Innovations in Language Education and Literature
Vol. 2, No. 1, 2025, pp. 8-20
ISSN: 3062-8962 (Print), 3062-8946 (Online)
DOI: https://doi.org/10.31605/ilere.v2i1.5033

Teaching Collocations and Academic Phrases with Corpus Linguistics: Applications for Specific and Academic Contexts

Misnawati Misnawati[1*], Saidna Zulfiqar Bin Tahir[2], Alwi Sibali[3], Widya Pertiwi Anwar[4], Nurfajriah Basri[5], Hikmanisa Bahtiar[6]

[1]*Universitas Cahaya Prima, Indonesia*
[2]*Universitas Iqra Buru, Indonesia*
[3]*Politeknik Maritim AMI Makassar, Indonesia*
[4]*Universitas Negeri Makassar, Indonesia*
[5]*Universitas Sawerigading Makassar, Indonesia*
[6]*Universitas Indonesia Timur, Indonesia*

**Corresponding Author: misnawati_amir@yahoo.com*

ARTICLE INFO

Received: 2025-05-02
Revised: 2025-06-10
Accepted: 2025-06-25

Keywords: Academic Collocations; COCA; Corpus Linguistics; Data-driven Learning; ESP Instruction

ABSTRACT

This study explores the integration of the Corpus of Contemporary American English (COCA) in teaching academic collocations and phrases in English for Specific and Academic Purposes (ESP/EAP) settings. Conducted with 15 first-semester undergraduate students in Indonesia, the six-week classroom-based action research employed corpus-based instructional tasks to develop students' lexical awareness and academic language proficiency. Data were collected through classroom observations, student writings, questionnaires, and semi-structured interviews. The findings show notable improvement in students' use of academic collocations and discipline-specific phrases, reflected in increased accuracy and fluency in both writing and speaking. Learners reported that COCA supported them in identifying more natural and context-appropriate word combinations, boosting their confidence and academic expression. Initial challenges with the COCA interface and corpus navigation were mitigated through instructional scaffolding. Over time, students demonstrated increased autonomy and began using COCA for independent and collaborative learning. The study affirms the pedagogical value of corpus linguistics for novice ESP learners and highlights the importance of technical and teacher support in integrating corpus tools into curriculum design.

Citation (APA): Misnawati, M., Tahir, S. Z. B., Sibali, A., Anwar, W. P., Basri, N., & Bahtiar, H. (2025). Teaching Collocations and Academic Phrases with Corpus Linguistics: Applications for Specific and Academic Contexts. *Innovations in Language Education and Literature*, *2*(1), 8-20.

INTRODUCTION

In recent years, the integration of corpus linguistics into language education has gained increasing attention among applied linguists and language educators, particularly in teaching English for Specific Purposes (ESP) and English for Academic Purposes (EAP) (Egbert et al., 2020; Meyer, 2023; Prasetya et al., 2020). Corpora—large, searchable collections of authentic language data—provide access to real-life examples of language use across genres and disciplines (Gaeta et al., 2022; Roslim et al., 2020; Stefanowitsch, 2020). This data-driven approach enables learners to explore patterns, frequency, and contextualized usage of words and expressions, promoting a deeper understanding of language beyond textbook rules (Barth & Schnell, 2021; Roslim et al., 2020; Yin & Li, 2021).

Despite its pedagogical potential, teaching collocations and academic phrases—essential for achieving fluency and academic competence—remain challenging in many ESP/EAP classrooms (Misnawati, 2024; Misnawati, Anwar, et al., 2024; Misnawati, Nur, et al., 2024). Traditional materials often fail to capture the nuanced and discipline-specific language patterns learners need to master (Misnawati et al., 2025). Moreover, language teachers may lack the training or resources to incorporate corpus tools into their instructional design (Misnawati et al., 2025). As a result, many students struggle to use appropriate and natural-sounding academic language in their writing and speaking tasks, especially in professional and academic settings.

While several studies have highlighted the usefulness of corpus-based instruction in vocabulary learning (Huang, 2018; Li et al., 2025; Sun & Park, 2023), few have addressed how learners interact with collocation-rich input or academic phrase lists in ESP and EAP environments (Çalışkan & Gonen, 2018; D. Liu & Jiang, 2009; Tsai, 2021). Most existing research focuses on general English contexts or advanced university-level learners, leaving a gap in understanding how corpus-informed activities work at various proficiency levels or in specific professional fields (Boulton & Tyne, 2015; Negm & Mandour, 2020; Zare & Al-Issa, 2024). To address this gap, the present study explores using corpus linguistics tools and resources to teach collocations and academic phrases in ESP and EAP classrooms. It seeks to answer the following research questions:

1. How does corpus-based instruction influence learners' ability to identify and use collocations and academic phrases in ESP/EAP contexts?
2. What are the learners' perceptions of using corpus tools in the classroom to enhance their academic and professional language skills?

The study aims to design and implement corpus-informed learning activities and evaluate their effectiveness in promoting lexical competence and discourse appropriateness among ESP/EAP students. By grounding the instructional approach in authentic language use and encouraging Learner autonomy through data-driven discovery, the study contributes to developing innovative pedagogical strategies in applied linguistics. The implications of this research are twofold: first, it offers practical guidance for language educators seeking to enrich their ESP and EAP curricula with corpus-informed techniques; second, it contributes to the growing body of literature advocating for evidence-based language teaching that bridges the gap between linguistic research and classroom practice.

The Integration of Corpus Linguistics into Language Instruction

Applied corpus linguistics has significantly influenced the teaching of collocations and academic phrases, especially within contexts involving English for Specific Purposes (ESP) (McEnery et al., 2019; Mukherjee, 2004; Reppen & Simpson-Vlach, 2019). Corpus-based approaches provide learners with authentic language input, enabling them to observe and analyze real-world usage patterns, which are particularly beneficial for mastering collocations and academic expressions (Biber et al., 2010; Cheng, 2013; Misnawati, Nur, et al., 2024).

Recent studies have demonstrated the effectiveness of corpus-based instruction in enhancing learners' collocational competence (Huang, 2018; Li et al., 2025; Sun & Park, 2023a). For instance, Liu and Gablasova (2023) conducted a study involving 100 Chinese learners of English, utilizing both freely available and self-compiled corpora with the #LancsBox analytical tool. The findings revealed that learners not only improved their knowledge of collocations but also retained three months post-instruction, reporting increased confidence in using collocations and continued use of corpus tools beyond the course duration. Similarly, a study by Fang et al. (2021) focused on senior secondary school students preparing for the IELTS exam in China. The research indicated that corpus-based training significantly improved students' use of collocations in writing tasks, highlighting the applicability of corpus tools across different proficiency levels.

In EAP, explicit instruction using corpora has shown positive outcomes. A study involving international EAP students employed the Corpus of Contemporary American English (COCA) to teach lexical collocations (Ayyat, 2024; Liontas et al., 2023; Salama & Altohami, 2022). The results demonstrated a strong positive correlation between improved collocational competence and enhanced academic writing performance. Furthermore, integrating corpus analysis into EAP instruction has been advocated to provide students with authentic academic language experiences. Previous researchers discussed how corpus-based instruction aids in vocabulary teaching, academic writing, and language awareness, emphasizing the role of corpus analysis in exposing learners to authentic language use and enhancing their academic language proficiency (Akbaş & Başal, 2024; Topkaya & Çelik, 2024; Zehir Topkaya & Çelik, 2024).

Despite these advancements, challenges persist in implementing corpus-based instruction, such as the need for teacher training in corpus tools and the development of learner-friendly interfaces. Addressing these issues is crucial for maximizing the benefits of corpus linguistics in ESP and EAP settings.

METHODS

Research Design

This study employed a qualitative classroom-based action research design to investigate how corpus-based instruction—particularly through Corpus of Contemporary American English (COCA)—enhances

students' mastery of collocations and academic phrases in ESP and EAP learning contexts. The intervention was embedded into a six-week instructional unit, combining lectures, guided discovery tasks, and reflective activities. Action research was chosen to allow iterative teaching, observation, and reflection cycles tailored to the classroom setting.

Participants

The participants were 15 third-semester undergraduate students enrolled at Universitas Cahaya Prima Bone, Indonesia. All students had taken general English courses and were currently enrolled in an ESP-focused class aimed at academic and professional communication. Participants were selected through purposive sampling and consented to participate in the study voluntarily.

Data Collection

Multiple data sources were used during the six-week instructional period:

1. Classroom Observations: The researcher took detailed field notes to capture students' engagement and interactions with COCA during task-based learning.
2. Student Artifacts: Samples of writing assignments, vocabulary logs, and worksheets requiring COCA consultation were collected and analyzed.
3. Post-Instruction Questionnaire: This survey measured students' perceptions of COCA, including perceived benefits and usability.
4. Semi-Structured Interviews: Ten students were interviewed to explore their experiences and reflections on using corpus tools to improve their academic language.

COCA was introduced in week 2 as the primary tool for data-driven learning. Students were taught how to use basic search functions (e.g., keyword in context, collocate search, frequency) and applied these to explore word combinations relevant to their academic writing and presentation topics.

Data Analysis

Data were analyzed using thematic content analysis by Braun and Clarke (2006). Student writings were examined for the correct and contextually appropriate use of collocations and academic phrases, both pre-and post-intervention. Observational notes and interview transcripts were coded inductively, while questionnaire data were tabulated for pattern identification. Particular attention was given to: 1) Frequency and variety of collocation use; 2) Appropriateness and naturalness of academic phrases; and 3) Student reflections on COCA usage. To ensure research validity, data triangulation was conducted across observations, student outputs, and interviews. Member checking was also used to confirm interview interpretations with participants. While peer debriefing involved consultation with two experienced ESP instructors to verify themes and coding accuracy.

Ethical Considerations

Ethical approval was obtained from the institutional ethics board of Universitas Cahaya Prima. All students were informed about the purpose of the study and provided written informed consent. Participation was voluntary, and confidentiality was maintained through pseudonymizing names and secure data handling. Students were informed that participation would not influence their grades or academic evaluation.

FINDINGS

This section presents the key findings from classroom observations, student artifacts, post-intervention questionnaires, and semi-structured interviews with 15 first-semester students at Universitas Cahaya Prima. Thematic analysis yielded five interrelated themes: (1) vocabulary growth and collocation awareness, (2) student perceptions of using COCA, (3) initial challenges in navigating corpus tools, (4) active language discovery, and (5) learner autonomy and pedagogical implications. Each theme is supported by direct student quotations (coded S1–S15) and interpretive commentary.

At the start of the intervention, most students demonstrated limited control over academic vocabulary and collocations in both speaking and writing tasks. Initial writing samples were characterized by frequent lexical errors, overuse of general-purpose words (e.g., *"good," "important," "thing"*), and minimal variation in expressions. Over six weeks of corpus-based instruction using COCA, most students showed notable improvement in lexical choice, phrase variation, and awareness of word partnerships. Classroom observations confirmed increased student engagement, especially during discovery tasks involving keyword-in-context (KWIC) searches and collocation lists.

Theme 1: Vocabulary Growth and Collocation Awareness

One of the most prominent outcomes of the intervention was the students' improved awareness and application of academic collocations relevant to their field. Before using COCA, most students relied heavily on literal translations from Indonesian or general verbs like *make*, *do*, and *get*, often leading to awkward or incorrect word combinations. Through consistent exposure to real usage data via the COCA collocate function, students began to recognize and adopt more natural, discipline-specific word partnerships such as "conduct research," "user-friendly interface," and "enhance the learning experience."

Student S1 and S9 reflected on this shift:

"I always wrote 'do research' before. After checking COCA, I found 'conduct research' is more natural." (S1)

"I didn't know that we say 'take a course' instead of 'follow a course.' Now I use it correctly." (S9)

These reflections suggest that the corpus exposed students to authentic academic phraseology, correcting long-standing misusages that were not effectively addressed through traditional instruction.

Several students began to revise previously learned phrases after observing how native speakers use collocations in authentic texts. S3 shared:

"Before, I wrote 'do mistake.' After seeing COCA examples, I use 'make a mistake.'"

The impact of COCA extended beyond individual vocabulary items to include set expressions and noun-verb combinations. S10 and S7 described how a single search helped him refine technical language:

"When I searched 'interface' in COCA, I learned 'user-friendly interface' is common. I use that now." (S10)

"I saw that people say 'improve performance,' not 'make better performance.' That helped me fix my sentence." (S7)

This pattern of self-correction and refinement points to the power of data-driven learning to instill a deeper understanding of language use that students may not otherwise acquire from memorized word lists.

Moreover, COCA helped students internalize collocations relevant to their disciplinary context—Information Technology and Education. S6 and S8 emphasized how the corpus supported his understanding of phrases specific to digital tools:

"COCA helped me understand how to use 'develop app' or 'build system.' That's very useful for my field." (S6)

"Now I know 'enhance learning experience' is better than just saying 'make study better.'" (S8)

Writing samples collected from weeks 1 and 6 were compared, revealing a marked increase in the frequency and accuracy of collocational usage. On average, post-intervention writing contained 38% more collocational combinations, especially in verb-noun and adjective-noun pairs. For instance, phrases like *"optimize user interface," "develop digital skills,"* and "*present research findings"* appeared more frequently and naturally in students' final assignments and group presentations.

These improvements indicate that even novice learners can internalize and apply academic collocations effectively when guided by structured tasks and exposed to authentic corpus data. This is particularly important in ESP and EAP contexts, where discipline-specific lexical precision is essential but often missing from generic coursebooks.

Theme 2: Student Perceptions of Using COCA

Students' perceptions of COCA as a corpus-based learning tool were overwhelmingly positive, particularly its authenticity, credibility, and potential to build confidence. For many participants, COCA represented a shift from traditional learning methods—such as memorization or translation—to an exploratory approach grounded in real-world language data. This perspective fostered a sense of intellectual engagement and academic identity, even among beginners.

Student S8 noted the value of using COCA as a language-checking tool:

"COCA helps me check if my sentence is really used by native speakers."

This statement underscores a growing awareness of usage norms, with COCA serving as a reference point for students to validate their language choices against naturally occurring English. Instead of relying solely on intuition or prescriptive grammar, learners began consulting COCA to ensure that their expressions aligned with authentic usage.

Some students reported a stronger sense of academic involvement while using corpus tools. S9 captured this shift vividly:

"I feel like a real researcher when I use COCA. It's not just learning English—it's exploring data."

This quote reflects the empowering nature of corpus work, which not only supports language development but also encourages students to adopt the mindset of investigators. Rather than passive recipients of knowledge, students saw themselves as active participants in the learning process.

Several participants also emphasized the *confidence-building* effect of seeing language in context. S5 shared:

"It gives me confidence because I see real examples, not just the teacher's explanation."

Observing authentic usage patterns helped students overcome their fear of making mistakes. This sense of assurance was vital for first-semester learners, who felt insecure about using English in formal contexts. S15 added:

"Now I use COCA when I want to write something formal. It helps me sound more academic."

By consulting COCA before writing reports, essays, or emails, students felt they were producing language closer to the academic register expected in university settings. This finding demonstrates that corpus tools can contribute to developing academic identity and stylistic control.

Even students who initially identified as "beginners" reported feeling capable of engaging with corpus data. S12 explained:

"Even if I'm still a beginner, I can understand patterns from the examples in COCA."

This reflects the user-friendly impact of the collocate and KWIC (Keyword in Context) features, which present information in structured and pattern-rich formats. With appropriate guidance, even students with modest proficiency can infer usage norms and build intuition for collocations.

Moreover, several participants compared COCA and more commonly used tools such as Google Translate. S2 noted:

"I prefer COCA over Google Translate. It shows real sentences and combinations."

This preference suggests developing a critical awareness of language tools. Students recognize that while translation apps may offer immediate answers, COCA provides insight into how language *works* in honest communication. The benefit lies not only in correctness but in contextual richness.

Lastly, student S14 emphasized the motivational aspect of corpus learning:

"COCA made me curious to find out more about how English is really spoken and written."

Such intrinsic motivation is vital for sustaining engagement in ESP and EAP courses, particularly when learners navigate unfamiliar content. COCA appeared to foster curiosity and a discovery-oriented mindset, encouraging students to go beyond surface-level understanding.

In sum, students viewed COCA as a reference tool and an academic resource that validated their language use, deepened their understanding, and strengthened their sense of ownership over the learning process. These perceptions are particularly significant in English for Specific Purposes, where learners often struggle to find resources that reflect authentic language in their field. The corpus-based approach bridged this gap and positioned learners as capable, confident, and inquisitive users of English.

Theme 3: Initial Challenges in Navigating Corpus Tools

While students ultimately reported positive outcomes and increasing comfort with COCA, their initial encounters with the tool were marked by confusion, intimidation, and technological unfamiliarity. These early-stage difficulties highlight the steep learning curve many first-semester non-English majors face when engaging with corpus tools for the first time, particularly in environments where exposure to digital academic resources is still limited.

For many students, the interface complexity and the abundance of unfamiliar functions caused hesitation. S15 described the sense of overload upon first accessing the platform:

"The first time I opened COCA, I didn't understand what to click. Too many options."

This experience was echoed by other students unfamiliar with the basic structure of corpus interfaces, especially those filled with data, filters, abbreviations, and search parameters that require both linguistic and technical literacy.

Students also reported confusion regarding the numerical data and collocation statistics presented by COCA. S4 remarked:

"I typed a word and saw strange numbers. I didn't know what they meant."

These "strange numbers" likely refer to frequency counts, range indicators, or statistical association measures such as MI (Mutual Information) scores, which were entirely new to most participants. Such data can seem meaningless or intimidating without a proper introduction, especially to learners unfamiliar with linguistic terminology.

Another barrier was understanding the specific functions within COCA, such as the collocate tab. S7 explained:

The "Collocate tab was confusing. I didn't know how to set the range or interpret the list."

This comment reflects the cognitive demands of interpreting collocational patterns, particularly when users face several filtering options like position (left/right), window size, and frequency cut-offs—all of which were initially opaque to students.

Some learners even perceived COCA as a tool meant only for language professionals or linguists. S13 noted:

"I thought COCA was only for experts. But after practice, it became easier."

This statement highlights the perception of exclusivity and complexity surrounding corpus use. However, it also demonstrates the value of sustained practice and structured support in transforming that perception into a sense of accessibility.

A key technical component that proved difficult was understanding the KWIC (Keyword in Context) display, a central feature of corpus-based learning. S11 shared:

"I needed time to understand what KWIC is. It looks technical at first."

Although KWIC is fundamental in corpus linguistics for identifying usage patterns, it can appear abstract or dense to novice users. This reinforces the importance of explicitly teaching the what and *why* behind each corpus function.

Notably, many students credited their successful navigation of COCA to instructor guidance and peer collaboration. Without this scaffolding, they might have given up early in the process. S9 admitted:

"Without the teacher's explanation, I think I would have stopped using COCA."

This comment illustrates how essential it is for instructors to anticipate and address the tool's cognitive and technical barriers through demonstration, repetition, and simplified guidance materials—especially in the early weeks of exposure.

Finally, language barriers within the interface also posed problems for students with lower English proficiency. S1 described:

"The site is all in English. I had to ask friends about the meaning of 'spoken,' 'fiction,' 'academic' tabs."

The absence of localized or multilingual support in COCA meant that navigating basic functions required additional language knowledge or peer assistance. This is particularly relevant for students studying in non-English-dominant contexts.

Taken together, these reflections highlight that although COCA offers rich learning potential, its effective use in beginner ESP/EAP classrooms depends heavily on providing scaffolded support, simplifying instructions, and peer-to-peer learning strategies. Addressing technical and linguistic challenges early in the instructional process enables students to overcome initial resistance and develop sustained corpus literacy.

Theme 4: Active Language Discovery

As students became familiar with COCA, their learning shifted from passive reception to active exploration. Rather than waiting for vocabulary to be provided by the teacher or relying on pre-packaged word lists, students started formulating their own questions and searching for answers independently within the corpus. This marked a significant pedagogical transition—students began viewing language not simply as something to memorize but as a system to investigate.

S2 exemplified this shift when reflecting on a COCA search related to technical vocabulary:

> *"I searched 'design interface' and found common verbs that I didn't learn before, like 'implement' or 'optimize.'"*

This quote shows that students were not merely confirming what they already knew but using the corpus to *expand* their repertoire. In this case, a simple search led to more precise and context-specific verbs commonly used in technology and design.

Some students became curious about grammatical patterns and verb positioning, not just isolated collocations. S3 explained:

> *"I like to search verbs and see what comes before and after. It makes me more aware."*

This behavior indicates the emergence of pattern noticing—a critical component of data-driven learning. By observing what typically surrounds a verb, students began to intuitively understand lexical environments, register constraints, and semantic preferences.

Students also engaged in comparative analysis between different word combinations, using frequency and co-text as a guide to appropriateness. S12 shared:

> *"I compared 'do assignment' and 'complete assignment' in COCA. Now I know which one is better for writing."*

Rather than accepting teacher instruction at face value, this student applied a corpus-informed judgment to distinguish between competing options. The ability to verify alternatives based on real usage reflects deeper engagement and growing confidence in decision-making.

S15 demonstrated a more metacognitive application, using COCA as a self-check tool:

> *"Sometimes I search my own sentences in COCA to check if they're common."*

This behavior reveals a high degree of language awareness. The student took the initiative to correct errors, test intuitions and revise based on evidence—a powerful habit in academic writing.

Some participants began to reconceptualize collocations as patterns of meaning shaped by context rather than just lists of word pairs. S6 captured this understanding:

> *"I learned that collocation is not just vocabulary. It's also about context and style."*

This quote underscores a growing sophistication in how students conceptualize language use. They began to appreciate that collocations carry stylistic and pragmatic implications, especially in formal academic or professional settings.

As students internalized these discoveries, they reported greater independence in their learning process. S7 reflected:

> *"Now I feel more independent. I don't need to ask the teacher every time."*

This sense of autonomy is a key outcome of corpus-based instruction. It suggests that learners are developing the skills and confidence to explore language beyond the classroom, taking charge of their progress rather than depending solely on external guidance.

Finally, students began critically evaluating their language use and recognizing why certain expressions might feel "off" in context. S12 noted:

> *"COCA helped me realize why some phrases sound weird. They're not used by native speakers often."*

This realization is a hallmark of corpus literacy—detecting and self-correcting unnatural constructions based on exposure to authentic usage. It also demonstrates an increasing ability to connect intuition with empirical evidence.

Collectively, these accounts reflect a significant evolution in the learner mindset. Students moved from compliance to inquiry, dependency to autonomy, and surface-level vocabulary acquisition to a deeper understanding of usage patterns. The integration of COCA in the classroom empowered learners to learn *about* English and explore and interact with it as a living, dynamic system.

Theme 5: Learner Autonomy and Pedagogical Implications

The final theme highlights a key long-term benefit of integrating COCA into ESP/EAP instruction: cultivating learner autonomy and extending corpus use beyond the classroom. As students became more comfortable and confident in navigating corpus tools, they adopted COCA as part of their independent learning strategies. Notably, several participants voluntarily applied COCA to tasks unrelated to the structured classroom activities, signaling both transfer of learning and sustained engagement.

Student S9 described how she continued using COCA in a class project outside the formal scope of English instruction:

"I use COCA when writing captions for my social media project in class. It helps me be more accurate."

This response illustrates how the tool's function as a lexical reference extended to creative and digital contexts, indicating a broader utility of corpus consultation for academic and semi-professional communication.

Students also began sharing the tool with peers, reinforcing their own understanding through informal teaching roles. S3 shared:

"I told my classmate in another course to try COCA. Now we both use it to check our writing."

Such peer-to-peer dissemination of tools reflects a growing sense of ownership over learning. It shows that students value COCA and feel responsible and competent enough to recommend and demonstrate its use to others.

S1 extended this behavior even further:

"After this class, I started using COCA for my English club activities too."

This voluntary application suggests that COCA has become part of the student's toolkit—relevant, accessible, and beneficial for different communicative purposes. It also indicates that the corpus had taken root as a lasting resource, not just temporary classroom exercise.

Students also documented their learning independently, creating personal phrase banks or vocabulary lists derived from their corpus searches. S2 explained:

"I made a note of common phrases I found in COCA and used them in my final presentation."

This quote illustrates the metacognitive dimension of corpus work—students were not just reading and reacting to language but recording, reflecting on, and applying what they found.

Several students expressed a sense of pride and empowerment in helping others access the tool, signaling the development of Learner identity as a source of knowledge. S13 reflected:

"I feel proud when I can help my friend check collocations using COCA."

This pride was not only in mastering the tool but in contributing meaningfully to their peers' learning journeys, suggesting a collaborative dimension to autonomy that is often overlooked.

Perhaps most significantly, students described a transformation in their perception of themselves as language learners. S4 articulated this change clearly:

"Now I feel like I can learn English by myself. I don't always wait for the teacher to correct me."

This shift from dependence to self-direction is a key goal in language education, particularly in higher education, where students are expected to develop into lifelong learners. As a discovery-based tool, COCA supported this trajectory by equipping students with both a method and a mindset.

Finally, students began to see the potential of corpus tools for cross-disciplinary application. S15 shared this aspiration:

"I hope we can use tools like COCA in other subjects, too. It helps me write better."

This sentiment suggests that students recognized the value of corpus linguistics in English learning and began to envision its relevance across academic disciplines—a crucial step toward disciplinary literacy.

In sum, the integration of COCA fostered better vocabulary use or collocational awareness and helped shape learners' attitudes toward English and themselves. It positioned students as capable users of authentic resources, encouraged self-reflection and promoted independent learning. These outcomes underscore the pedagogical potential of corpus-based instruction to move beyond language accuracy and into learner empowerment and educational transformation.

DISCUSSION

The integration of the Corpus of Contemporary American English (COCA) into the instructional practices of first-semester non-English major students at Universitas Cahaya Prima Bone has yielded multifaceted insights into the efficacy of corpus-based approaches in English for Specific Purposes (ESP) and English for Academic Purposes (EAP) contexts. This study's findings corroborate existing literature and extend our understanding of how corpus tools can be effectively utilized to enhance language acquisition, learner autonomy, and pedagogical practices.

One of the most salient outcomes observed was the significant improvement in students' academic vocabulary and collocation awareness. This aligns with Data-Driven Learning (DDL) principles, where learners engage directly with authentic language data to discover linguistic patterns (Biber et al., 2010; Cheng, 2013; Misnawati, Nur, et al., 2024). The student's ability to identify and use appropriate collocations, such as *"conduct research"* instead of *"do research,"* demonstrates the effectiveness of corpus consultation in refining lexical choices. This finding corroborates previous studies highlighting the role of corpora in improving collocational competence (Misnawati, 2024; Misnawati, Anwar, et al., 2024; Misnawati, Nur, et al., 2024).

Moreover, improvements in accurate collocational usage observed in post-intervention essays underscore the potential of corpus tools like COCA in facilitating vocabulary acquisition and natural language use, particularly in ESP and EAP contexts. This quantitative improvement is consistent with the findings of Huang (2018), Li et al. (2025), Sun and Park (2023), who reported significant gains in verb-noun collocation usage among EFL learners following corpus-based instruction.

Students' perceptions of COCA were overwhelmingly positive, noting its authenticity and usefulness in providing real-life language examples. This perception aligns with the constructivist approach to language learning, where learners build knowledge through active exploration (Cobb, 2007). Students reported increased confidence in their writing, attributing it to the ability to verify language usage through COCA. This self-assurance is crucial in academic settings, where precision and appropriateness are paramount. The findings echo those of McEnery et al. (2019), Mukherjee (2004), and Reppen and Simpson-Vlach (2019), who emphasized the motivational impact of DDL on learners.

Furthermore, students' preference for COCA over traditional tools like Google Translate indicates a shift toward more analytical and evidence-based language learning strategies. This transition reflects a deeper engagement with language learning processes, fostering critical thinking and self-directed learning (Barth & Schnell, 2021; Roslim et al., 2020; Yin & Li, 2021).

Despite the benefits, students initially faced challenges in using COCA, including difficulties with its interface and understanding corpus-specific terminology. These obstacles are consistent with findings from Misnawati et al. (2025), who identified technical complexity as a barrier to effective corpus use among beginners. The necessity of instructor guidance and peer support in overcoming these challenges highlights the importance of scaffolding in DDL implementation (Misnawati et al., 2025). Tailored instruction and collaborative learning can mitigate the initial intimidation associated with corpus tools, facilitating smoother integration into the learning process.

As students became more comfortable with COCA, they transitioned from passive recipients of information to active language explorers. Engaging in tasks such as analyzing verb-noun collocations and comparing phrase frequencies fostered a deeper understanding of language patterns. This shift aligns with learner autonomy in DDL, where students take control of their learning by formulating hypotheses and testing them against corpus data (Boulton & Tyne, 2015; Negm & Mandour, 2020; Zare & Al-Issa, 2024). The development of metalinguistic awareness observed in students reflects the potential of corpus-based approaches to promote critical thinking and self-directed learning, essential skills in academic contexts.

The study's findings indicate that integrating COCA into instruction enhances language proficiency and fosters learner autonomy. Students continued use of COCA beyond classroom requirements, including in extracurricular activities and peer tutoring, demonstrates the tool's impact on their learning habits. This autonomy is a key objective in ESP and EAP education, preparing students to navigate language challenges in their academic and professional pursuits independently. The results support the notion that corpus tools can serve as catalysts for developing lifelong learning skills (Boulton, 2010). The findings affirm that COCA significantly aids in acquiring appropriate collocations, enhancing students' academic writing skills (Ayyat, 2024; Liontas et al., 2023; Salama & Altohami, 2022). Additionally, students perceive COCA as a valuable resource contributing to their confidence and autonomy in language learning.

The positive outcomes suggest that incorporating corpus tools like COCA into language instruction can be highly beneficial, even for learners at the beginner level. Educators should consider integrating DDL approaches into their curricula, providing adequate support to navigate initial challenges. Furthermore, the development of learner autonomy observed in the study underscores the importance of fostering environments that encourage exploration and self-directed learning. By equipping students with the skills to utilize resources like COCA, educators can empower them to take ownership of their language development.

The implications of these findings are significant for language educators, curriculum designers, and institutions aiming to modernize English instruction for specific and academic purposes. First, using corpus linguistics tools can help bridge the gap between textbook language and authentic use, particularly in subject-specific domains where conventional materials often fall short. Second, integrating COCA into classroom activities supports a learner-centered approach that enhances vocabulary acquisition and digital and analytical literacy. Third, as learners engage with corpus data, they are encouraged to develop metacognitive strategies and critical awareness of language use, which are transferable to other academic and professional communication domains. These benefits align with current trends in applied linguistics that advocate empirical, technology-enhanced, and learner-driven pedagogies.

However, this study is not without its limitations. As a small-scale classroom-based action research project, the sample size was limited to 35 students from a single institution, which may affect the generalizability of the findings. The study also focused on perceptions and qualitative improvement in student writing without extensive quantitative measurement of long-term language retention. Additionally, while COCA was chosen for its accessibility and reliability, some students found the interface challenging, particularly in the initial sessions. These technical and cognitive barriers underscore the need for adequate training and simplified instructional resources when introducing corpus tools to novice users.

Future research can address these limitations by expanding the participant pool across institutions, disciplines, and proficiency levels to examine how corpus-based instruction functions in more diverse settings. Longitudinal studies are also needed to assess the retention of collocational knowledge and the continued use of corpus tools after formal instruction ends. It would be valuable to explore the integration of other learner-friendly corpora or interfaces—such as SKELL, ECOPA, or BYU iWeb—and to examine how different types of corpus tasks (e.g., guided vs. unguided, individual vs. collaborative) influence learning outcomes. Moreover, further investigation is needed into the role of teacher training in corpus linguistics, as effective facilitation significantly shapes students' success with DDL approaches.

In conclusion, the findings of this study contribute to the growing body of evidence supporting the use of corpus linguistics in applied educational settings. COCA, when integrated thoughtfully into ESP and EAP classrooms, offers more than access to authentic language; it opens a pathway for learners to become informed, autonomous, and reflective users of English. With appropriate pedagogical scaffolding, even first-semester students can harness the power of corpus tools to develop meaningful linguistic competence tailored to their academic and professional needs.

CONCLUSIONS

This study has demonstrated the practical value and pedagogical potential of integrating the Corpus of Contemporary American English (COCA) into ESP and EAP instruction for beginner-level students, particularly those outside English language majors. By incorporating data-driven learning (DDL) into classroom practice, students were exposed to authentic language use and empowered to explore linguistic patterns relevant to their academic and professional disciplines. The findings highlighted clear improvements in students' ability to recognize and apply appropriate collocations and academic phrases, as seen in their writing tasks and verbal responses. Through regular use of COCA, learners developed a deeper understanding

of how words are naturally combined in context, moving beyond rote vocabulary memorization to meaningful language use informed by real-life data. Beyond lexical development, this study also revealed positive shifts in learner attitudes. Students expressed high confidence and engagement as they began to take control of their learning. Their reflections indicated a growing sense of independence; some even adopted COCA as a tool for peer support and self-study beyond the classroom. These outcomes strongly support the integration of corpus tools in language learning environments, especially where fostering learner autonomy is a curricular goal. Importantly, this study confirms that even novice learners, when appropriately scaffolded, can benefit from advanced digital tools traditionally reserved for higher-level or linguistically mature students.

REFERENCES

Akbaş, E., & Başal, A. (2024). Enhancing Academic English Teaching and Learning: The Integration of Corpus Analysis in EAP. In *Teaching English for Academic Purposes: Theory into Practice* (pp. 249–272). Springer.

Ayyat, A. (2024). Implementing a Three-Stage Approach to Teach Writing Skills Using the Corpus of Contemporary American English in a Foundation Program Context. *Arab World English Journal*, *15*(4), 241–254. https://doi.org/10.24093/awej/vol15no4.15

Barth, D., & Schnell, S. (2021). *Understanding corpus linguistics*. Routledge.

Biber, D., Reppen, R., & Friginal, E. (2010). Research in corpus linguistics. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed, pp. 548–570). Oxford University Press.

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x

Boulton, A., & Tyne, H. (2015). Corpus-based approaches to language learning: Towards DDL for all? *Language Learning & Technology*, *19*(3), 1–24.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Çalışkan, G., & Gonen, S. I. K. (2018). Training teachers on corpus-based language pedagogy: Perceptions on using concordance lines in vocabulary instruction. *Journal of Language and Linguistic Studies*, *14*(4), 190–210. http://www.jlls.org/index.php/jlls/article/download/1020/430

Cheng, W. (2013). Corpus-based linguistic approaches to critical discourse analysis. *The Encyclopedia of Applied Linguistics*, 1353–1360.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, *11*(3), 38–63. https://doi.org/10125/44117

Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge University Press.

Fang, L., Ma, Q., & Yan, J. (2021). The effectiveness of corpus-based training on collocation use in L2 writing for Chinese senior secondary school students. *Journal of China Computer-Assisted Language Learning*, *1*(1), 80–109.

Gaeta, L., Angster, M., Cioffi, R., & Bellante, M. (2022). Corpus linguistics for low-density varieties. Minority languages and corpus-based morphological investigations. *Corpus*, *23*.

Huang, L.-S. (2018). Taking stock of corpus-based instruction in teaching English as an international language. *RELC Journal*, *49*(3), 381–401. https://doi.org/10.1177/0033688217698294

Li, D., Noordin, N., Ismail, L., & Cao, D. (2025). A Systematic Review of Corpus-Based Instruction in EFL Classroom. *Heliyon*.

Liontas, J. I., Bangun, I. V, & Li, S. (2023). Lexical Collocational Instruction in EAP Writing via COCA. *Teaching English with Technology*, *23*(1), 80–106.

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

Liu, T., & Gablasova, D. (2023). Data-driven learning of collocations by Chinese learners of English: a longitudinal perspective. *Computer Assisted Language Learning*, 1–26.

McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus linguistics, learner corpora, and SLA: Employing technology to analyze language use. *Annual Review of Applied Linguistics*, *39*, 74–92.

Meyer, C. F. (2023). *English corpus linguistics: An introduction*. Cambridge University Press.

Misnawati, M. (2024). *Corpus linguistics in English for public administration in Indonesian higher education context* [Disertation]. Universitas Negeri Makassar.

Misnawati, M., Anwar, W. P., & Astri, Z. (2024). A corpus linguistics study of frequent collocations in public administration. *Proceedings of ICoLT-Hybrid Conference*, *1*(1), 40–52.

Misnawati, M., Nur, S., & Tahir, S. Z. (2024). Corpus linguistics today: A qualitative approach. *Research and Innovation in Applied Linguistics [RIAL]*, *2*(1), 45–62. https://doi.org/10.31963/rial.v2i1.4486

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. *The Qualitative Report*, *30*(1), 3024–3042. https://doi.org/https://doi.org/10.46743/2160-3715/2025.6867

Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective*. Rodopi.

Negm, M. S. N., & Mandour, W. S. M. (2020). The Use of an Open Access Corpus Web Tool in a DDL Writing Model. *CDELT Occasional Papers in the Development of English Education*, *70*(1), 127–142.

Prasetya, E. P., Ekawati, A. D., Nugraha, D. S., Marzuq, A., & Darlis, T. S. (2020). Corpus linguistics, language corpora and language teaching . *English Journal*, *14*(2), 75–83.

Reppen, R., & Simpson-Vlach, R. (2019). Corpus linguistics. In *An introduction to applied linguistics* (pp. 91–108). Routledge.

Roslim, N., Abdullah, M. H. T., Aziz, A., Nimehchisalem, V., & Almuddin, A. (2020). Corpus research applications in second language teaching: A look into corpus-informed materials development. *International Journal of Modern Languages and Applied Linguistics (IJMAL)*, *4*(2), 15–37.

Salama, A. H. Y., & Altohami, W. M. A. (2022). Enhancing EFL Students' COCA-Induced Collocational Usage of Coronavirus: A Corpus-Driven Approach. *International Journal of Advanced Computer Science and Applications*, *13*(2).

Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.

Sun, W., & Park, E. (2023a). EFL Learners' Collocation Acquisition and Learning in Corpus-Based Instruction: A Systematic Review. *Sustainability*, *15*(17), 13242.

Sun, W., & Park, E. (2023b). EFL Learners' Collocation Acquisition and Learning in Corpus-Based Instruction: A Systematic Review. *Sustainability*, *15*(17), 13242.

Topkaya, E. Z., & Çelik, H. (2024). Introduction to Teaching English for Academic Purposes: Theory. *Teaching English for Academic Purposes: Theory Into Practice*, 3.

Tsai, Y.-R. (2021). Exploring the effects of corpus-based business English writing instruction on EFL learners' writing proficiency and perception. *Journal of Computing in Higher Education*, *33*(2), 475–498. https://doi.org/10.1007/s12528-021-09272-4

Yin, X., & Li, S. (2021). Lexical bundles as an intradisciplinary and interdisciplinary mark: A corpus-based study of research articles from business, biology, and applied linguistics. *Applied Corpus Linguistics*, *1*(1), 100006. https://doi.org/https://doi.org/10.1016/j.acorp.2021.100006

Zare, J., & Al-Issa, A. (2024). Exploring task engagement strategies in DDL-enhanced tasks: Insights from EFL learners. *Contemporary Educational Psychology*, *78*, 102299.

Zehir Topkaya, E., & Çelik, H. (2024). Introduction to Teaching English for Academic Purposes: Theory into Practice. In *Teaching English for Academic Purposes: Theory into Practice* (pp. 3–18). Springer.