced and detailed understanding of the subject matter.

Research Design

Our present study employed a qualitative research design, underpinned by a constructivist paradigm that acknowledged the subjective nature of reality as perceived and experienced by individuals (Creswell & Poth, 2016). This qualitative approach allowed for an in-depth exploration of personal experiences and perceptions, providing rich, contextual insights. Our approach was particularly suited to exploring the nuanced perspectives and experiences of Vietnamese scholars in integrating corpus linguistics into English education. The theoretical framework guiding our study was rooted in two key areas: (1) the sociocultural theory of language education and (2) the diffusion of innovations theory.

Sociocultural theory, as proposed by Vygotsky (1978), emphasized the importance of social context and interaction in the process of learning. By focusing on the social aspects of learning, this theory helped to understand the collective and individual processes involved in adopting new educational practices. This perspective was critical in understanding how educational practices, like the adoption of corpus linguistics, are shaped by and adapted to the specific cultural and institutional contexts within which Vietnamese scholars operate. It provided a lens to examine how these scholars negotiated and made meaning of their experiences within the broader social and educational milieu of Vietnam.

The Diffusion of Innovations theory, articulated by Rogers et al. (2014), offered a framework to analyze how new ideas, practices, or technologies are adopted and disseminated within a community. This theory helped to identify the stages and factors that influenced the acceptance and spread of new methodologies. This theory assisted in exploring the factors influencing the adoption of corpus linguistics among Vietnamese scholars, including the perceived advantages, compatibility with existing practices, and the role of social networks and communication channels in spreading these innovations.

Participants

Our study focused on a purposively selected group of nine Vietnamese linguistics scholars, representing a cross-section of career stages, to provide a diverse range of perspectives on the integration of corpus linguistics in English education. This purposeful selection ensured a comprehensive understanding across different levels of academic experience. These participants were affiliated with two prominent educational institutions in Vietnam, known for their active engagement in English language education and research. The selection was deliberately stratified to include three novice scholars (those in the early stages of their academic careers), three mid-career scholars (with substantial experience and established research profiles), and three near-end career scholars (who possessed extensive experience and were considered veterans in the field). This stratification ensured a comprehensive understanding of how the adoption and perception of corpus linguistics may vary across different career stages.

The inclusion of participants from two different institutions also allowed for the exploration of potential institutional influences on the adoption and application of corpus linguistics. These institutions were chosen based on their distinct educational environments and research cultures, which were expected to provide varied contexts for the study.

Addressing ethical considerations was a cornerstone of this research. Ensuring the ethical integrity of the study was crucial for the validity and reliability of the findings. Prior to conducting the interviews, we provided all participants with detailed information about the study's purpose, the nature of their involvement, and the ways in which their data would be used. Informed consent was obtained from each participant, ensuring that they were willingly participating and understood their rights, including the right to withdraw from the study at any point without any adverse consequences.

Confidentiality and anonymity were of paramount importance in this research. To protect the identity of the participants, pseudonyms were used in place of real names in any published work or presentations arising from our study. Additionally, any potentially identifying information was carefully omitted or altered in the presentation of our findings.

Data Collection

The data collection for our study was primarily conducted through semistructured interviews, chosen for their flexibility and depth. Semi-structured interviews allowed for the exploration of specific topics while providing the flexibility to delve deeper into unexpected areas of interest. To ensure the reliability and relevance of the interview process, several specific actions and processes were undertaken.

Prior to the main data collection phase, we conducted a pilot study with two Vietnamese linguistics scholars who were not part of the main study. The purpose of the pilot study was to test the interview protocol, ensuring that the questions were clear, relevant, and effectively elicited the required information. Based on the feedback from the pilot study, necessary revisions were made to the interview questions to enhance their clarity and relevance to the study's objectives.

The revised interview questions included queries such as, "How do you perceive the role of corpus linguistics in current English language education in Vietnam?" and "Can you describe any challenges you have faced in integrating corpus linguistics into your research or teaching?" These questions were designed to elicit detailed responses about the participants' experiences and perceptions regarding the use of corpus linguistics in English education.

Each interview lasted approximately 60–90 minutes, providing ample time for participants to elaborate on their experiences and insights. During the interviews, participants were encouraged to provide specific

examples and anecdotes to illustrate their points, ensuring a rich and detailed data set. The interviews were conducted in places convenient for the participants, such as their offices or a quiet location within their respective institutions, ensuring a comfortable environment conducive to open and detailed discussions.

Considering the linguistic context of the participants, all interviews were conducted in Vietnamese. Conducting interviews in the participants' native language ensured that responses were genuine and reflective of their true thoughts and experiences. This approach was essential to capture the nuanced meanings and cultural contexts accurately. This choice was made to ensure that participants could express themselves freely and accurately, without the constraints of a language barrier. For participants who were more comfortable in English, they were given the option to respond in English if they preferred. This approach not only facilitated ease of communication but also ensured that nuanced meanings and cultural contexts were accurately captured.

Data Analysis

We used thematic analysis, a method well-suited for identifying, analyzing, and reporting patterns (themes) within data, to analyze the data collected from the semi-structured interviews. Thematic analysis provided a structured yet flexible approach to analyzing qualitative data, allowing for the identification of key themes and patterns. This approach allowed for a rich and detailed, yet complex, account of the data. Thematic analysis in our study was guided by the six-phase framework described by Braun et al. (2023), which included (1) familiarizing oneself with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report.

The first phase involved a thorough immersion into the data through repeated reading of the interview transcripts, enabling a deep familiarity with the depth and breadth of the content. During this process, we noted initial ideas for coding. The subsequent phase of generating initial codes was carried out systematically across the entire data set. This involved identifying and labeling significant features of the data that were relevant to the research questions. These codes were then collated into potential themes, considering the frequency and significance of the codes.

Following the initial coding, we began a more focused phase of searching for themes. This involved sorting the different codes into potential themes and gathering all data relevant to each potential theme. This phase was crucial for organizing the data into meaningful categories that reflect the research questions. Themes were developed by combining codes that shared similar patterns or meanings. The next phase involved reviewing these themes, ensuring they worked in relation to the coded extracts and the entire data set, and forming a coherent pattern. This step involved refining the themes by considering their coherence and distinctiveness. This step sometimes led to the refinement or combination of themes.

The fifth phase, defining and naming themes, involved a detailed analysis of each theme, identifying the essence of what each theme captured about the data and determining how it related to a research question. Each theme was clearly defined and named to reflect its central concept and contribution to the overall analysis. Defining and naming themes helped to articulate the core messages and insights derived from the data. This phase was crucial in refining the specifics of each theme and the overall story the analysis told, determining the 'why' behind each theme.

Finally, the last phase involved the production of the report, where we wrote the analysis. The report-writing phase synthesized the findings and connected them to the broader research context. This included a detailed analysis of each theme, supported by relevant quotes from the participants, and a discussion of how the themes related to the research questions and the literature previously reviewed.

Findings

Challenges Incorporating Corpus Linguistics Into Teaching Practices

One prominent theme that emerged was the challenge of incorporating corpus linguistics into teaching practices. This theme was reflected in the responses of all nine participants across the different career stages, highlighting its universal relevance. The specifics of these challenges, however, varied depending on the career stage, the type of teaching, and the nature of the corpus linguistics tools being used. Notably, six participants, including two novices, two mid-career, and two near-end scholars, specifically emphasized the difficulty of integrating corpus linguistics tools and methodologies into the existing curriculum.

A novice scholar, Dr. Huyen, expressed, “As a newcomer in the field, I find it challenging to incorporate corpus-based tools in my teaching. There is a gap between the theoretical understanding of corpus linguistics and its practical application in the classroom.” This illustrated the difficulty faced by less-experienced scholars in moving from theory to practice. This sentiment was echoed by a mid-career scholar, Dr. Minh, who noted, “Despite my experience, adapting these tools to suit our traditional teaching methods is tough. It is like trying to fit a square peg into a round hole.” This comment highlighted the challenge of reconciling new methodologies with entrenched teaching practices. From the perspective of a near-end scholar, Dr. Phuong remarked, “Over the years, I have seen many changes in teaching methodologies. However, integrating something as data-intensive as corpus linguistics into our existing framework is a daunting task.” This underscored the persistent difficulty of integrating new, complex tools into a well-established educational system.

The challenges also varied depending on the type of corpus linguistics being employed. For example, some scholars found it particularly challenging to use specialized corpora for teaching specific language skills, such as collocations or idiomatic expressions, because these required advanced analytical skills and understanding that are not typically covered in standard teacher training programs. Dr. Linh, a novice scholar, noted, “Using specialized corpora for teaching collocations is difficult because it requires a deep understanding of both the tool and the linguistic feature being taught.” Additionally, the kind of teaching influenced the nature of the challenges. Those teaching higher education students found it somewhat easier to incorporate corpus-based approaches compared with those teaching at the secondary level. Dr. Tuan, a mid-career scholar, mentioned, “In higher education, students are more receptive to new tools, but at the secondary level, it is harder to introduce these methods because the curriculum is more rigid and exam focused.”

These excerpts demonstrated the pervasive nature of the challenge, irrespective of the career stage. Through the lens of sociocultural theory, this finding could be interpreted as a reflection of the complex interplay between the scholars’ existing teaching practices, which are deeply rooted in the educational culture of Vietnam, and the innovative approach of corpus linguistics. The tension between traditional and new methods underscored the influence of social and cultural contexts in the adoption of new educational practices.

From the perspective of the Diffusion of Innovations theory, this challenge could be seen as a matter of compatibility, one of the key factors influencing the adoption of an innovation. The difficulty in aligning corpus linguistics with existing teaching practices suggested a lack of compatibility, which can hinder the adoption process. This was particularly evident in the case of novice scholars like Dr. Huyen, who, despite being theoretically knowledgeable about corpus linguistics, struggled with its practical application in a teaching environment that had not traditionally embraced such methodologies.

Perceived Benefits of Corpus Linguistics in Enhancing Language Proficiency

Another significant theme that emerged from the analysis was the perceived benefits of corpus linguistics in enhancing language proficiency. All nine participants consistently mentioned this theme, indicating a universal recognition of the value of corpus linguistics in language education. Specifically, seven participants, including

three novices, two mid-career, and two near-end scholars, highlighted the positive impact of corpus linguistics on improving students' language proficiency, particularly in understanding real-world language use.

The aspects of corpus linguistics used by the cited scholars varied and directly related to the statements in the previous section. For instance, Dr. Linh, a novice scholar, shared, "Using corpus linguistics in my classes has helped students grasp how English is used in real contexts. They seem to understand idiomatic expressions better." This indicated the use of authentic language examples from corpora to teach idiomatic expressions, which aligned with the challenge of integrating specialized corpus tools. This observation was reinforced by a mid-career scholar, Dr. Tuan, who stated, "My students have shown noticeable improvement in their language intuition. They are more adept at picking up language patterns and usage." Dr. Tuan's use of corpus linguistics focused on recognizing language patterns and usage, which reflected the general challenge of adapting corpus-based methods to traditional teaching practices. Similarly, a near-end career scholar, Dr. Anh, commented, "Regarding my experiences, I have seen how exposure to authentic language through corpus examples enhances students' practical language skills, especially in comprehension and usage." Dr. Anh's emphasis on comprehension and usage aligned with the broad challenge of integrating data-intensive corpus methods into existing frameworks.

These excerpts underlined the perceived effectiveness of corpus linguistics in enhancing language proficiency, a view shared across the career spectrum. From the perspective of sociocultural theory, this finding could be understood as the scholars recognizing the value of engaging with authentic language, a key component of corpus linguistics, which aligned well with the theory's emphasis on the importance of real-world context in learning. This approach facilitated a deeper understanding of language as it is naturally used, fostering a more contextual and practical learning experience for students.

In terms of the Diffusion of Innovations theory, this positive perception of the benefits of corpus linguistics could be seen as a factor that promoted its adoption among educators. According to Rogers, the perceived relative advantage of an innovation greatly influenced its adoption rate. In this case, the belief that corpus linguistics could significantly enhance students' language proficiency acted as a compelling advantage that encouraged its integration into teaching practices.

Institutional Support and Resources for Corpus Linguistics Research

The third key theme identified was institutional support and resources for corpus linguistics research. This theme emerged prominently in the discussions, with all nine participants commenting on it. Notably, the level of emphasis on institutional support varied across career stages: All three novice scholars, two mid-career scholars, and only one near-end scholar specifically highlighted the critical role of institutional backing in facilitating corpus linguistics research.

Dr. Bao, a novice scholar, mentioned, "Starting in this field, I quickly realized the need for institutional support, especially in terms of access to corpus resources and training." This perspective was shared by a mid-career scholar, Dr. Quynh, who noted, "The success of integrating corpus linguistics in our work largely depends on the resources and support provided by our institutions. Funding, access to software, and training make a significant difference." In contrast, Dr. Tam, a near-end scholar, reflected a different viewpoint, "Over my career, I have seen gradual improvements in institutional support, but more is needed, especially for young researchers."

These excerpts illustrated the perceived importance of institutional support in corpus linguistics research, with a particular emphasis on the challenges faced by less-experienced scholars. Sociocultural theory provided a framework for understanding this finding, as it emphasized the role of social and institutional structures in shaping educational practices. The need for institutional support, as highlighted by the participants, reflected

the interaction between individual scholars and their institutional environments, impacting their ability to effectively engage in corpus linguistics research.

From the viewpoint of the Diffusion of Innovations theory, the emphasis on institutional support could be interpreted as a factor influencing the rate of adoption and implementation of corpus linguistics in research. The theory posited that organizational facilitation, including the provision of resources and support, was crucial for the adoption of new innovations. This was particularly evident in the case of novice scholars like Dr. Bao, who required more institutional support to overcome the barriers to accessing and utilizing corpus linguistics tools and resources.

Discussion

Challenges Incorporating Corpus Linguistics Into Teaching Practices

This theme resonated with existing literature, yet also offered unique insights specific to the Vietnamese context. Previous studies in various geographical and educational settings have acknowledged the challenges associated with integrating corpus linguistics into language education. For instance, research by Curado Fuentes (2023) highlighted similar concerns among European educators regarding the practical application of corpus tools in teaching. However, our present study contributed new dimensions to this discourse by underscoring the specific challenges faced in a Vietnamese educational setting, which included reconciling traditional teaching methods with corpus linguistics approaches.

The findings of our study aligned with those of Nguyen (2018), who noted the growing interest in corpus linguistics among Vietnamese scholars. However, our study extended this observation by revealing the nuanced difficulties in integrating corpus linguistics into existing teaching frameworks, regardless of the educator's career stage. This topic was particularly noteworthy, as it contrasted with some of the existing literature that suggested such challenges were predominantly faced by less-experienced educators. For example, studies by Burkhauser and Lesaux (2017) in the United States highlighted that novice teachers faced greater challenges in adopting corpus methods due to their limited experience. In contrast, our study found that even experienced, near-end career scholars in Vietnam encountered significant challenges, indicating a broader systemic issue rather than an individual level of experience.

Through the lens of sociocultural theory (Vygotsky, 1978), our study's findings emphasized the significant impact of the social and cultural context of Vietnamese education on the adoption of new teaching methodologies. This perspective was somewhat underexplored in the existing literature, which often focused more on the technical or resource-based challenges of integrating corpus linguistics (e.g., Zhao & Fan, 2018). Our study contributed to the field by highlighting the cultural and institutional dimensions of these challenges.

Furthermore, the Diffusion of Innovations theory provided a framework for understanding the adoption process of corpus linguistics in Vietnam. While previous studies have applied this theory to understand the adoption of educational innovations (e.g., Raynard, 2017), our current study uniquely applied it to the context of corpus linguistics in a non-native English-speaking country. It underscored the importance of compatibility as a crucial factor in the adoption of innovations, a point that was not extensively explored in existing studies focusing on corpus linguistics.

Perceived Benefits of Corpus Linguistics in Enhancing Language Proficiency

The perceived benefits of corpus linguistics in enhancing language proficiency reflected a growing consensus in the field of language education, yet the findings from our study added a unique dimension by contextualizing these benefits within the Vietnamese educational landscape. Previous research, such as

studies by O’Keeffe and McCarthy (2010), has similarly identified the advantages of using corpus linguistics for language learning, particularly in improving learners’ understanding of real-world language use. However, our current study extended this understanding by highlighting the specific ways in which Vietnamese scholars across different career stages perceived and experienced these benefits.

The universal recognition among the study’s participants of the value of corpus linguistics aligned with international research, yet it contrasted with some studies that suggested variations in perceptions based on teaching experience. For instance, research in an American context by O’Keeffe et al. (2007) indicated that more experienced educators were likely to report the benefits of corpus linguistics in language education. In contrast, our study found that even novice scholars in Vietnam recognized and appreciated the practical language skills gained through corpus linguistics, suggesting a more uniformly positive perception across career stages.

Our findings also enriched the application of sociocultural theory in language education research. While the theory has been used to understand the role of authentic language contexts in learning, our study provided empirical evidence from Vietnam, highlighting how exposure to real-life language usage through corpus linguistics enhanced language proficiency. This aligned with Vygotsky’s (1978) emphasis on the importance of interaction with the authentic language environment in the learning process but was contextualized within the specific Vietnamese educational setting.

Furthermore, the application of the Diffusion of Innovations theory in this context was particularly insightful. The theory, as applied in previous studies (e.g., Greenhalgh et al., 2005; Grgurovic, 2010), often focused on technological or methodological innovations in broader educational contexts. Our current study, however, applied it specifically to the field of corpus linguistics in language education. It illustrated that the perceived benefits of corpus linguistics, as identified by the Vietnamese scholars, served as a key factor in its adoption, consistent with the emphasis on the perceived relative advantage of an innovation as discussed by Rogers et al. (2014).

Institutional Support and Resources for Corpus Linguistics Research

The current theme highlighted a critical aspect of academic research that was often echoed in the broader literature, yet our study brought unique insights into the specific context of Vietnamese academia. The emphasis on institutional support aligned with findings from various international studies that underscored the importance of institutional backing in fostering innovative research practices, such as the work by Engwall and Hedmo (2016) on the role of universities in supporting corpus linguistics. However, our study extended this understanding by specifically addressing the varying levels of institutional support experienced across different career stages in Vietnam, which is a novel contribution to the field.

The findings suggested that, while there is a general acknowledgment of the importance of institutional support among all career stages, novice scholars, in particular, felt a more acute need for such support. This contrasted with some existing studies that implied a more uniform perception of institutional support across career stages. For example, studies in European and North American contexts (e.g., Flowerdew, 2012) have generally not differentiated the perception of institutional support based on career stage. Our current study, therefore, contributed new knowledge by highlighting how the needs and perceptions regarding institutional support could vary significantly depending on the researcher’s experience level in the Vietnamese academic setting.

From a sociocultural theory perspective, these findings emphasized the influence of institutional structures and cultures on scholarly practices. While existing research often focused on the individual scholar’s capabilities and motivations, our study highlighted the critical role of the institutional environment, aligning with Vygotsky’s (1978) assertion of the importance of social context in learning and development. This

perspective was particularly relevant in understanding the experiences of Vietnamese scholars, where institutional culture and resources play a pivotal role in shaping research practices.

Furthermore, applying the Diffusion of Innovations theory (Rogers et al., 2014), our study shed light on how organizational support and facilitation were key factors in the adoption of corpus linguistics in research. While the importance of institutional support is a well-documented factor in the adoption of innovations (Jun & Weare, 2011), the specific challenges and needs identified by Vietnamese scholars at different career stages provided a more nuanced understanding of these dynamics. This was especially significant for novice researchers, who required more tangible support in the form of resources and training to overcome barriers in adopting new methodologies.

Conclusion

Our study set out to explore the integration of corpus linguistics within the field of English education research among Vietnamese scholars. Recognizing the growing global influence of corpus linguistics in language education and the unique position of Vietnam in this context, our study aimed to provide insights into how Vietnamese scholars adopted and adapted this methodology in their teaching and research practices. Employing a qualitative research design, our study used semi-structured interviews with nine Vietnamese linguistics scholars from different career stages and two institutions, which offered a comprehensive perspective across a diverse range of experiences. The methods employed, including a pilot study followed by thematic analysis of the interview data, were instrumental in uncovering rich, nuanced insights. The thematic analysis allowed for a detailed exploration of the themes that emerged from the data, which provided a deep understanding of the participants' perspectives and experiences.

The findings of our study revealed three key themes: (1) the challenges of incorporating corpus linguistics into teaching practices, (2) the perceived benefits of corpus linguistics in enhancing language proficiency, and (3) the crucial role of institutional support and resources for corpus linguistics research. These themes highlighted the complexity and multifaceted nature of integrating corpus linguistics in the Vietnamese context. The challenges, universally acknowledged across career stages, underscored the tension between traditional teaching methods and innovative corpus linguistics approaches. Conversely, the recognized benefits of corpus linguistics in enhancing language proficiency reflected a universal appreciation of its practical value in real-world language education. Last, the emphasis on institutional support and resources revealed the significant impact of institutional factors on the adoption and effective implementation of corpus linguistics, particularly for novice scholars.

The implications of our study are multifaceted, reflecting the complex interplay of pedagogical innovation, cultural context, and institutional frameworks. First, the challenges in incorporating corpus linguistics into teaching practices underscored the need for professional development initiatives. Educational institutions and policymakers should consider investing in targeted training programs that bridge the gap between theoretical knowledge of corpus linguistics and its practical application in the classroom. These programs should not only focus on the technical aspects of corpus linguistics tools but also on pedagogical strategies to integrate these tools effectively within existing curricula.

Second, the study's findings regarding the perceived benefits of corpus linguistics in enhancing language proficiency have significant implications for curriculum design. The positive impact of corpus linguistics on students' understanding of real-world language usage suggested that incorporating corpus-based analyses and materials into language courses could be highly beneficial. This integration could take various forms, such as the use of corpus-derived examples in teaching materials, corpus-based exercises, and projects. Such an approach would align well with contemporary language teaching methodologies that emphasize authentic, contextualized language use.

The emphasis on the critical role of institutional support and resources for corpus linguistics research pointed to a broader implication for university administration and funding bodies. There was a clear indication that successful integration of corpus linguistics in research and teaching requires not only financial investment in resources, such as software and corpora, but also in creating an institutional culture that supports and values innovative research methodologies. For institutions, this might mean re-evaluating resource allocation, offering grants or funding opportunities specifically for corpus linguistics projects, and fostering a research environment that encourages interdisciplinary collaboration and methodological innovation.

Furthermore, our study's findings have implications for the broader academic community, particularly in non-native English-speaking countries. The insights gained from the Vietnamese context can inform similar efforts in other countries where English is taught as a second or foreign language. Our study highlighted the universality of certain challenges and benefits, suggesting that the strategies developed in Vietnam could be adapted to similar contexts elsewhere.

Limitations and Recommendations for Further Studies

Our study, while providing valuable insights, had certain limitations that should be acknowledged. First, the sample size, though diverse in career stages and institutions, was relatively small and limited to two educational institutions in Vietnam. This confines the generalizability of the findings, as they may not fully represent the experiences and perceptions of all Vietnamese scholars working with corpus linguistics. Additionally, the qualitative nature of our study, while offering depth in understanding individual perspectives, lacked the broader representation that quantitative methods might provide. Another limitation was the focus on English language education, which might overlook how corpus linguistics is being integrated into the study of other languages in Vietnam. Finally, the study relied solely on self-reported data from interviews, which may be subject to biases and may not have fully captured the practical implementation of corpus linguistics in classroom settings.

In light of these limitations, several recommendations for future research can be made. Future studies could benefit from a larger and more diverse sample, possibly including a wider range of institutions across different regions of Vietnam, to enhance the generalizability of the findings. Incorporating quantitative methods, such as surveys or experimental designs, could provide a broader overview of the adoption and impact of corpus linguistics in language education. Further research could also explore the application of corpus linguistics in the teaching and learning of languages other than English, providing insights into its versatility and adaptability in different linguistic contexts. Additionally, observational studies or action research that document the actual implementation of corpus linguistics in classroom settings could complement the self-reported data and provide a more comprehensive view of its practical challenges and benefits. Such studies would contribute to a more holistic understanding of the role of corpus linguistics in language education, both within Vietnam and in similar educational contexts globally.

Acknowledgments

The research team utilized ChatGPT (OpenAI, 2024) for language editing, as the researchers are not native English speakers. When prompted with "Check grammar of this sentence and suggest a better way to write it," ChatGPT checked and suggested another way to write the sentences. Then, the researchers would consider whether they should change as suggested or retain the original language. Researchers, however, ensured that all data were reliable and valid, with no fabrication involved.

References

- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Braun, V., Clarke, V., & Hayfield, N. (2023). *Thematic analysis: A reflexive approach*. SAGE Publications.
- Burkhauser, M. A., & Lesaux, N. K. (2017). Exercising a bounded autonomy: Novice and experienced teachers' adaptations to curriculum materials in an age of accountability. *Journal of Curriculum Studies*, 49(3), 291–312. <https://doi.org/10.1080/00220272.2015.1088065>
- Chau, T. H. H., & Truong, V. (2019). The integration of intercultural education into teaching English: What Vietnamese teachers do and say. *International Journal of Instruction*, 12(1), 441–456. <https://doi.org/10.29333/iji.2019.12129a>
- Conrad, S. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34(3), 548–560. <https://doi.org/10.2307/3587743>
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Curado Fuentes, A. (2023). Corpus affordances in foreign language reading comprehension. In K. Harrington & P. Ronan (Eds.), *Demystifying corpus linguistics for English language teaching* (pp. 99–118). Springer International Publishing.
- Engwall, L., & Hedmo, T. (2016). The organizing of scientific fields: The case of corpus linguistics. *European Review*, 24(4), 568–591. <https://doi.org/10.1017/S1062798716000259>
- Flowerdew, J. (2012). *Discourse in English language education*. Routledge.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., & Peacock, R. (2005). Storylines of research in diffusion of innovation: A meta-narrative approach to systematic review. *Social Science & Medicine*, 61(2), 417–430. <https://doi.org/10.1016/j.socscimed.2004.12.001>
- Grgurovic, M. (2010). Technology-enhanced blended language learning in an ESL class: A description of a model and an application of the Diffusion of Innovations theory. [Dissertation, *Iowa State University* 10.18870/hlrc.v14i2.1557]. <https://dr.lib.iastate.edu/entities/publication/123e2afe-aff4-4ebe-98e3-fe93aecd5966>
- Hashimoto, K., & Nguyen, V. T. (2018). *Professional development of English language teachers in Asia: Lessons from Japan and Vietnam*. Routledge.
- Jun, K. N., & Weare, C. (2011). Institutional motivations in the adoption of innovations: The case of e-government. *Journal of Public Administration Research and Theory*, 21(3), 495–519. <https://doi.org/10.1093/jopart/muq020>
- Kirkpatrick, A. (2010). Researching English as a lingua franca in Asia: The Asian Corpus of English (ACE) project. *Asian Englishes*, 13(1), 4–18. <https://doi.org/10.1080/13488678.2010.10801269>
- Lang, J., & Jing-Schmidt, Z. (2022). Building collocational knowledge between lexicon and grammar. In F. Yuan, B. He, & W. Hu (Eds.), *Pedagogical grammar and grammar pedagogy in Chinese as a second language* (pp. 34–55). Routledge.
- Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177–190. <https://doi.org/10.1515/jccall-2021-2008>
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Meyer, C. F. (2023). *English corpus linguistics: An introduction*. Cambridge University Press.

- Nghi, T. T., & Phuc, T. H. (2022). A corpus-based study on the most frequently used English prepositions: A conceptual transfer perspective. *Theory and Practice in Language Studies*, 12(2), 213–220. <https://doi.org/10.17507/tpls.1202.01>
- Ngoc, B. M., & Barrot, J. S. (2023). Current landscape of English language teaching research in Southeast Asia: A bibliometric analysis. *The Asia-Pacific Education Researcher*, 32(4), 517–529. <https://doi.org/10.1007/s40299-022-00673-2>
- Nguyen, T. T. T. (2018). A corpus-based study on cross-cultural divergence in the use of hedges in academic research articles written by Vietnamese and native English-speaking authors. *Social Sciences*, 7(4), 70. <https://doi.org/10.3390/socsci7040070>
- O’Keeffe, A., & McCarthy, M. (2010). *The Routledge handbook of corpus linguistics* (Vol. 10). Routledge.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Raynard, M. (2017). Understanding academic e-books through the diffusion of innovations theory as a basis for developing effective marketing and educational strategies. *The Journal of Academic Librarianship*, 43(1), 82–86. <https://doi.org/10.1016/j.acalib.2016.08.011>
- Rogers, E. M., Singhal, A., & Quinlan, M. M. (2014). Diffusion of innovations. In D. W. Stacks & M. B. Salwen (Eds.), *An integrated approach to communication theory and research* (pp. 432–448). Routledge.
- Sarvandy, E., & Ekstam, J. (2018). Principles of the ‘Lingua Franca Approach’ and their implications for pedagogical practice in the Iranian context. *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, 6(1), 39–52. <https://doi.org/10.22049/jalda.2019.26265.1073>
- Van Nguyen, K., Van Huynh, T., Nguyen, D. V., Nguyen, A. G. T., & Nguyen, N. L. T. (2022). New Vietnamese corpus for machine reading comprehension of health news articles. *Transactions on Asian and Low-Resource Language Information Processing*, 21(5), 1–28. <https://doi.org/10.48550/arXiv.2006.11138>
- Vu, T. N. (2019). Theoretical constructs and practical strategies for intercultural communication. *Journal of Curriculum Studies Research*, 1(1), 43–53. <https://doi.org/10.46303/jcsr.01.01.4>
- Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.
- Yilmaz, N., & Koc, D. K. (2020). Developing pragmatic comprehension and production: Corpus-based teaching of formulaic sequences in an EFL setting. *Journal of Language and Linguistic Studies*, 16(1), 474–488. <https://doi.org/10.17263/jlls.712880>
- Zhao, Y., & Fan, B. (2018). Exploring open government data capacity of government agency: Based on the resource-based theory. *Government Information Quarterly*, 35(1), 1–12. https://www.researchgate.net/publication/323121687_Exploring_open_government_data_capacity_of_government_agency_Based_on_the_resource-based_theory

The *Higher Learning Research Communications (HLRC)* is a peer-reviewed, online, interdisciplinary journal indexed in Scopus, ERIC, JGATE and Directory of Open Access Journals (DOAJ). It is an open access journal with an international focus published by Walden University, USA. Its aim is to disseminate both high quality research and teaching best practices in tertiary education across cultures and disciplines. *HLRC* connects the ways research and best practice contribute to the public good and impact the communities that educators serve. *HLRC* articles include peer-reviewed research reports, research briefs, comprehensive literature reviews, and books reviews.