SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 4888-4900
Open Access. Online & Print

DOI: 10.5281/zenodo.19100336

TRACING COLLOCATIONAL CHANGE IN UNDERGRADUATE APPLIED LINGUISTICS RESEARCH AT IMSIU (2022–2025): A CORPUS-BASED DIACHRONIC STUDY

Mohammad Abdullah Alhammad[1*]

[1]Department of English Language and Literature, College of Languages and Translation, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia, https://orcid.org/0009-0007-0203-9346

Received: 01/02/2026
Accepted: 05/03/2026

Corresponding Author: Mohammad Abdullah Alhammad

ABSTRACT

The current study investigates diachronic changes in collocational patterns in undergraduate research projects in applied linguistics at Imam Mohammad Ibn Saud Islamic University (IMSIU) between 2022 and 2025. The corpus used depended on students' titles, abstracts, and other analytical sections. It was divided into four annual sub-corpora to examine lexical and thematic development. Using this type of corpus based on an analytical workflow, collocations of two or three words were extracted and evaluated using Mutual Information and t-score measures within symmetrical ±5-token windows. This was also supplemented by dispersion analyses to reduce topic-burst effects. Collocations were then classified by grammatical pattern (Adj +N, N+N, V+N). After that, they were mapped onto specific thematic domains such as language skills, pedagogy, technology/AI, translation, and bilingualism. The study's findings highlighted a gradual shift from traditional pedagogy-oriented bundles (such as language learning, EFL students, and motivation to learn) in 2022 toward combinations that focused on media and interaction in 2023 (e.g., listening and speaking skills, classroom games). This led to a strong rise in the use of collocations related to technology and Artificial Intelligence (AI) in 2024–2025 (such as machine translation and AI-assisted writing). These changes in collocational use show that EFL undergraduates developed stronger academic language skills and also changed their research interests. The current study also discussed implications for curriculum design, especially genre modeling and academic-lexis instruction, and for assessment, including the incorporation of evaluation criteria that were collocation-sensitive. The study provided an easy-to-repeat and low-cost approach for monitoring student research writing. It also introduced a reusable system to interpret changes in collocation over time.

KEYWORDS: Collocation; Corpus Linguistics; EFL Writing; AI-Assisted Learning; Saudi Higher Education; Diachronic Analysis

1. INTRODUCTION

It is well known that corpus-based approaches play an important role in the field of applied linguistics by providing systematic methods for analyzing language use, variation, and development across educational and professional contexts (Biber & Reppen, 2015; Gries, 2021). In the field of EFL education, corpus-driven research has been highly effective in identifying lexical and phraseological patterns that distinguish expert academic writing from learner production (Paquot & Granger, 2021). On the same hand, collocations, especially in term of recurrent lexical combinations as well as semantic and grammatical features, are widely recognized as key indicators of advanced proficiency and academic writing maturity (Nesselhauf, 2003; Sanosi, 2025). There is empirical evidence that corpus-informed instruction improves learners' acquisition and use of collocations across different proficiency levels (Abdellah, 2015, Du, Afzaal, & Al Fadda, 2022; Sun & Park, 2023). Although there is growing interest in learner collocations, most of the existing studies focus on classroom essays, online learner corpora, or controlled elicitation tasks (Bestgen & Granger, 2018; Paquot, 2019). Therefore, authentic academic genres, such as undergraduate research projects, remain underexplored especially in Arab EFL contexts, where English-medium undergraduate research is expanding. Most previous studies have discussed challenges faced by Saudi and Jordanian learners in mastering grammatical and lexical collocations, especially verb–noun and prepositional combinations (Alfaqara, 2023; Alsulayyi, 2024). Such studies reflected the need for context-sensitive, longitudinal investigation.

In Saudi Arabia, recent research has highlighted the role of online and blended learning environments in enhancing students' linguistic development, academic engagement, and research practices. On the other hand, learners' interactions with learning management systems and digital writing tools showed effects on lexical sophistication and academic discourse development (Ali El Deen & Mahmoud, 2025). However, there were few studies that tackled such developments or reflected longitudinally in extended undergraduate research writing within one specific institutional context. At Imam Mohammad Ibn Saud Islamic University (IMSIU), undergraduate students majoring in applied linguistics produce English-medium research projects every year on topics related to language pedagogy and learner motivation to artificial intelligence, machine translation, and digital learning analytics. These projects constitute an excellent diachronic dataset, which can be systematically analyzed in terms of collocational development and changing academic literacy and disciplinary orientation (Hyland, 2018; Römer, 2020). Research based on corpus demonstrated the way in which continuous work with a discipline-specific phraseology could support the development of academic writing (Birhan et al., 2024; Misnawati et al., 2025). In this way, they underscored the relevance of analyzing collocational patterns in extended undergraduate research texts.

The current study attempted to fill in this gap through using a corpus-based diachronic analysis of undergraduate applied linguistics research projects produced at IMSIU between 2022 and 2025. By using this annually constructed sub-corpora, the current study tried to extract two- and three-word collocations, evaluate their statistical strength using Mutual Information and t-score measures, and examine structural preferences (e.g., Noun + Noun, Adjective + Noun, Verb + Noun).

The current study aims to identify lexical, structural, and thematic shifts that reflect the development of students' academic discourse competence and research orientations. Accordingly, the study tried to answer the following research questions:

1. What are the most frequent collocational patterns used by undergraduate applied linguistics students between 2022 and 2025?
2. How do these collocational patterns change over time in terms of frequency, structure, and thematic orientation?
3. What do these diachronic shifts reveal about the development of academic discourse competence and research culture among EFL students at IMSIU?

The current study extended descriptions of learner language to institutional academic writing by tracing collocational change that happened across four academic years. The study is expected to contribute theoretically by providing authentic research genres rather than classroom tasks. This would be implemented through offering data-driven insights for curriculum design, research-writing instruction, and assessment in EFL higher education (Römer & O'Donnell, 2022).

2. LITERATURE REVIEW

This section is assigned to review research on collocations in academic writing, corpus-based analyses of learner collocational use, developmental perspectives from learner corpora, and diachronic institutional approaches. It also frames the rationale

for examining collocational change in undergraduate applied linguistics research writing.

A. Collocations And Academic Language Competence

Collocation is commonly defined as the statistically significant co-occurrence of words. It has been central to linguistic theory since Firth's (1957) assertion that "you shall know a word by the company it keeps." Then a subsequent development occurred in corpus linguistics. This demonstrated that collocations play an important role in fluency, idiomaticity, and disciplinary identity (Sinclair, 1991; Hoey, 2005). Another body of empirical research confirmed a strong relationship between collocational competence and overall language proficiency, as well as with the production of coherent, lexically sophisticated writing (Bestgen & Granger, 2018; Nesselhauf, 2003). On the other hand, Corpus-based studies usually show that lower-proficiency learners use fewer varied collocations, but this was usually accompanied by reduced semantic precision (Du, Afzaal, & Al Fadda, 2022). More systematic reviews further showed that corpus-informed instruction facilitates learners' acquisition of both lexical and grammatical collocations (Sun & Park, 2023). Such results findings suggest that collocational knowledge reflects not only lexical breadth but also writers' epistemic positioning within academic discourse (Hyland, 2018).

B. Corpus-Based Analyses of Collocations in Academic Writing

In the 1990th learner corpora exported and corpus-based methods became necessary in the study of collocational usage in EFL and ESL academic writing (Gries, 2021; Römer & O'Donnell, 2022). AntConc and Sketch Engine are some of the tools that helped the researchers to record systematic variance between expert and learner writing. In the same breath, disciplinary variation in collocational use was also well reported. However, most of the research was restricted to academic publications or postgraduate writing only when the undergraduate research writing particularly the research projects were comparatively unexplored and this hampered our understanding of early-phase academic phraseological development.

C. Collocational Development in Learner Corpora and EFL Contexts

The approach of development became the more popular one in the sphere of learner corpora research. It investigated the development of the collocational knowledge over the proficiency levels (Vyatkina, 2020). Also, longitudinal studies indicated that learners internalize discipline specific collocations over time when exposed to these collocations together with feedback and instructions. Nevertheless, the process of development was frequently nonlinear and influenced by the cognitive and instructional and contextual elements (Durrant & Schmitt, 2009; Paquot, 2019). It is no secret that EFL learners often struggle with abstract nouns, academic verbs, and prepositional phrases, indicating discontinuities in academic lexical patterning. Research on Arab EFL settings also indicated that difficulties with learning grammatical collocations, especially verb-noun and prepositional constructions, are common (Alfaqara, 2023; Alsulayyi, 2024). Although English-based academic writing is applied extensively in Saudi universities, the collocational sophistication is frequently low as EFL students use high frequency and general-purpose combinations extensively. Simultaneously, the studies on the topic, in turn, identified that online participation and virtual learning could have a positive impact on the process of lexical development and writing quality (Ali El Deen & Mahmoud, 2025). As later research found, instructional interventions targeting the use of academic phraseology could positively influence its use in context-specific situations among learners (Misnawati et al., 2025).

D. Diachronic And Institutional Perspectives on Collocational Shifts

There are relatively few studies that adopted a diachronic perspective on collocational use compared with cross-sectional research. Biber and Gray (2016) indicated that tracking phraseological patterns over time showed deep changes in academic discourse, including increased grammatical compression and lexical sophistication. Römer and O'Donnell (2022) similarly argued that diachronic collocational analysis could illustrate how disciplinary lexis was gradually internalized across learner cohorts. In Saudi higher education, there were some institutional changes, especially in terms of the expansion of digital pedagogy and learning management systems. These changes had a big impact on both academic writing practices and literacy development (Al-Mubireek et al., 2022). However, most of the diachronic studies used large-scale or multi-institutional corpora, left localized institutional contexts without exploration. Currently, corpus-informed instruction has turned to become increasingly embedded in higher education (Birhan & Nurie, 2024; Alenizi & Adawi, 2024; Liu &

Gablasova, 2023), so institution-level diachronic analyses are urgently needed to discover these developmental trajectories in authentic academic genres. Overall, existing research studies confirm the main role of collocations in academic literacy and their value in distinguishing expert from non-expert writing. Nevertheless, much of the literature remains cross-sectional as it only focuses on advanced or published genres, or relies on short classroom tasks. There are only few studies that examine collocational development longitudinally within a single institution, particularly in Arab EFL contexts. Therefore, the current study tried to address this gap by examining four years (2022–2025) of undergraduate applied linguistics research writing at IMSIU using diachronic corpus methods.

3. METHODOLOGY

A. Research Design and Approach

The current study employed a corpus-based design which is both descriptive and diachronic as well. This specific design was mainly used to examine changes in lexical and collocational patterns in undergraduate applied linguistics research produced at Imam Mohammad Ibn Saud Islamic University (IMSIU) between 2022 and 2025. This specific design included both quantitative and qualitative procedures. As for the quantitative analysis, it identified recurrent two- and three-word collocations whereas the qualitative analysis tried to give interpretation of their linguistic and thematic significance. This mixed approach helped the researcher to make systematic investigation of academic language development within a single institutional context. It also supported transparency and replicability (Gries, 2021; Römer & O'Donnell, 2022). The current study adopted an analytical framework that drew upon Sinclair's (1991) idiom principle, which conceptualized meaning as emerging from recurrent lexical combinations. Also, Biber and Gray's (2016) model of phraseological development was also used. Therefore, collocations were treated as indicators of academic literacy development and disciplinary alignment among EFL learners.

B. Corpus Description (Participants and Data)

The corpus used in the current study consisted of 20 undergraduate English-medium research projects that were submitted by applied linguistics students at IMSIU between 2022 and 2025. The current study selected only projects that were formally submitted for evaluation. They should also exceed 3,000 words (excluding references and appendices), and tackle linguistics or language-learning topics rather than literary analysis. The resulting corpus contained about 140,000 words. It was divided into four annual sub-corpora. They reflected shifts in thematic focus across four years: pedagogy and motivation (2022), interactive and media-based learning (2023), technology-oriented studies involving artificial intelligence and translation tools (2024), and cognitive and neuro-linguistic topics, including autism and AI-assisted writing (2025). Although this number could be relatively small, the corpus represented authentic academic writing produced under rigorous institutional conditions and provided a suitable basis for diachronic analysis.

C. Instruments And Tools

This paper employed the use of the AntConc 4.0 (Anthony, 2024) program to perform the corpus processing and analysis and to provide analytical consistency and accuracy. The researcher cross-tested a few or selected frequency lists and output collocations via Sketch Engine. These tools assisted the researcher to derive systematical extraction, filtering, and categorization of collocations across the four sub-corpora.

D. Data Collection and Corpus Preparation

All the texts were anonymized before the process of analysis began. It was also done by the researcher, deleting tables, references, appendices, and formatting artifacts to establish uniformity. Each research project was processed separately by the researcher, and then combined into its corresponding annual sub-corpus (20222025). Afterward, the entire process of data collection was meticulously recorded so that it would be transparent and repeatable.

E. Data Analysis Procedures

The researcher depended on words that frequently co-occurred within a ±5-token window to identify collocations. It can be stated that both two-word and three-word combinations were included to capture common academic expressions such as language learning, EFL students, and improve writing skills. On the other hand, function words (such as and, of, the) were also excluded to focus on semantically meaningful lexical combinations. The researcher used frequency thresholds and association measures, specifically Mutual Information (MI) and t-score to ensure and evaluate collocational strength. This combined method helped the researcher to identify word combinations that were both

statistically important and commonly used (Gries, 2021). After that the researcher extracted collocations and categorized them according to grammatical structure and linguistic function, following Durrant and Schmitt (2009), as illustrated in Table 1.

Table 1: Categorization Of Collocations by Grammatical Structure and Linguistic Function.

Grammatical Pattern	Example	Linguistic Function
Adjective + Noun	second language, Saudi learners	Descriptive or classificatory
Noun + Noun	language learning, EFL students	Core conceptual terminology
Verb + Noun	develop proficiency, enhance skills	Process-oriented expressions
Noun + Preposition + Noun	impact of technology, role of media	Analytical and causal framing

The researcher also classified the collocations into semantic domains such as pedagogy, technology, motivation, and AI-related discourse to examine thematic change over time. The researcher then conducted cross-year comparisons to identify shifts in lexical preference and topical orientation.

F. Reliability, Validity, And Ethical Considerations

Various processes used by the researcher were employed to ensure reliability and validity. As an illustration, random samples of extracted collocations were cross-checked in Sketch Engine, and two practitioners of applied linguistics reviewed the selected grammatical categorization and thematic assignments themselves. Their inter-rater reliability was a Cohen 0.87 which corresponds to a high-degree of agreement. Ethical approval was also obtained from the IMSIU College Research Ethics Committee. The researcher was keen to choose all projects that had been publicly presented and archived before starting any data collection. Personal information, grades, and supervisor comments were anonymized as instructed by the COPE guidelines and IMSIU research integrity policies.

4. RESULTS & DISCUSSION

The diachronic corpus analysis of undergraduate applied linguistics research writing at IMSIU from 2022 to 2025 will be presented and discussed in this specific section. The researcher integrated both results and interpretation to pinpoint changes in collocational frequency, grammatical structure, thematic orientation, and academic discourse maturity across cohorts.

A. Overall Diachronic Trends in Collocational Use (2022–2025)

It was quite clear that across the four-year period, the corpus showed a clear developmental trajectory in collocational use as it shifted from formulaic and pedagogically oriented expressions toward disciplinary, technological, and interdisciplinary phraseology. On the same hand, the density of using collocation increased steadily from 4.7% in 2022 to 6.2% in 2023, approximately 7.4–8.1% in 2024, and peaked at 8.1% in 2025. Such results showed expanding lexical connectivity and more frequent use of multi-word expressions. It was quite obvious that early-stage writing (2022) relied heavily on high-frequency general academic collocations such as language learning, second language, and Saudi students. In contrast, later cohorts increasingly used semi-technical and technical collocations (such as learning outcomes, AI-assisted writing, cognitive development). This reflected growing disciplinary engagement and academic literacy. In addition, this progression aligns with corpus-based research, and indicated that developing writers gradually replaced generic academic language with discipline-specific phraseology due to getting more research experience (Hyland, 2018; Römer & O'Donnell, 2022).

B. 2022 Results & Discussion: 2022 Corpus

In the 2022 sub-corpus, the initial stage of undergraduate academic writing at IMSIU was clearly reflected. The results showed how heavily they used general academic collocations with limited lexical diversity and minimal disciplinary terminology. The following tables showed clearly the dominant collocational patterns and thematic orientation of this cohort. Table 2, for example, displayed the most frequent collocational patterns in the 2022 corpus, and showed exactly students' reliance on general academic expressions rather than discipline-specific terminology.

Table 2: Most Frequent Collocational Patterns in the 2022 Corpus.

Rank	Collocation	Frequency (per 10k words)	Category
1	language learning	42	Noun + Noun
2	second language	35	Adjective + Noun
3	Saudi students	31	Adjective + Noun

4	English language	28	Noun + Noun
5	motivation in learning	21	Noun + Preposition + Noun

It was quite clear the dominance of collocations such as language learning, second language. Saudi students relied on familiar classroom and textbook language rather than specialized applied linguistics terminology. This reflected their early-stage academic socialization as they gave priority to general academic expressions to communicate basic ideas. There was a high frequency of adjective-noun and noun-noun patterns, suggesting limited lexical complexity and a tendency toward descriptive rather than analytical writing. In the following table (Table 3), the thematic distribution of collocational clusters in the 2022 corpus is displayed to illustrate the dominant topics that shaped students' lexical choices.

Table 3: Thematic Distribution of Collocational Clusters in the 2022 Corpus.

Theme	Representative Collocations	Interpretation
Pedagogy and Learning	language learning, English skills, teaching methods	Focus on classroom-based instruction and skill acquisition
Learner Identity	Saudi students, EFL learners, foreign language	Emphasis on local learner context and national identity
Motivation and Attitude	learning motivation, positive attitude, student performance	Reflection of affective and attitudinal dimensions of SLA

In consideration of the thematic distribution, 2022 cohort was found to write about pedagogy, learner identity, and motivation. This observation surpassed the view of early academic writing as being based on personal experience of learning, problems in the classroom (Paquot, 2019). This emphasized absence of focus on the international research discourses, but this emphasis on the identity of the learner and the motivation was also a manifestation of local contextual orientation and suggested that students were writing within a familiar frame of reference. The summary of the main findings of the corpus, 2022, can be found in Table 4. It has indicated the lexical, grammatical and pedagogical features of the cohort.

Table 4: Summary of Key Findings from the 2022 Corpus.

Feature	Observation	Interpretation
Collocational focus	Repetitive pedagogical combinations	Limited lexical diversity
Lexical sophistication	Low (general academic vocabulary)	Reflects novice EFL writing stage
Dominant grammar types	Adjective + Noun, Noun + Noun	Typical of early academic discourse
Thematic domains	Pedagogy, motivation, learner identity	Classroom-centered orientation
Pedagogical implication	Need for explicit instruction in academic phraseology	Supports targeted corpus-based teaching

Table 4 confirmed the results of the 2022 cohort being an early phase of collocational development. Nevertheless, the students were still in the process of building their academic vocabulary, which was manifested in low lexical sophistication and low collocational density. In general, the 2022 cohort showed early functional writing in academics that was not disciplinary.

C. 2023 Results & Discussion: 2023 Corpus

In the 2023 sub-corpus, it was quite clear that there was an increase in lexical experimentation and contextual awareness, particularly in media-based learning and autonomous learning. The following tables present the dominant collocations and thematic clusters of this cohort. For example, Table 5 showed the dominant collocational patterns in the 2023 corpus, and highlighted a shift toward media-related and experiential expressions.

Table 5: Dominant Collocational Patterns in the 2023 Corpus.

Rank	Collocation	Frequency (per 10k words)	Category
1	English movies	38	Noun + Noun
2	TV shows	34	Noun + Noun
3	language exposure	29	Noun + Noun
4	bilingual students	26	Adjective + Noun
5	self-learning	23	Noun + Noun
6	motivation level	21	Noun + Noun
7	listening skills	20	Noun + Noun

As observed, the emergence of collocations like English movies and TV shows and the exposure to

the language also resulted in students adopting the experience with digital media in academic texts. This observation demonstrated how it was changing the classroom-based writing to more immersive and contextualized research topics. Noun phrases were also more aptly used. This was a sign of an increase in lexical independence as students were able to stop using general academic constructions and shift to more specific and context-dependent language. Table 6 on the other hand showed the thematic clusters in corpus 2023. It demonstrated that the changes of the topic and discourse style were reflected in lexical patterns.

Table 6: Thematic Clusters and Lexical Expansion in the 2023 Corpus.

Theme	Representative Collocations	Interpretation
Media-Based Learning	English movies, TV shows, visual input	Use of entertainment media for vocabulary and listening practice
Autonomous Learning	self-learning, learning independently, motivation level	Emphasis on learner agency and self-regulation
Bilingual Experience	bilingual students, language exposure, native speakers	Recognition of mixed linguistic environments and cross-linguistic influence
Skill Development	listening skills, speaking improvement, vocabulary gain	Focus on measurable language outcomes

In terms of the thematic clusters, the 2023 cohort was mostly represented by media-based learning, autonomous learning, bilingual experience, and skill development. This outcome proved that the students started to connect academic writing with real language experience and online space. The agency of learners and self-regulation were also highlighted as points of such presence of autonomous learning themes. Indeed, both are considered important characteristics of upper-level academic writing. The most important findings of the 2023 corpus were summarized by Table 7. It showed patterns on lexical diversity, grammatical patterns, and collocational density.

Table 7: Summary of Key Findings from the 2023 Corpus.

Feature	Observation	Interpretation
Lexical trend	Increased diversity; inclusion of media- and autonomy-related collocations	Reflects contextual awareness and lexical experimentation
Dominant structures	Noun + Noun; Verb + Noun	Movement toward process-oriented and causal expression
Thematic domains	Media-based learning, autonomous learning, bilingualism	Reflects digital and experiential influences
Collocational density	6.2% (↑ from 4.7%)	Expanding lexical cohesion
Pedagogical implication	Learners developing contextual and autonomous lexical control	Indicates readiness for advanced academic discourse instruction and targeted corpus-based pedagogy

Table 7 confirmed that the collocational density increased to 6.2%, indicating an increase in lexical cohesion and the more frequent occurrence of multi-word expressions. This change in the direction of media-related and autonomous learning collocations indicated that students were starting to generate more context-sensitive academic language. One can say that this stage of development showed the preparedness to more advanced instruction in academic discourse since it involved more specific corpus-based instruction that facilitated the development of lexical knowledge and rhetoric (Sun & Park, 2023).

In the 2024 sub-corpus, an intense shift towards tech-oriented, AI-related academic discourse occurred. Students were more evaluative and research-oriented. This was an indication of their methodological sensitivity and enculturation into the discipline. The tables below summarize the dominant collocations, thematic clusters, and key features which defined this cohort. An example of the most common collocational patterns in the 2024 corpus was presented in Table 8. It demonstrated the development of technology and AI-related lexical decisions.

D. 2024 Results & Discussion: 2024 Corpus

Table 8: Dominant Collocational Patterns in the 2024 Corpus.

Rank	Collocation	Frequency (per 10k words)	Category
1	artificial intelligence	36	Adjective + Noun
2	translation tools	33	Noun + Noun

3	writing evaluation	28	Noun + Noun
4	language technology	24	Noun + Noun
5	AI tools	22	Noun + Noun
6	automated feedback	21	Adjective + Noun
7	learning efficiency	18	Noun + Noun

There was significant dominance of collocations such as artificial intelligence, translation tools, and writing evaluation, and this reflected a strong shift from earlier pedagogical themes toward technology-driven research. This result illustrated how students were engaging with contemporary educational trends and applying disciplinary terminology rather than relying on generic academic phrases. This was apparent in the frequent appearance of noun phrases and compound terms that reflected greater lexical condensation and technical precision, typical of more advanced academic writing (Biber & Gray, 2016). Overall, it can be said that these patterns demonstrated that the 2024 cohort started to align their writing with research-oriented discourse. Table 9 revealed the thematic clusters related to AI in EFL, automated writing evaluation, and digital translation tools. It indicated advanced academic discourse.

Table 9: Thematic Clusters in the 2024 Corpus.

Theme	Representative Collocations	Interpretation
Artificial Intelligence in EFL	AI tools, artificial intelligence, AI applications	Integration of intelligent systems into language learning
Automated Writing Evaluation	writing feedback, automated systems, language accuracy	Focus on AI-supported writing assessment
Digital Translation and Corpus Tools	translation engines, online translators, machine accuracy	Technological mediation in translation learning
Audio-Media Learning	English podcasts, listening improvement, language exposure	Multimodal approaches to input and comprehension

It became very evident that the thematic clusters in Table 9 indicated that students did not simply apply vocabulary related to technology; they also modeled their research on evaluation, automated assessment, and digital mediation. Meanwhile, more emphasis was placed on measurement and effectiveness, suggesting increased awareness of methodology, as indicated in the cluster of automated writing evaluation. The availability of digital translation and corpus tools, on the other hand, indicated the interdisciplinary approach that bridged linguistics and educational technology. All these themes affirmed that the 2024 cohort transcended descriptive writing to the level of evidence-based, analytical academic writing. Table 10 confirmed that the transition toward evaluative and data-driven academic writing was high in the cohort of 2024.

Table 10: Comparison Of Key Features Across The 2022–2024 Corpora.

Feature	2022	2023	2024
Main Focus	Pedagogical & motivational	Media and bilingual exposure	AI & digital tools
Dominant Collocations	language learning, Saudi students	English movies, self-learning	AI tools, writing evaluation
Collocational Density	4.7%	6.2%	8.1%
Lexical Character	General academic	Contextual & experiential	Disciplinary & technical
Rhetorical Style	Descriptive	Analytical	Empirical & data-driven

The evolution of undergraduate writing in this developmental path was well demonstrated in Table 10. It was clear that the 2024 cohort, compared to the 2022 and 2023 ones, was more inclined towards empirical and data-driven writing and its density of collocations was higher (8.1) and the choice of more specific lexicon. This shift in focus from contextual and experience-based themes to AI and digital technology indicated that students were increasingly using more discipline-specific vocabulary and research-based language, which is perceived as one of the main indicators of academic enculturation (Römer & O'Donnell, 2022). Such a comparison enhanced the thesis that the writing of the students became more advanced and corresponded with the trends in the world academic writing. Moreover, a significant change in the 2024 sub-corpus was the move of the academic writing more to technology-oriented and research-oriented writing. These prevailing collocational patterns were focusing on artificial intelligence, automated feedback and digital tools. They were also indicative of a stronger alignment with current trends in educational technology. The key results of the 2024 corpus were

summarized in Table 11, and the lexical, structural, and thematic changes defining this generation were highlighted.

Table 11: Summary Of Key Findings from the 2024 Corpus.

Feature	Observation	Interpretation
Lexical trend	Integration of technical collocations and AI terminology	Reflects disciplinary and conceptual advancement
Collocational density	8.1%	Significant increase in lexical connectivity
Dominant grammar types	Noun + Noun; Adjective + Noun	Reflects lexical condensation and precision
Thematic domains	AI, automated feedback, translation, digital media	Marks interdisciplinary engagement
Pedagogical implication	Students demonstrate near-professional academic discourse	Evidence of lexical and cognitive growth

As it was demonstrated in Table 11, the 2024 cohort revealed a significant rise in collocational density (8.1%). This was a sign of more lexical connectivity and more common use of multi-word expressions. Besides, the widespread use of technical collocations and AI terms demonstrated that there was a shift towards disciplinary and conceptual progress. Conversely, the prevalence of noun phrases was a confirmation that there was further condensation of lexical characteristics of higher academic writing. The formation of the interdisciplinary research orientations was also conditioned by the presence of the obvious thematic focus on the AI, automated feedback, translation, and digital media. In general, these results indicated that in 2024, such students were able to use near-professional academic discourse. This in fact implied serious lexical and cognitive developments in contrast to previous cohorts.

E. 2025 Results & Discussion: 2025 Corpus

The highest level of collocational development was the one that took place in the 2025 sub-corpus. Students were interdisciplinary in their use of AI, ethics, and cognitive science in this sub-corpus. This was the sign of well-developed scholarly positioning and research preparedness. High-frequency collocations of AI-assisted writing and ethical implications were evident in Table 12 and this represented high disciplinary and conceptual maturity.

Table 12: Dominant Collocational Patterns In AI-Assisted Writing Contexts.

Rank	Collocation	Frequency (per 10k words)	Pattern	Typical Context
1	AI-assisted writing	37	Adj + Noun	Integration of generative tools into composition tasks
2	automated feedback	33	Adj + Noun	Evaluation of AI systems for writing support
3	cognitive development	28	Adj + Noun	Link between language learning and neuro-cognition
4	autism spectrum	26	Noun + Noun	Focus on language acquisition in ASD contexts
5	learning processes	24	Noun + Noun	Broader pedagogical discussion
6	language processing	22	Noun + Noun	Neuro-linguistic perspective
7	student interaction	21	Noun + Noun	Collaborative learning in AI environments
8	ethical implications	19	Adj + Noun	Concerns about AI fairness and bias
9	teacher supervision	17	Noun + Noun	Human oversight in AI-mediated learning
10	data-driven learning	15	Adj + Noun	Adoption of analytics-based pedagogy

Domination of AI-related collocations (e.g., AI-assisted writing, automated feedback) was present. This showed that students were not merely informed about current technological trends but they could also be able to explain to an academic context what the implication of the same would be. Furthermore, the use of cognitive and neuro-linguistic terms (cognitive development, language processing) demonstrated that it had been transformed into a more interdisciplinary investigation, rather than only pedagogical issues, and this indicated the presence of higher-level conceptual thinking and research direction. Furthermore, the critical awareness of AI limitations and ethical responsibility was rather evident since the use of the evaluative collocations like ethical implications and teacher supervisions was also evident. This tendency largely corresponded to the discussion of the stance and disciplinary positioning by Hyland and Tse (2023) in which the writers were more authoritative and more evaluative. Thus, significant themes such as AI-mediated writing, cognitive research, ethics, and data-driven pedagogy were pointed out in Table 13.

Table 13: Thematic Clusters In AI-Assisted Writing Contexts.

Theme	Representative Collocations	Interpretation
AI-Mediated Writing and Assessment	AI-assisted writing, automated feedback, teacher supervision	Integration of AI tools for drafting and evaluating student writing
Cognitive and Neuro-Linguistic Research	cognitive development, language processing, autism spectrum	Expansion into interdisciplinary topics linking language and brain functions
Ethics and Human Oversight	ethical implications, data privacy, teacher supervision	Critical reflection on responsible AI use in education
Analytic and Data-Driven Pedagogy	data-driven learning, research-based practice, measurable improvement	Orientation toward evidence and quantification of outcomes

On the other hand, the thematic clusters in Table 13 showed that 2025 students were able to expand their focus beyond technology to include its cognitive and ethical dimensions. This finding indicated students' high maturation in academic argumentation as they did not only describe phenomena but they also evaluated implications which is considered a key marker of disciplinary enculturation (Römer & O'Donnell, 2022). The presence of such data-driven pedagogy themes helped the research to know that learners were able to frame their research in terms of measurable outcomes and evidence-based practices. Therefore, Table 14 reflected clearly the peak collocational density and advanced syntactic structures, and confirmed that the 2025 cohort reached research-ready academic literacy.

Table 14: Summary of Key Findings from the 2025 Corpus.

Feature	Observation	Interpretation
Lexical trend	Integration of AI, cognitive, and ethical collocations	Culmination of disciplinary and conceptual maturity
Dominant structures	Complex N + Prep + N and Verb + that-clauses	Advanced syntactic and rhetorical control
Thematic domains	AI-assisted learning, neuro-linguistics, ethics	Interdisciplinary orientation
Collocational density	8.1% (↑ from 7.4%)	Peak lexical sophistication
Pedagogical implication	Learners demonstrate research-ready academic literacy	Supports integration of AI literacy in EFL curricula

This significant increase in collocational density to 8.1% demonstrated richer lexical connectivity and more frequent use of multi-word expressions. There was also dominance of complex structures such as N + Prep + N and Verb + that-clauses. This revealed a shift toward more sophisticated academic reasoning and argumentation, typical of postgraduate-level writing (Biber & Gray, 2016). Such important results showed that students were able to move from descriptive to analytical writing, and they also demonstrated stronger disciplinary voice and research competence.

F. Summary Of Diachronic Development

It can be summarized in this subsection the findings reached through this four-stage developmental trajectory:

1) The 2022 represented the foundational stage which was characterized with formulaic, pedagogical, and locally grounded collocations
2) The 2023 represented the experiential stage which was distinguished with media-based and autonomy-oriented lexical expansion
3) The 2024 represented the analytical stage which was marked by technology-focused, evaluative, and empirical discourse
4) The 2025 represented the integrative stage which used interdisciplinary, ethical, and research-driven academic language

It can be said that the four-year analysis revealed a clear developmental trajectory from foundational academic writing toward interdisciplinary, AI-integrated academic discourse.

5. CONCLUSION AND PEDAGOGICAL IMPLICATIONS

A. Summary Of Findings

5) The authors in the current study analyzed lexical and collocational reaches in four categories of undergraduate applied linguistics research writing at IMSIU (2022-2025). In this diachronic study, a clear change in lexical sophistication, grammatical complexity, and disciplinary engagement was evident. First of all, the 2022 cohort was a preparation phase characterized by the reliance on high-frequency pedagogical collocations, such as language-learning materials. This stage indicated that the beginner had mastered academic discourse rules. Secondly, academic writing of students advanced to an experience level in 2023 because they began to use more expressions

referring to the media and autonomy (self-learning, TV shows). It implied greater contextual sensitivity and experimentation of lexicon. Thirdly, the 2024 cohort had entered an analytical stage that was characterized by the emergence of semi-technical collocations in technology and assessment (learning outcomes, digital tools, translation engines). This phase was a transition to evidence-based scholarly writing. Fourthly, the 2025 study showed that students could access an integrative stage by coping with lexical resources related to AI, cognition, and ethics (e.g., writing with AI assistance, cognitive growth, ethical concerns). This was a sign of complete involvement in interdisciplinary scholarly conversation. In these four years, the collocational density grew between 4.7 and 8.1, and grammatical structures appeared not only in simple adjective-noun combinations but also in more intricate noun-preposition-noun and verb + that-clauses. These results indicated a developmental transition in which academic phrases were mechanically copied into conceptual, disciplinary, and rhetorical complexity.

6) In general, the results of the present study might be beneficial to corpus-related applied linguistics in the following aspects. First, the study could provide a rare diachronic perspective on undergraduate lexical development within a single Saudi institutional context. By doing so, it complements predominantly cross-sectional learner-corpus research (Paquot, 2019). Second, the results of the current study provided empirical evidence for integrating AI literacy into academic discourse, illustrating how emerging technologies reshape lexical repertoires and writing practices. Third, the findings showed the local–global interface in Saudi EFL academic writing, and showed how learners progressively aligned with international academic norms while retaining context-specific thematic orientations.

B. Pedagogical Implications

This study identified different diachronic patterns. They suggested several pedagogical priorities for undergraduate applied linguistics programs. First, it must be stated that explicit collocational instruction should be integrated into academic writing courses. Second, instructors should encourage using AI-enhanced writing pedagogy through guided and supervised implementation of generative tools (such as Grammarly and ChatGPT). This could foster critical awareness rather than unreflective dependence. Third, universities need to promote data-driven learning approaches, such as small-scale learner-corpus projects to enhance metalinguistic reflection and learner autonomy. This could enable students to discover lexical patterns in authentic data. In addition, curricula designers and developers should incorporate ethical and metalinguistic training such as modules on research integrity, AI ethics, and academic honesty to make sure that linguistic development is aligned with responsible academic practice.

C. Limitations And Future Research

The current study was limited to undergraduate research writing from one English department at IMSIU over a four-year period. Therefore, the results could not be generalized across other different educational institutions or proficiency levels. There is an urgent need to more future research that could extend this work by incorporating postgraduate corpora, cross-institutional comparisons, and multilingual datasets. As for the methodology adopted in this research, other research could employ mixed-methods approaches that use both corpus analysis and interviews, questionnaires, or think-aloud protocols. This could provide deeper insight into learners' metacognitive awareness of collocations and lexical choices (Vyatkina, 2020). More investigation into task design, digital writing environments, and AI-mediated academic practices is also needed to improve understanding of how emerging technologies have impacted lexical development (Ali El Deen & Mahmoud, 2025).

What was fairly evident is that in the timeframe between 2022 and 2025, undergraduate students evolved their application of collocations to reproduce familiar lexical frames to build knowledge using data-driven, ethically sensitive, and AI-mediated discourses. It is a giant shift in both Saudi and international EFL settings (Du et al., 2022; Alfaqara, 2023; Alenizi & Adawi, 2024). On the whole, the findings of the study suggested that the construction of collocational competence as a linguistic and epistemic accomplishment was capable of facilitating the necessity of long-term instruction based on the corpus to facilitate the enhancement of advanced, discipline-specific academic literacy (Hyland, 2021).

Declaration of AI Using Assisted Technologies: DeepSeek and Grammarly were used for language

proofreading in this article. Neither tool was involved in idea generation or content development. Any remaining errors are solely the authors' responsibility.

Funding Statement: This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-DDRSP2602).

Data Availability: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Competing Interests: The author declares no competing interests

REFERENCES

Abdellah, A.S. (2015). The Effect of a Program Based on the Lexical Approach on Developing English Majors' Use of Collocations. Journal of Language Teaching and Research. Vol. 6 (4). http://dx.doi.org/10.17507/jltr.0604.08.

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://dx.doi.org/10.2139/ssrn.4666773.

Alfaqara, W. M. (2021). Learning and comprehension of English grammatical collocations with prepositions by Jordanian EFL learners at Mutah University. Journal of Language and Linguistic Studies. 17(3), 1748-1762 https://www.jlls.org/index.php/jlls/article/view/4233/1215

Ali El Deen, A.M.M & Mahmoud, M.M.A. (2025). Emotion regulation strategies and emotional experiences in blended EFL contexts: A case study of Saudi university learners. Forum for Linguistic Studies, 7(11), 728–741. http://doi.org/10.30564/fls.v7i11.11382.

Al-Mubireek, S., Mahmoud, M. M. A., Ali El-Deen, A.M.M, Moumene, A., Younis, A.(2022). Investigating EFL instructors' perceptions of using Blackboard in TEFL at IAU preparatory year. International Journal of Learning, Teaching and Educational Research, 22(1), Article 19. https://doi.org/10.26803/ijlter.22.1.19.

Alsulayyi, M. (2024). The use of grammatical collocations by advanced Saudi EFL learners in the UK and KSA. International Journal of English Linguistics. 5(1) 32-43 https://doi.org/10.5539/ijel.v5n1p32

Bestgen, Y., & Granger, S. (2018). Quantifying the development of phraseological complexity in L2 English writing: An automated approach. Journal of Second Language Writing, 41, 19–30. https://doi.org/10.1016/j.jslw.2014.09.004.

Biber, D., & Gray, B. (2016). Grammatical complexity in academic English: Linguistic change in writing. Cambridge University Press.

Birhan, A.T., Nurie, Y. (2024). Developing engineering students' engagement in academic writing classes using corpus based instruction. Asian Pacific Journal of Second and Foreign Language Education, 9, Article 11. https://doi.org/10.1186/s40862-023-00232-2

Du, X., Afzaal, M., & Al Fadda, H. (2022). Collocation use in EFL learners' writing across multiple language proficiencies: A corpus driven study. Frontiers in Psychology, 13, 752134. https://doi.org/10.3389/fpsyg.2022.752134

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? International Review of Applied Linguistics in Language, 47(2), 157-177

Gries, S. T. (2021). Analyzing linguistic data: A practical introduction to statistics using R (2nd ed.). Cambridge University Press.

Hoey, M. (2005). Lexical priming: A new theory of words and language. Routledge.

Hyland, K. (2018). Metadiscourse: Exploring interaction in writing (2nd ed.). Bloomsbury.

Hyland, K. (2021). Second language writing. Cambridge University Press.

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

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). https://doi.org/10.31605/ilere.v2i1.5033

Nesselhauf, N. (2003) The use of collocations by advanced learners of English and some implications for teaching. Applied Linguistics, 24(2), 223–242. https://doi.org/10.1093/applin/24.2.223.

Paquot, M. (2019). Phraseology in learner writing: A corpus-based interdisciplinary perspective. Cambridge University Press.

Pellicer-Sanchez, A. (2017). Learning L2 collocations incidentally from reading. Language Teaching Research. Vol. 21 (3) https://doi.org/10.1177/1362168815618428.

Römer, U. (2020). Language learning from a corpus linguistic perspective: Recurrent patterns and implications for teaching. Annual Review of Applied Linguistics, 40, 80–100. https://doi.org/10.1017/S0267190520000095

Römer, U., & O'Donnell, M. (2022). Investigating academic phraseology across time and disciplines. Corpora, 17(3), 431–456. https://doi.org/10.3366/cor.2022.0256

Sanosi, B.A. (2025). The present study answers the research question: A corpus-based analysis of collocate directionality in academic English writing. Heliyon. Vol.11 (2) https://doi.org/10.1016/j.heliyon.2025.e42088.

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

Sun, W., & Park, E. (2023). EFL Learners' Collocation Acquisition and Learning in Corpus-Based Instruction: A Systematic Review. Sustainability, 15(17), 13242. https://doi.org/10.3390/su151713242

Vyatkina, N. (2020). Formulaic language development in L2 German: Longitudinal learner corpus evidence. Language Learning, 70(S1), 150–182. https://doi.org/10.1111/lang.12366