Ed Process Int J

www.edupij.com

ISSN 2147-0901 (Print), 2564-8020 (Online)

EDUCATIONAL PROCESS
INTERNATIONAL JOURNAL

Corpus-Driven Innovation in Saudi Arabian EFL Teaching Practices

Abdullah Al Fraidan, Rahaf Alkuwaity

Research Article

Cite this article: Al Fraidan, A., & Alkuwaity, R. (2025). Corpus-Driven Innovation in Saudi Arabian EFL Teaching Practices. *Educational Process: International Journal, 19*, e2025553. https://doi.org/10.22521/edupij.2025.19.553

Received July 21, 2025
Accepted November 4, 2025

Keywords: Corpus Linguistics, Corpus-based Instruction, COCA, Saudi EFL, Pedagogical Innovation, Language Teaching.

Author for correspondence:
Abdullah Al Fraidan
afridan@kfu.edu.sa
King Faisal University, Alahsa, Saudi Arabia

Abstract

Background/purpose. Traditional language instruction in Saudi EFL classrooms often limits learners' exposure to authentic language, reducing communicative competence. This review article explores how corpus-based approaches, particularly using the Corpus of Contemporary American English (COCA), can bridge this gap and foster pedagogical innovation.

Materials/methods. This study employed a systematic literature review of relevant empirical studies and theoretical works obtained from Google Scholar and academic databases. It focused on identifying the instructional potential of linguistic corpora, with an emphasis on COCA. The review also presents practical lesson scenarios and instructional applications in Saudi educational contexts.

Results. Findings indicate that corpus-based methods enhance learners' understanding of vocabulary, grammar, and discourse markers, particularly when used to teach linking adverbs like therefore and however. COCA's concordance lines and frequency tools offer valuable input for contextualized language learning. Saudi EFL students have shown improved lexical awareness and positive attitudes toward such approaches.

Conclusion. Corpus linguistics, particularly through tools like COCA, holds significant promise for transforming EFL instruction in Saudi Arabia. Despite challenges such as limited training and digital access, strategic integration of corpora can promote more authentic, engaging, and data-driven learning environments for both teachers and students.

1. Introduction

1.1. Corpus Linguistics in Language Teaching

Corpus linguistics has emerged as a significant methodological innovation in language research over the past few decades, closely tied to advances in technology (Youssef & Omar, 2020). To fully understand corpus linguistics, it is beneficial to define it both broadly and narrowly. Broadly, a "corpus" refers to a systematically organized collection of texts. In literary studies, a corpus typically includes texts produced by authors of a specific genre or period. However, within linguistics, a corpus may encompass any systematically gathered linguistic data—even a single sentence—collected for linguistic research purposes (Stefanowitsch, 2020).

The application of corpora has notably expanded into pedagogical contexts, particularly within language teaching and learning (Xodabande & Nazari, 2022). This paper is a review article that synthesizes previous research on corpus linguistics and COCA applications, particularly in the context of Saudi Arabian EFL education. It aims to bridge the gap between theory and practice by offering instructional insights grounded in existing literature.

. There are numerous advantages to implementing corpora in educational contexts; for example, they illuminate common lexical items and collocations and provide extensive authentic language examples (Hunston, 2002; Römer, 2005). Nevertheless, there remains limited awareness—and occasionally resistance—from students, teachers, and materials developers regarding the use of corpora in Saudi classrooms (Misnawati et al., 2025).

Despite the growing body of international research on corpus linguistics, there remains a noticeable gap in synthesizing how such approaches are practically applied within the Saudi EFL educational context. Most existing studies focus on isolated experiments or small-scale implementations, without offering a cohesive overview of COCA's instructional potential, especially for vocabulary and discourse instruction. Moreover, Saudi Arabia's Vision 2030 emphasizes educational innovation and digital transformation—yet corpus-based methods are rarely featured in curriculum planning or teacher training. This review aims to fill this gap by compiling and evaluating empirical findings, proposing practical applications, and contextualizing them for Saudi classrooms.

The present study specifically examines the use of COCA in Saudi EFL instruction. The Corpus of Contemporary American English (COCA) comprises approximately one billion words across nearly 500,000 texts collected from 1990 to 2019. These texts are evenly distributed across various genres, including spoken discourse, fiction, magazines, newspapers, academic journals, blogs, webpages, and TV/movie subtitles. Furthermore, the corpus includes roughly 20 million words annually from each genre, ensuring balanced representation over time (Davies, 2020). Such corpora serve as invaluable resources, providing authentic and abundant language input and represent a transformative pedagogical trend in EFL classrooms (Sinclair, 2004).

Beyond merely examining COCA's potential to improve Saudi EFL instruction, this paper also proposes practical lesson ideas incorporating concordance lines, vocabulary-building activities, and collocation exercises. These strategies aim to help teachers effectively integrate authentic language use into classroom practice.

In the Saudi context, English-language instruction is undergoing a significant transformation, driven by the national Vision 2030 agenda, which emphasizes educational innovation, digital literacy, and global competitiveness. While technologies such as AI and digital platforms are gaining traction, corpus-based instruction remains underutilized despite its proven potential. Many Saudi EFL classrooms still rely heavily on traditional grammar-translation methods, often neglecting authentic language exposure and learner autonomy. Moreover, institutional barriers, such as limited teacher training, a lack of corpus literacy, and infrastructural challenges, continue to impede the adoption of

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

innovative linguistic tools. By situating this review within the realities of Saudi education, the paper underscores the urgent need to integrate corpus linguistics into teacher development, instructional materials design, and classroom practices to achieve more effective, modern language education. These conclusions align with Saudi-based work that connects corpus-informed pedagogy to strategy-based assessment, AI-enabled learning, and immersive environments (Al Fraidan, 2024a, 2024b, 2025; Al Fraidan & Alaliwi, 2024; Al Fraidan & AlMarri, 2024; Al Fraidan & Al Salman, 2023; Al Fraidan & Olaywi, 2024)

2. Literature Review

Corpora can be employed either directly or indirectly in the English classroom—indirectly through curriculum and materials development or directly through classroom-based activities (Römer, 2011).

2.1. Corpora in Vocabulary Instruction

A recent corpus-based study by Boontam and Phoocharoensil (2018) involving 30 Thai learners assessed the effectiveness of corpus-based instructional methods. After six weeks of instruction (one 50-minute class per week), pre- and post-test results demonstrated vocabulary improvements attributed to corpus-based activities (Youssef & Omar, 2020). Similarly, research involving two groups of 24 Saudi EFL female students found that the corpus approach significantly enhanced vocabulary acquisition, with the experimental group outperforming the control group. Despite initial difficulties, which could be mitigated through practice and training, most students expressed favorable attitudes toward corpus-based methods, particularly the use of COCA (Alenizi & Adawi, 2024). Additional studies have consistently shown that corpus-based vocabulary instruction is superior to traditional textbook methods, fostering positive student attitudes (Hussan & Kalim, 2024; Lu & Chen, 2019; Binkai, 2012; Elsherbini & Ali, 2017; Ashkan & Seyyedrezaei, 2016; Rizvić-Eminović & Neslanović, 2022).

2.2. Corpora in Grammar Instruction

Grammar instruction, especially for academic writing among ESL/EFL learners, presents considerable challenges for language teachers (Alamri, 2022). Vannestål and Lindquist (2007) found that students generally favored corpus-based grammar instruction over traditional textbook approaches. Gaskell and Cobb's (2004) investigation yielded intriguing results: after students received feedback on their writing, using a concordancer significantly improved their grammatical accuracy. Similarly, Boulton (2009) demonstrated that corpora effectively support learning linking adverbials in English, outperforming traditional teaching methods, particularly for lower-level students.

Moreover, a study involving Iranian EFL learners found significant improvements in grammar among students using concordancers compared to those using a thesaurus-based approach (Hanafiyeh & Keshi, 2013). In Saudi Arabia, grammar instruction has traditionally followed a teacher-centered, grammar-translation method. Sofi (2015), however, advocates for a communicative, student-centered approach integrating grammatical input dynamically.

Incorporating corpus-based and contextualized lexico-grammatical methods has been shown to enhance vocabulary acquisition, critical thinking, and exploratory learning skills in ESL and EFL contexts (Liu & Jiang, 2009). An investigation by Girgin (2011) at Erciyes University demonstrated that lower-level students effectively used corpus-based activities to learn grammar, yielding outcomes comparable to those of traditional textbook instruction and showing more positive attitudes. Additionally, Kayaoglu's (2013) study indicated that even brief training with COCA enabled students to better distinguish between synonyms, underscoring the efficacy of corpora in grammatical instruction.

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

2.3. Corpora in Discourse Instruction

Over the past two decades, corpus linguistics has been increasingly applied to discourse marker analysis (Alsharif, 2017; Altenberg & Tapper, 2014; Min, 2011). For instance, Altenberg and Tapper (2014) compared conjunct usage in essays written by advanced Swedish learners and native English speakers, revealing lower conjunct frequency among the former. Jaihow and Sinwongsuwat's (2024) research similarly demonstrated the benefits of corpus-based techniques in improving pragmatic competence in spoken discourse. Burns (2016) emphasized the importance of developing corpora of spoken English to help learners meet authentic communication needs, thereby enhancing speaking skills through exposure to natural linguistic features.

2.4. Corpus-Based Instruction in Saudi EFL Classrooms

Recently, corpus-based instruction has gained notable attention in Saudi EFL classrooms. Educators and researchers are actively exploring how authentic linguistic data can strengthen vocabulary and grammar skills, ultimately enhancing academic writing. A study at the University of Tabuk demonstrated the positive impact of a Data-Driven Learning (DDL) approach on grammar instruction for preparatory-year Saudi students. Unique to this study was the implementation of a four-stage training project for teachers, culminating in self-created, corpus-informed instructional materials (Zohairy, 2013).

Similarly, research conducted at the University of Jeddah highlighted the benefits of integrating corpus-based approaches into academic writing instruction, significantly improving students' lexical and grammatical integration, independence, and confidence (Al-Qahtani, 2021). Alharbi's (2012) investigation at King Saud University further confirmed that direct corpus usage by students not only improved their linguistic accuracy but also heightened their problem-solving awareness and skills related to grammatical and lexical challenges.

3. Methodology

This article employs a narrative literature review methodology, focusing on peer-reviewed studies on corpus linguistics and its applications in English language teaching, particularly in Saudi and regional EFL contexts. A comprehensive search was conducted across academic databases including Google Scholar, Scopus, ERIC, and Web of Science, using combinations of the following keywords: *"corpus linguistics," "COCA," "data-driven learning," "Saudi EFL," "vocabulary," "grammar," "discourse markers,"* and *"language teaching."*

Studies published between 2000 and 2025 were prioritized to reflect both foundational and contemporary perspectives. Selection criteria included empirical or conceptual relevance to the Saudi EFL context, application of corpus tools (especially COCA), and pedagogical significance. Articles lacking practical applications or focused solely on theoretical linguistics were excluded.

The final corpus consisted of 34 sources, including journal articles, book chapters, and thesis-based case studies. Studies were categorized thematically under vocabulary, grammar, and discourse instruction. Additional Saudi-based studies were reviewed separately to identify regional implementation patterns and challenges. This review synthesizes the findings across these categories and draws practical instructional recommendations for EFL practitioners in Saudi Arabia.

4. Practical Applications of the Corpus of Contemporary American English (COCA)

The Corpus of Contemporary American English (COCA) is one of the largest corpora of American English, with over one billion words from various sources. The COCA interface offers filters and functions that address several categories of linguistic traits, allowing users to search for words in certain contexts. Depending on the user's choices, the frequency of a word or phrase can also be displayed in a variety of ways, including lists or charts. These are just a handful of the interface options

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

offered by COCA; when combined and layered, its capabilities can yield significant outcomes for English teachers (Veloso, n.d.).

This paper includes a step-by-step guide to using COCA. The first step is registering. Visit https://www.english-corpora.org/coca/ to access the corpus within a browser on the internet. To access the "LOGIN" page, click the yellow ID icon in the upper right corner. Locate and click the gray "REGISTER" button beneath the "Log in" icon. This will direct users to a page that resembles the one shown in Figure 1.

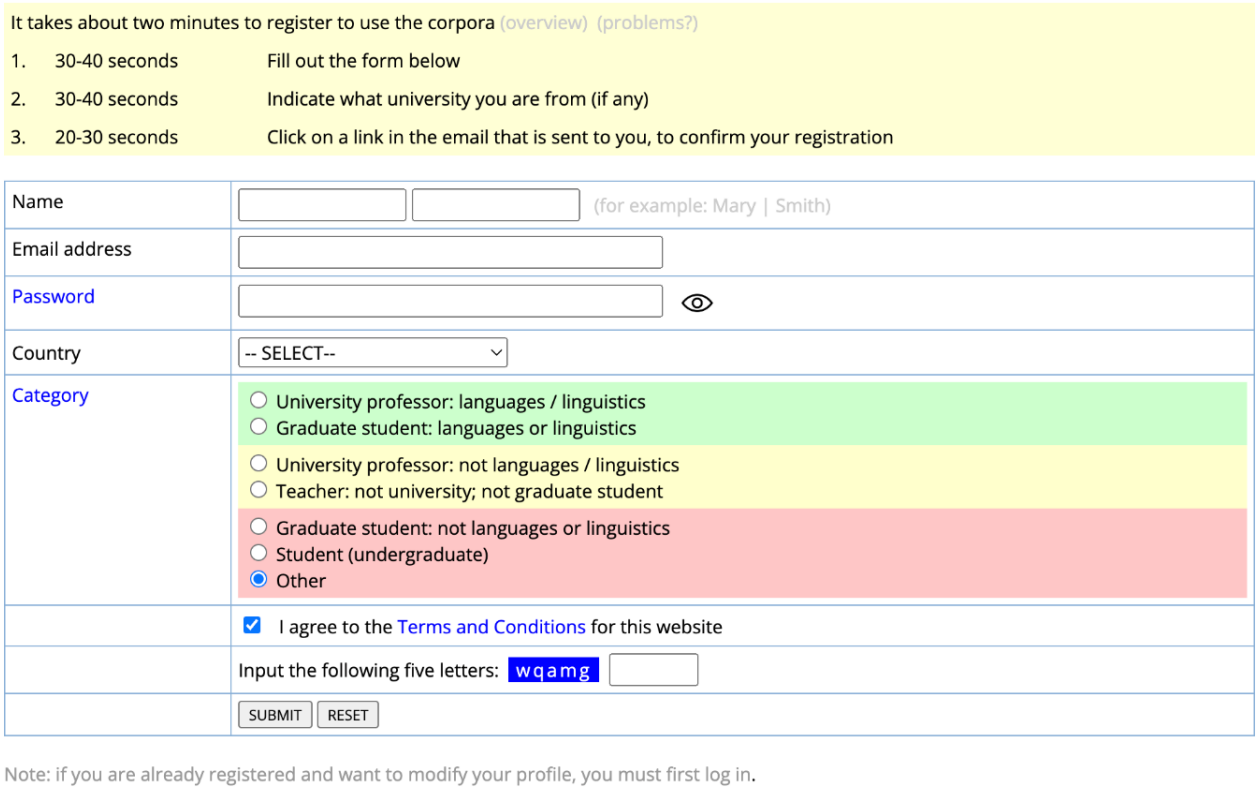


Figure 1. Registering

To create an account on the English Corpora platform, users are required to provide their full name, email address, password, country of residence, and a category that reflects their professional or academic status. After agreeing to the Terms and Conditions, users must click "Submit" to proceed to the confirmation stage. A verification email will then be sent from admin@english-corpora.org, containing a link that must be followed to activate the account.

As a registered user logged into the website, your homepage will land on the search tab as illustrated in Figure 2.

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

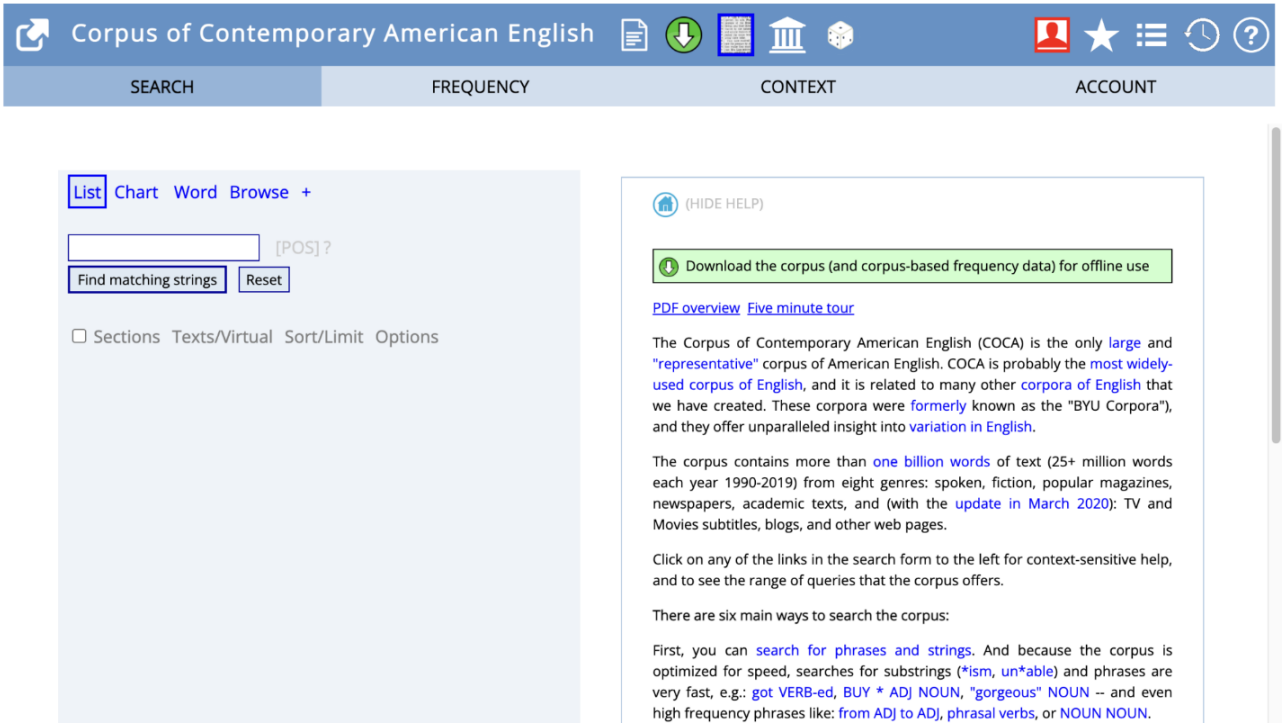


Figure 2. COCA Home Page

The top right corner icons are additional resources for the corpus; Figure 3 demonstrates these icons' names.

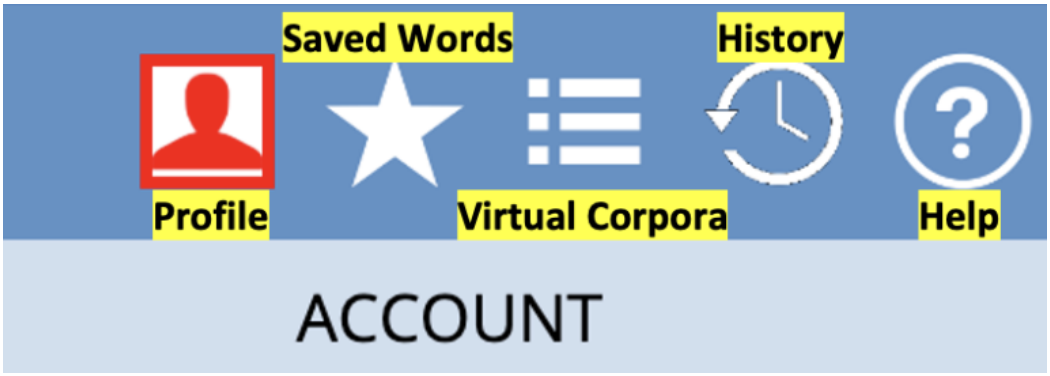


Figure 3. Top Right Corner Icons

The Profile icon allows users to view their account information, make changes, and view their statistical usage of the corpus. The Saved Words icon directs users to saved words in previous searches. The Virtual Corpora icon enables users to create their own corpora from COCA entries. The History icon displays recent searches, and the Help icon explains each icon's functions. At the top left corner of the screen are the main tabs of the website: SEARCH, FREQUENCY, CONTEXT, and ACCOUNT. These tab options will shift and change depending on the website function. When a word, phrase, or string is searched using the SEARCH tab, users can evaluate the frequency of their input within the corpora against a variety of parameters by selecting the FREQUENCY tab. The CONTEXT tab displays a search's concordance lines from the corpus. Users can view words related to their searches and analyze their use depending on the information supplied. The ACCOUNT tab contains a brief description of the corpus and its general functionalities. Users can use this page to learn about the many functions available on the corpus website. To search for a word or phrase within the corpus, the users type the word, phrase or string into the textbox in the 'List' on the top left of the home page, and then they click the "Find matching strings" button below. This step is depicted in Figure 4.

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

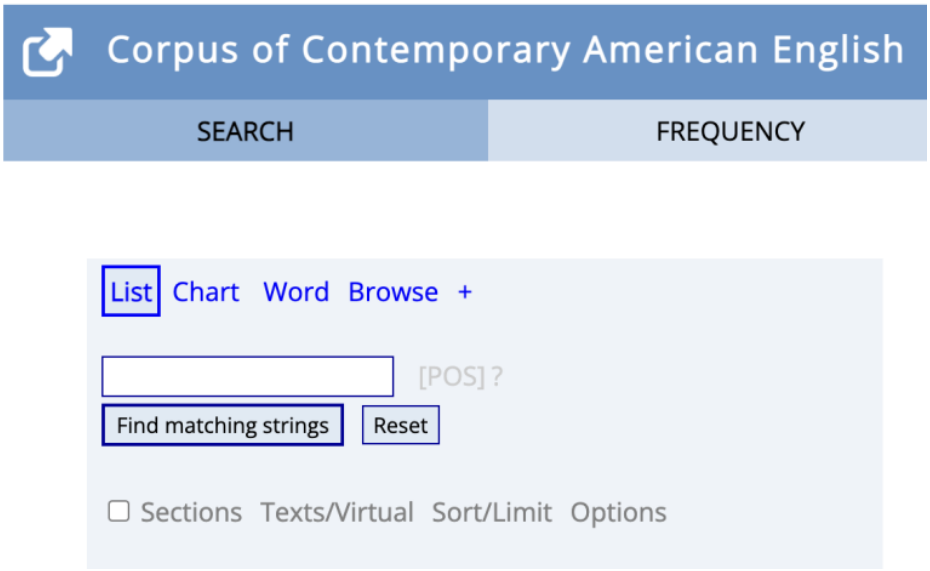


Figure 4. List

The search will redirect the users to the results on the "FREQUENCY" tab, where the other list options can be found. Check Figure 5 to fully grasp this particular step.

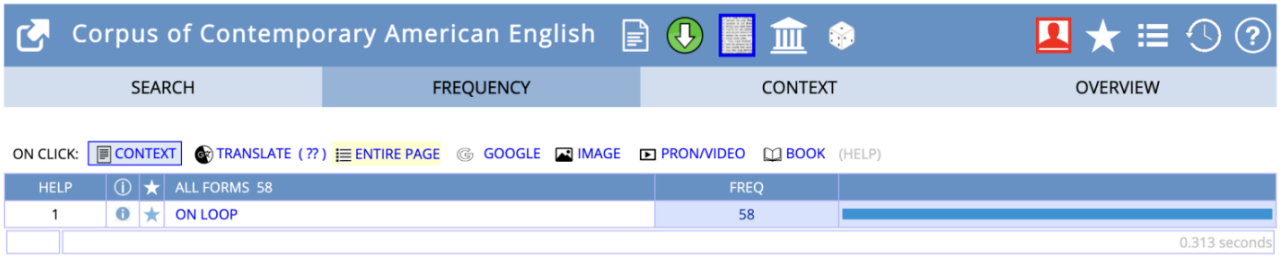


Figure 5. List Options

The interface displays all relevant forms of the queried word or phrase in the "ALL FORMS" column, as well as their related frequencies in the "FREQ" column. Users can save specific word forms by clicking the star symbol next to each result and adding them to their "Saved Words" list (as explained in Section 4.1). Furthermore, selecting a listed word form (e.g., "ON LOOP") takes the user to the "CONTEXT" page, where the item is displayed in its original usage contexts within the corpus. See figure 6.

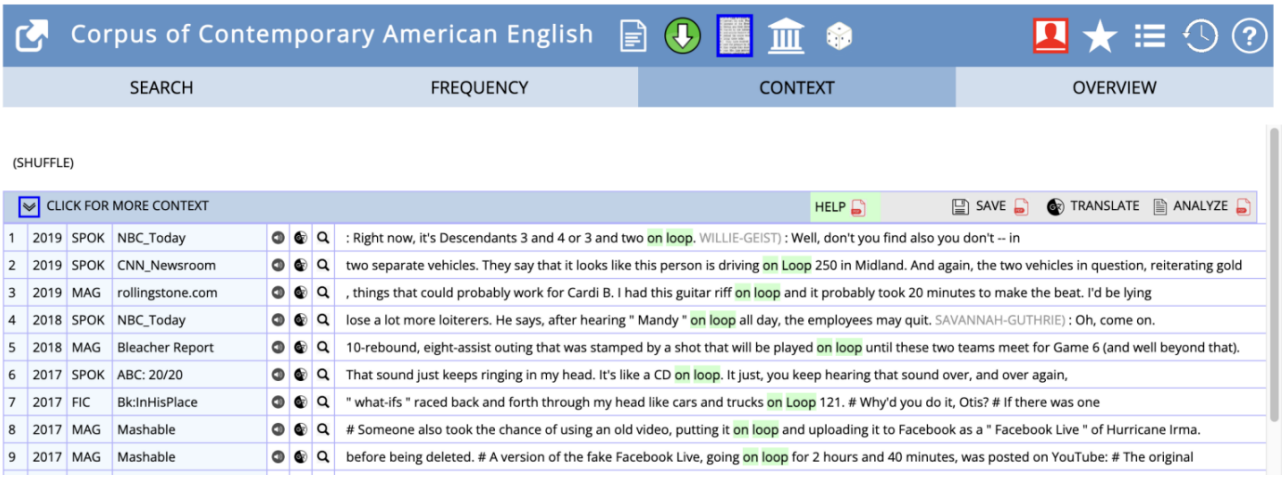


Figure 6. Context Page.

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

Coca enables its users to display their search results in several sentences. These are called concordance lines. The Context page in Figure 6 also shows the date of the source, its type (magazine, spoken, fiction, etc.), and the source name. To restart the search procedure on all tabs, users should return to the "SEARCH" tab and click the "Reset" button below the text box.

Following the aforementioned procedure to search for certain words can be hectic. Thus, the Coca search bar is equipped with a couple of features that help in narrowing down the searches. An example of those features is wildcard search, which allows users to see different results by utilizing part of a word or morpheme. The way to do is by using the "*" wildcard. For example, if we attach "*" directly to the end of "be," we receive words from COCA that start with the word "be," as shown in Figure 7.

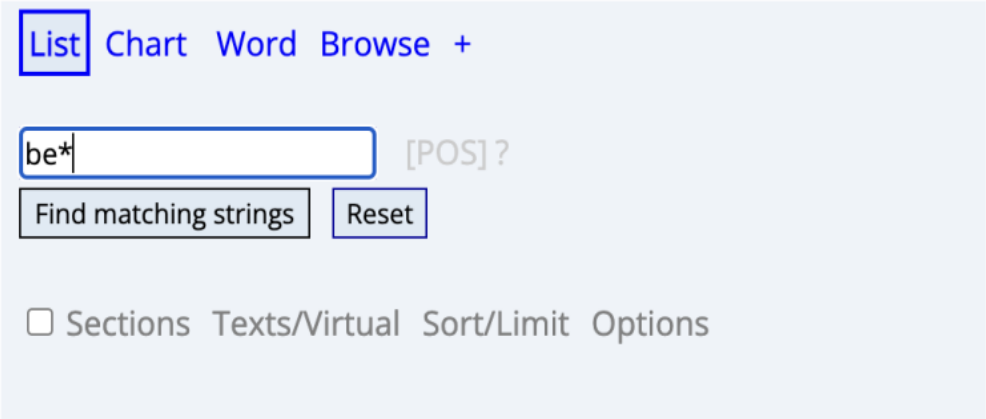


Figure 7. Wildcard Search.

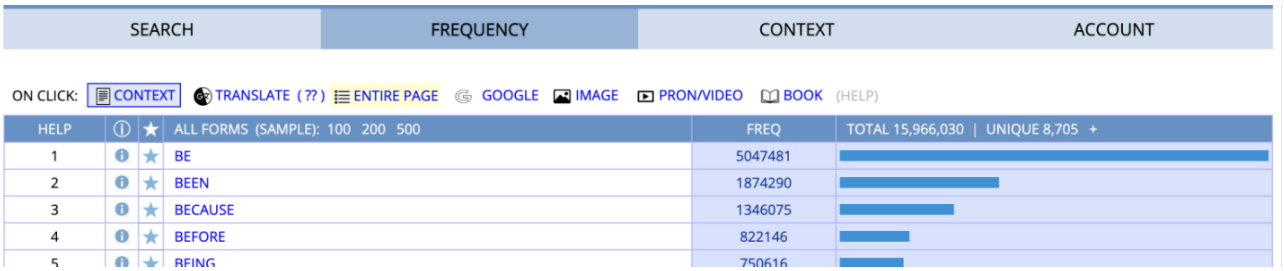


Figure 8. Examples of Wildcard Search.

Furthermore, when separating the "*" with "be," the results will be of phrases that start with the word "be," as displayed in Figure 9.

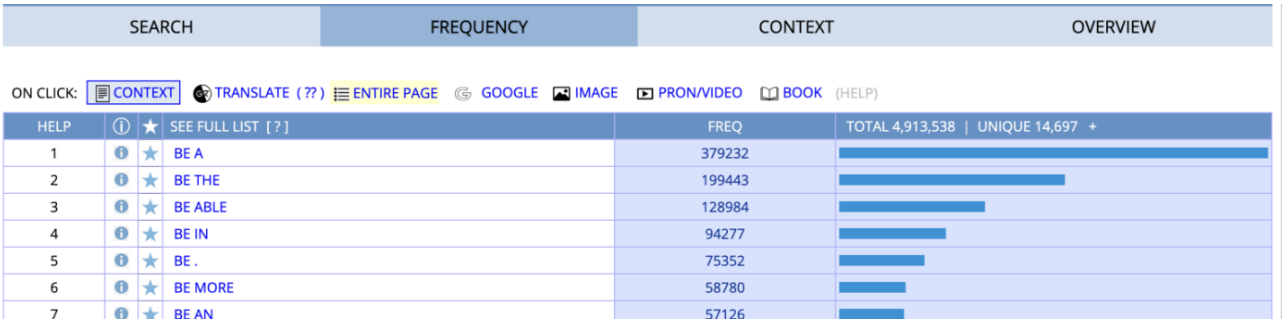


Figure 9. Search with Space Before Wildcard.

An asterisk (*) can be used before or after a search phrase to obtain word variants that frequently appear with that prefix or suffix. For example, if a teacher wants to find terms that begin with the prefix "im," they can do a search for "im" to get relevant results.

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

4.1. Identifying Frequent Academic Words

Following the previously described steps, once users have logged into COCA, they can utilize the "List" or "Frequency" tabs to explore frequently used words within a specific genre. For instance, selecting the "academic" genre can help highlight vocabulary particularly relevant to academic contexts. COCA enables users to view the most frequent words along with their parts of speech. Results are displayed as a ranked list according to frequency within academic texts. Such lists are beneficial for identifying target vocabulary for instructional or research purposes. Additionally, COCA allows users to save specific words by marking them with a star or to export lists for use in classroom activities or curriculum development.

4.2. Vocabulary-Building Activities

According to Krashen (1981), "The best methods [for language acquisition] are those that supply 'comprehensible input' in low-anxiety situations." Consequently, it is crucial to select vocabulary appropriate for students' proficiency levels to ensure optimal learning progress in second-language acquisition.

COCA serves as a powerful tool for creating effective vocabulary activities by identifying and sorting the most frequently used words. Vocabulary selection should align with the curriculum and the learners' proficiency level. Providing students with engaging reading materials that match their language competence is essential to sustaining motivation and maximizing learning potential. Tools such as Text Inspector can be employed to evaluate the complexity of reading texts.

Students should be encouraged to read without frequent pauses for dictionary consultation; instead, they should use morphological and contextual clues to infer meanings, mirroring natural first-language reading practices to foster fluency. Active reading strategies, such as annotating texts by underlining unfamiliar words and making marginal notes, help students maintain engagement. After reading, students can revisit unknown vocabulary using dictionaries as needed.

At this stage, COCA becomes especially valuable for examining how frequently and contextually new vocabulary is used. For instance, students can use COCA to prioritize learning common words like *surprised* over less frequent alternatives like *flabbergasted*. Students may also record selected vocabulary in personalized notebooks, including definitions, example sentences, collocations, and usage notes derived directly from COCA searches. Regular vocabulary reviews and self-assessment through these notebooks support long-term retention and lexical development.

4.3. Using COCA for Discourse Marker Exercises

Saudi EFL learners often encounter difficulties mastering discourse markers in academic writing (Alsharif, 2017). Therefore, employing COCA to design exercises focused on discourse markers can effectively aid students in identifying, analyzing, and using these markers appropriately.

A practical application involves creating a class-specific corpus by compiling students' argumentative essays into a single file. The instructor then uses COCA's concordance tool to analyze this corpus, identifying the frequency and types of discourse markers utilized by students. Following this analysis, students categorize the identified discourse markers into functional groups such as additive (e.g., *also*, *in addition*) or adversative (e.g., *however*, *on the other hand*).

Next, inspired by Alsharif's (2017) methodology, students can compare the frequency and variety of discourse markers from their class corpus with those typically found in native English-speaker writing. This comparison facilitates a classroom discussion highlighting any instances of overuse, underuse, or incorrect usage. Finally, students are encouraged to revise their essays, intentionally focusing on proper and varied discourse marker usage to improve coherence, cohesion, and overall writing quality.

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

The figures provided below illustrate how to verify the frequency of common discourse markers such as *also*, *however*, and *therefore* using COCA.

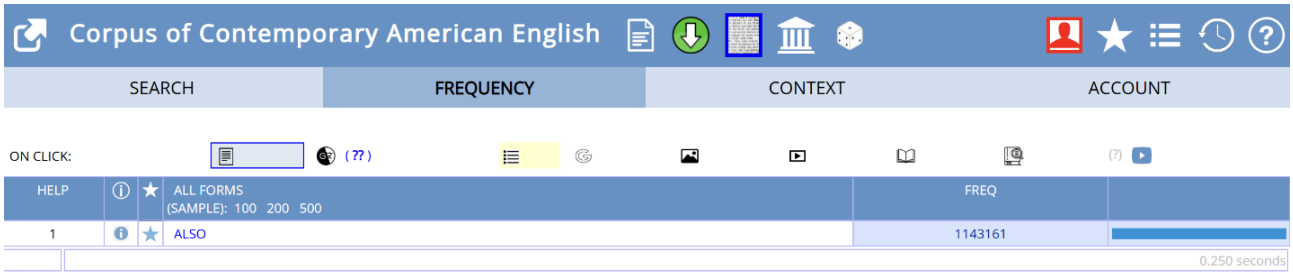


Figure 10. The Frequency of 'Also'

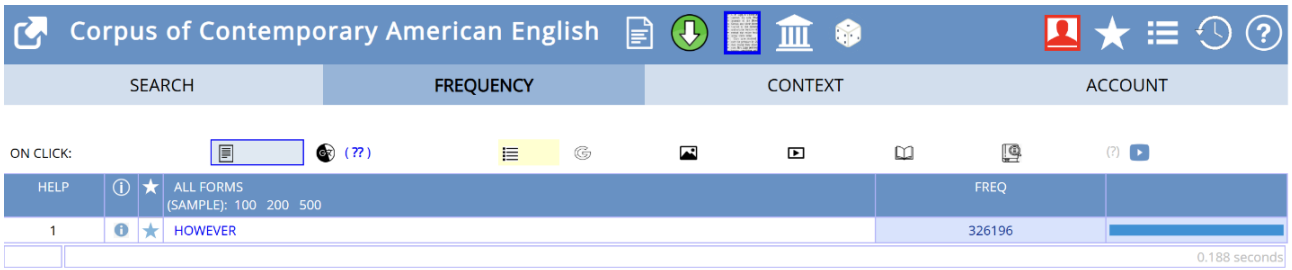
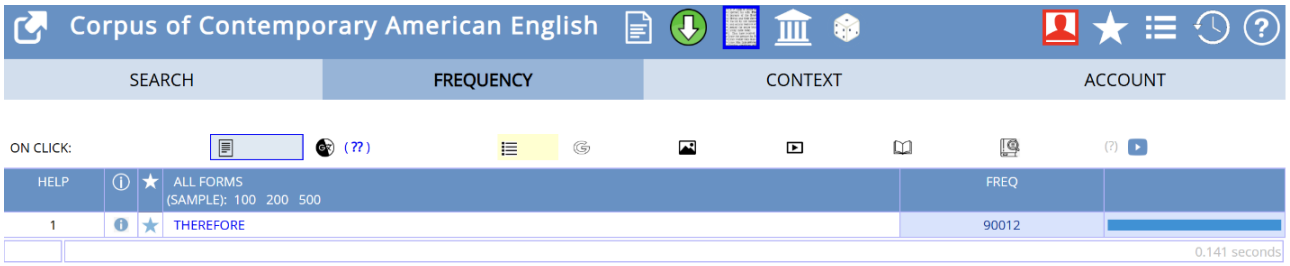


Figure 11. The Frequency of 'However'.



Figured 12. The Frequency of 'Therefore'

Since this exercise is purely hypothetical, Table 1 presents comparative data from Alsharif's (2017) study, demonstrating the use of discourse markers by Saudi EFL students and native speakers, which resembles one of the steps in the exercise. SNNS refers to Saudi Non-Native Speakers, and NS refers to native speakers.

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

Table 1. The Top Ten Frequently Used Discourse Markers by Both Groups Saudi Non-Native Speakers and Native Speakers

SNNS				NS		
raw %	per 10,000	words		raw %	per 10,000	words
1. so	123	18.5	24.7	9	4.4	5.4
2. also	102	15.3	20.5	53	26.2	32.1
3. however	39	5.8	7.8	30	14.8	18.1
4. in addition	29	4.3	5.8	-	-	-
5. for example	27	4	5.4	15	7.4	9
6. as a result	25	3.7	5	-	-	-
7. then	24	3.6	4.8	8	3.9	4.8
8.moreover	21	3.1	4.2	-	-	-
9.therefore	18	2.7	3.6	8	3.9	4.8
10. finally	18	2.7	3.6	-	-	-
-	-	-	-	15	7.4	9 (now)
-	-	-	-	5	2.4	3 (of course)
-	-	-	-	4	1.9	2.4 (yet)
-	-	-	-	4	1.9	2.4 (rather)
-	-	-	-	4	1.9	2.4 (still)

4.4. A Lesson Plan Scenario in a Saudi EFL Classroom

The previously described vocabulary-building activity and discourse marker exercises can be effectively integrated into a structured lesson plan. Before beginning the lesson activities, students should be clearly informed about the lesson's purpose and objectives. Specifically, the lesson aims to assist students in identifying, comprehending, and accurately employing logical connectors (e.g., *however*, *therefore*) in academic writing through authentic usage examples drawn from the Corpus of Contemporary American English (COCA).

Conducting this type of lesson requires computers and reliable internet access, enabling students to register and access COCA. A worksheet containing guided activities will facilitate the smooth delivery of instructions. Initially, the teacher should begin with a brief discussion on the meanings and usages of the logical connectors *however* and *therefore*, which serves as a warm-up to prepare students for the lesson. Next, the teacher should demonstrate how to navigate and use COCA, dedicating sufficient time to ensure all students are comfortable with the interface and understand the task clearly.

The primary activity involves students completing a worksheet provided by the teacher, featuring four to five short, incomplete academic sentences. For instance, sentences might include:

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

- "The data indicate a clear trend; _____, additional analysis is necessary."
- "Existing theories offer one explanation; _____, our findings suggest an alternative interpretation."

Students must accurately complete each sentence with either *however* or *therefore*, subsequently justifying their choice. Such tasks guide students toward understanding the connectors' functions and appropriate positioning within sentences (e.g., sentence-initial, mid-sentence following a semicolon). Before finalizing their answers, students can verify the correct usage of these connectors by consulting examples from COCA.

The lesson can conclude with the previously described vocabulary-building activity or alternatively with a homework assignment requiring students to produce a short text incorporating each connector. Students would subsequently verify their connector usage by comparing their sentences to authentic examples in COCA. By the end of this instructional sequence, students will be better equipped to use logical connectors naturally and accurately in academic writing, benefiting from authentic linguistic examples provided by COCA.

5. Discussion

This study examined the use of corpora (specifically COCA) in language teaching and found both teachers and students exhibited positive attitudes toward incorporating corpus-based approaches into language instruction (Xodabande & Nazari, 2022). Römer (2011) highlights how corpus linguistics has transformed language instruction through innovations such as the development of dictionaries (Mayor, 2009), reference grammars (Biber et al., 1999), coursebooks (Lee & Swales, 2006; McCarthy et al., 2005), and data-driven learning (DDL) methodologies, which allow direct student access to corpora through concordancers (Vyatkina, 2016).

Despite these evident benefits, the adoption of corpus-based technologies in classrooms remains challenging for many educators (Angeli, 2005). Language instructors need sufficient proficiency in technological tools and resources available to them. Breyer (2009) emphasizes that integrating corpora into language teaching is ultimately the teacher's decision, underscoring teachers' central role in bringing corpus-informed practices to classroom settings. Although freely accessible corpora exist, their potential in language teaching and learning contexts has not yet been fully utilized (Römer, 2011; Zareva, 2016). Additionally, educators often report environmental constraints, limited time, unreliable internet access, classroom management challenges, and students' limited access to mobile devices as significant barriers hindering the practical use of corpus resources (Xodabande & Nazari, 2022).

Considering these obstacles, the current study offers practical recommendations to enable effective implementation of corpus-based approaches within EFL classrooms.

The findings from this review confirm and consolidate the growing consensus in the literature that corpus-based instruction significantly enhances learners' linguistic competence—particularly in vocabulary acquisition, grammatical awareness, and discourse coherence. Across the studies reviewed, COCA emerged as a versatile tool capable of supporting both receptive and productive skills development. These results align with international findings (e.g., Boulton, 2009; Gaskell & Cobb, 2004), while reinforcing region-specific studies from Saudi Arabia (e.g., Alenizi & Adawi, 2024; Al-Qahtani, 2021).

However, what distinguishes this review is its dual focus: (1) providing a synthesized framework of evidence-based corpus applications, and (2) contextualizing them within the Saudi educational system, addressing both national reform goals and classroom-level constraints. The paper also presents a visual, step-by-step COCA tutorial and an integrated lesson plan, offering EFL instructors

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press

practical tools to overcome implementation challenges. In this way, the review contributes not only to theoretical understanding but also to pedagogical innovation by transforming research into action.

5.1. Practical Recommendations for Teachers

Teachers should actively raise student awareness about effective corpus usage. For instance, as demonstrated in Alsharif's (2017) study, instructors integrated corpus-based activities into writing assignments, helping students identify common patterns in the use of logical connectors. Additionally, EFL teachers are encouraged to develop hands-on tasks where students independently investigate word usage, collocations, and frequency through COCA, thereby fostering learner autonomy and active engagement.

As illustrated by the previously described lesson scenario, integrating corpus-based activities can transform routine lessons into stimulating, innovative learning experiences. Incorporating such data-driven pedagogical methods within traditional Saudi EFL curricula enables teachers to offer more effective, engaging, and authentic language instruction.

6. Conclusion

Through a systematic review of relevant literature and an exploration of the Corpus of Contemporary American English (COCA), this study examined the impact of integrating corpus-based activities into Saudi EFL classrooms. Findings indicate that corpus linguistics provides numerous instructional advantages for both educators and students. Specifically, COCA's accessible and versatile search options significantly enhance practical classroom applications, creating a more dynamic and authentic learning environment.

Although initial barriers—such as unfamiliarity with corpus tools and insufficient training—may exist, these can be addressed through dedicated professional development, tailored educator training, practical experience, and gradual integration into classroom routines. Finally, corpora should be viewed not merely as supplementary teaching resources, but as vital tools for enhancing teacher autonomy, instructional effectiveness, and professional growth—particularly among non-native English-speaking instructors (Xodabande & Nazari, 2022).

Declarations

Conflicts of Interest. The author declares no conflict of interest.
Funding. This work was funded and supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, [Grant, KFU253536].
Data Availability. Data is available upon request.
Acknowledgments. I would like to acknowledge all the people and family members who facilitated this project including administrators for their cooperation. Special acknowledgments to AlAnoud Alwasmi who facilitated a lot of this research processes and also for Hiam Al Fraidan.

References

Alamri, B. (2022). The role of corpus linguistics in grammar instruction: A review of literature. *International Journal of Linguistics, 14*(6), 158. https://doi.org/10.5296/ijl.v14i6.20500

Alenizi, A., & Adawi, R. (2024). Investigating the effectiveness of using corpus-based developed materials in vocabulary learning for Saudi EFL students. *Forum for Linguistic Studies, 6*(3), 721–745. https://doi.org/10.30564/fls.v6i3.6551

Al Fraidan, A. A. (2025). The enhanced adaptive PPP model: A novel framework for revolutionizing test-taking strategies in language assessment. *Forum for Linguistic Studies, 7*(1), 298–312. https://doi.org/10.30564/fls.v7i1.7918

Al Fraidan, A. A. (2024b). Beyond the bubble: Unveiling the multifaceted landscape of test wiseness and their operationalization among English-language majors. *Theory and Practice in Language Studies, 14*(6), 1735–1744. https://doi.org/10.17507/tpls.1406.14

Al Fraidan, A. (2024a). Unveiling students test strategizing through the lens of Monitor Theory: Teaching insights. *Forum for Linguistic Studies, 6*(2), 1181. https://doi.org/10.59400/fls.v6i2.1181

Al Fraidan, A., & Alaliwi, M. (2024). Digital transformation for sustainable English language learning: Insights from Saudi Arabia and global perspectives. *Forum for Linguistic Studies, 6*(6), 439–449. https://doi.org/10.30564/fls.v6i6.7754

Al Fraidan, A., & AlMarri, N. (2024). Repetition, difficulty, and AI: Enhancing vocabulary acquisition for second language learners. *African Journal of Biomedical Research, 27*(4S), 8088–8095. https://doi.org/10.53555/AJBR.v27i4S.5152

Al Fraidan, A., & Al Salman, A. (2023). The utilisation of test-taking strategies by Saudi EFL high schoolers in Al-Ahsa region. *World Journal of English Language, 13*(2), 414–429. https://doi.org/10.5430/wjel.v13n2p414

Al Fraidan, A., & Olaywi, M. (2024). Scenarios to implement Metaverse into the Saudi educational system. *Forum for Linguistic Studies, 6*(4), 180–193. https://doi.org/10.30564/fls.v6i4.6853

Alharbi, W. H. (2012). Learners' interaction with online applications: Tracking language-related episodes in computer-assisted L2 writing. *Frontiers of Language and Teaching, 3*, 96–107.

Al-Qahtani, A. A. (2021). The integration of corpus-based approach with EFL academic writing in the Saudi context. *International Journal of English Language and Literature Studies, 10*(1), 22–45. https://doi.org/10.18488/journal.23.2021.101.22.45

Alsharif, M. (2017). The frequently used discourse markers by Saudi EFL learners. *Arab World English Journal, 8*(2). https://dx.doi.org/10.24093/awej/vol8no2.28

Altenberg, B., & Tapper, M. (2014). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *Learner English on computer* (pp. 80–93). Routledge. https://doi.org/10.4324/9781315841342-6

Angeli, C. (2005). Transforming a teacher education method course through technology: Effects on preservice teachers' technology competency. *Computers & Education, 45*(4), 383–398. https://doi.org/10.1016/j.compedu.2004.06.002

Ashkan, L., & Seyyedrezaei, S. H. (2016). The effect of corpus-based language teaching on Iranian EFL learners' vocabulary learning and retention. *International Journal of English Linguistics, 6*(4), 190–196. https://doi.org/10.5539/ijel.v6n4p190

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Longman. https://doi.org/10.5539/elt.v5n4p131

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

Boontam, P., & Phoocharoensil, S. (2018). Effectiveness of English preposition learning through data-driven learning (DDL). *3L: Language, Linguistics, Literature, 24*(3), [insert page numbers]. [Insert DOI or URL] https://doi.org/10.17576/3l-2018-2403-10

Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL, 21*(1), 37–54. https://doi.org/10.1017/s0958344009000068

Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning.* https://doi.org/10.1080/09588220902778328

Burns, A. (2016). Research and the teaching of speaking in the second language classroom. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 242–256). Routledge. https://doi.org/10.4324/9781315716893-18

Davies, M. (2020). The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/

Elsherbini, S. A. H., & Ali, A. D. (2017). The effects of corpus-based activities on EFL university students' grammar and vocabulary and their attitudes toward corpus. *Journal of Research in Curriculum Instruction and Educational Technology, 3*(1), 133–161. https://jrciet.journals.ekb.eg/article_24482.html

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*(3), 301–319. https://doi.org/10.1016/j.system.2004.04.001

Girgin, U. (2011). Corpus-based activities at lower levels of EFL proficiency: The effectiveness of using concordance lines on grammar learning [Master's thesis, Bilkent University].

Hanafiyeh, M., & Keshi, A. K. (2013). The effect of corpus-based instruction and thesaurus-based teaching on Iranians EFL learners' grammatical knowledge. *Journal of Basic and Applied Scientific Research, 3*(2), 167–179.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press. https://doi.org/10.1017/cbo9781139524773

Hussan, S., & Kalim, S. (2024). The efficacy of corpus-based materials in enhancing vocabulary acquisition among Pakistani EFL learners. *Qlantic Journal of Social Sciences and Humanities, 5*(4), 11–27. https://doi.org/10.55737/qjssh.139943559

Incelli, E. (2025). Exploring the future of corpus linguistics: Innovations in AI and social impact. *International Journal of Mass Communication, 3*, 1–10. https://doi.org/10.6000/2818-3401.2025.03.01

Jaihow, S., & Sinwongsuwat, K. (2024). Integrating corpus-based approaches in teaching English conversation: A study on enhancing pragmatic competence and conversational skills. *World Journal of English Language, 14*(2), 57–69. https://files.eric.ed.gov/fulltext/EJ1415583.pdf

Krashen, S. (1981). *Second language acquisition*. Second Language Learning.

Kyaoglu, M. N. (2013). The use of corpus for close synonyms. *The Journal of Language and Linguistics Studies, 9*(1), 128–144.

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. https://doi.org/10.1016/j.esp.2005.02.010

Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal, 93*(1), 61–78. https://doi.org/10.1111/j.1540-4781.2009.00828.x

Lu, X., & Chen, B. (Eds.). (2019). *Computational and corpus approaches to Chinese language learning*. Springer. https://doi.org/10.1007/978-981-13-3570-9

Mayor, M. (2009). *Longman Dictionary of Contemporary English*. Pearson Longman.

McCarthy, M. J., McCarten, J., & Sandiford, H. (2005). *Touchstone Level 1: Student's book with audio CD/CD-ROM*. Cambridge University Press.

Min, S. (2011). A corpus-based analysis of EFL learners' use of discourse markers in cross-cultural communication. *English Language & Literature Teaching, 16*(4), 177–194.

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). https://doi.org/10.46743/2160-3715/2025.6867

Rizvić-Eminović, E., & Neslanović, V. (2022). The effect of corpus-based activities on vocabulary acquisition of A1–A2 level seventh graders: A case study. *European Journal of English Language Teaching, 7*(4), 1–14. https://oapub.org/edu/index.php/ejel/article/view/4369

Römer, U. (2005). *Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. John Benjamins. https://doi.org/10.1075/scl.18

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics, 31*, 205–225. https://doi.org/10.1017/S0267190511000055

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Sofi, L. (2015). Teaching English in Saudi Arabia through the use of multimedia [Master's thesis, University of San Francisco]. https://repository.usfca.edu/capstone/138/

Stefanowitsch, A. (2020). *Corpus linguistics*. Language Science Press. https://langsci-press.org/catalog/book/269

Veloso, I. (n.d.). *COCA tutorial*. Montclair State University. https://pressbooks.pub/testbook123/chapter/coca-tutorial/

Vannestål, M. E., & Lindquist, H. (2007). Learning English grammar with a corpus: Experimenting with concordancing in a university grammar course. *ReCALL, 19*(3), 329–350. https://doi.org/10.1017/s0958344007000638

Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology, 20*(3), 159–179. https://doi.org/10.10125/44487

Xodabande, I., & Nazari, M. (2022). Impacts of a corpus linguistics course on in-service EFL teachers' corpus literacy. *Computer-Assisted Language Learning Electronic Journal, 23*(1), 318–346.

Youssef, A. F. F. A., & Omar, M. A. A. (2020). The effectiveness of corpus-based approach on vocabulary learning gains and retention in Saudi tertiary EFL context. *Journal of English Language Teaching and Applied Linguistics, 2*(3), 1–15.

Zareva, A. (2016). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal, 71*(1), 69–79. https://doi.org/10.1093/elt/ccw045

Zohairy, S. (2013). *Corpus linguistics applications, data-driven learning approach, and elementary students' achievement and motivation* [Master's thesis, University of Tabuk]. https://www.academia.edu/38262696

About the Contributor(s)

Prof. Abdullah Al Fraidan, King Faisal University, Alahsa, Saudi Arabia
Email: afridan@kfu.edu.sa
ORCID: https://orcid.org/0000-0002-5152-9794

Rahaf Alkuwaity, King Faisal University, Alahsa, Saudi Arabia
Email: rahafalkuwaiti@hotmail.com
ORCID: https://orcid.org/0009-0000-5042-2509

https://doi.org/10.22521/edupij.2025.19.553 Published online by Universitepark Press