

A CORPUS-BASED STUDY OF VOCABULARY USE IN PAKISTANI HIGH SCHOOL STUDENTS' ENGLISH ESSAYS

Ayesha Zubair

MPhil Linguistics University of Sahiwal, Punjab, Pakistan

Email: ayeshauos20@gmail.com

Alina Ahmad

MPhil Linguistics, University of Sahiwal, Punjab, Pakistan

Email: leenyahmad1@gmail.com

Siddiqua Bibi

MPhil Linguistics, University of Sahiwal, Punjab, Pakistan

Email: siddiquabeyg05@gmail.com

Muqaddar Zaka (Corresponding Author)

MPhil Linguistics, University of Sahiwal, Punjab, Pakistan

Email: muqadarzaka@gmail.com

Abstract

This study investigates the vocabulary use of Pakistani high school students in English essay writing using a corpus-based approach. The purpose of the study is to identify common vocabulary patterns, frequency of word use, and the level of academic vocabulary used by students. A small learner corpus was compiled, consisting of essays written by high school students from Pakistani schools. The essays were converted into digital text and analyzed with corpus analysis software such as AntConc. The analysis focused on word frequency, lexical diversity, and the use of academic vocabulary based on the Academic Word List. The findings reveal that students frequently rely on common everyday vocabulary and show limited use of academic words. Repetition of certain words and limited lexical variation were also observed. The results highlight the need to improve students' academic vocabulary and writing skills. This study contributes to research in Corpus Linguistics and provides useful insights for teachers and researchers working in English language learning in Pakistan.

Keywords: *vocabulary, corpus-based approach, Essay writing, Academic Word List, KWIC, collocation, concordance.*

Introduction

Vocabulary is one of the most important components of language learning and academic writing. A strong vocabulary allows learners to express their ideas clearly, accurately, and effectively. In academic contexts, students are expected to use appropriate and varied vocabulary to present arguments, explain concepts, and develop well-structured essays. However, learners of English as a second language often face difficulties in using a wide range of vocabulary in their writing. Limited vocabulary knowledge may lead students to rely on simple and repetitive words, which can affect the quality of their academic writing.

In higher education, English plays a significant role in Pakistan because it is widely used as a medium of instruction in many schools. High school students are frequently required to complete written assignments, essays, and reports in English. Despite studying English for many years, many students still struggle to use appropriate academic vocabulary in their writing. Their essays may show repetition of common words, lack of lexical variety, and limited use of academic

expressions. These problems highlight the need to examine how students actually use vocabulary in their academic writing.

One useful method for investigating language patterns is **Corpus Linguistics**. Corpus linguistics involves the analysis of large collections of authentic texts, known as corpora, with the help of computer software. This approach allows researchers to study real language use by identifying patterns such as word frequency, collocations, and lexical diversity. By analyzing a collection of students' essays as a learner corpus, researchers can better understand the vocabulary choices and writing practices of language learners.

Several corpus analysis tools are commonly used for this purpose, including **AntConc**, which helps researchers identify frequent words, concordances, and lexical patterns in a text collection. Through such tools, it becomes easier to observe how often certain words appear and how students organize vocabulary in their essays.

Another important aspect of academic writing is the use of academic vocabulary. Academic vocabulary refers to words that are commonly used in scholarly texts across different disciplines. A well-known resource for identifying such vocabulary is the **Academic Word List**, which includes words that frequently occur in academic texts but are less common in everyday language. The use of words from this list can indicate the level of academic language competence among students.

Previous studies in second language writing have shown that learners often rely heavily on general vocabulary while using fewer academic words. This can limit the effectiveness of their writing and reduce the clarity of their arguments. Therefore, examining vocabulary use in students' essays is important for understanding their linguistic abilities and identifying areas that require improvement.

In the context of Pakistani schools, there is still limited corpus-based research focusing on vocabulary use in high school students' writing. Investigating students' essays through a corpus-based approach can provide valuable insights into their vocabulary patterns, lexical diversity, and use of academic vocabulary. Such findings can help teachers design more effective strategies for teaching vocabulary and improving students' academic writing skills.

Therefore, the present study aims to analyze the vocabulary used in essays written by Pakistani high school students by developing a small learner corpus and examining it through corpus analysis techniques. The study focuses on identifying common vocabulary patterns, measuring lexical diversity, and investigating the presence of academic vocabulary in students' essays.

Research Objectives

1. To examine the vocabulary patterns used in essays written by Pakistani high school students.
2. To identify the most frequently used words in students' essays using corpus analysis tools such as AntConc.
3. To investigate the use of academic vocabulary in students' essays based on the Academic Word List.

Research Questions

1. What vocabulary patterns are found in essays written by Pakistani high school students?
2. What are the most frequently used words in high school students' essays?
3. To what extent do Pakistani high school students use academic vocabulary from the Academic Word List in their essays?

Literature Review

Vocabulary plays a central role in second language learning and academic writing. Researchers in the fields of Corpus Linguistics and **second language acquisition** have emphasized that learners' ability to use appropriate vocabulary significantly affects the quality of their academic writing. Corpus-based studies provide an effective method for examining vocabulary patterns in authentic texts because they allow researchers to analyze large collections of language data using computational tools.

One important concept in academic writing research is **academic vocabulary**. Academic vocabulary refers to words that frequently appear in academic texts across different disciplines. A widely recognized resource for identifying such words is the Academic Word List, which contains 570-word families commonly used in academic discourse. These words are considered essential for students who aim to succeed in higher education because they occur frequently in academic texts but less frequently in everyday language. Studies have shown that knowledge and use of academic vocabulary play an important role in improving students' reading comprehension and writing abilities.

Corpus-based approaches have been widely used to examine vocabulary usage in academic texts and learner writing. In corpus linguistics research, a corpus is a large electronic collection of texts that can be analyzed using computer software to identify patterns of language use. Corpus analysis tools help researchers examine features such as word frequency, lexical diversity, collocations, and distribution of vocabulary in texts. This method provides empirical evidence about how language is actually used rather than relying only on intuition or theoretical assumptions.

Several studies have used corpus-based methods to investigate vocabulary patterns in academic writing. For example, a corpus-driven study of student essays analyzed the frequency and distribution of academic vocabulary in learners' writing. The results indicated that students tend to rely heavily on a limited set of common words while using fewer discipline-specific academic terms. The study also showed that some academic words are overused while many others appear rarely in student essays. These findings highlight the importance of explicit vocabulary instruction to help learners expand their lexical repertoire and improve the quality of their academic writing.

Another corpus-based study examined the use of academic vocabulary in first-year ESL high school students' writing. The research analyzed essays written by university students and investigated how academic vocabulary density and diversity relate to writing quality. The findings revealed that students who used a wider range of academic vocabulary tended to produce higher-quality essays. The study also found that vocabulary diversity was strongly connected to students' writing performance, suggesting that the ability to use varied academic vocabulary is an important indicator of writing proficiency.

Researchers have also investigated the role of corpus-based approaches in vocabulary learning. A study on university students' vocabulary acquisition demonstrated that using corpus-based learning methods significantly improved students' ability to understand and use academic vocabulary. Students who engaged with corpus materials were able to recognize and apply relevant vocabulary more effectively than those who relied on traditional learning methods. The study concluded that corpus-based learning can support vocabulary development and enhance language learning in academic contexts.

Corpus-based vocabulary research has also been conducted in specific educational contexts. For instance, a corpus-based analysis of English textbooks used in colleges in Punjab examined the frequency and distribution of academic words in instructional materials. The results showed that the coverage of academic vocabulary in the textbooks was relatively limited, suggesting that students may not receive sufficient exposure to academic vocabulary through textbooks alone. The researchers emphasized the need for better vocabulary selection and instructional strategies to support students' language development.

In addition to analyzing student writing, corpus-based studies have also explored vocabulary patterns in academic research articles across different disciplines. For example, a large corpus study of chemistry research articles analyzed millions of words to identify frequently used academic vocabulary in scientific writing. The study found that discipline-specific vocabulary plays an important role in academic communication and that specialized word lists can help students better understand and produce academic texts in their fields.

Another important line of research focuses on how corpus tools can support language learning and vocabulary development. Studies have shown that exposure to corpus data helps learners understand word meanings, collocations, and grammatical patterns more effectively. Through corpus analysis, students can observe authentic language examples and develop a deeper understanding of vocabulary usage in context. As a result, corpus-based instruction has been recognized as an effective method for improving vocabulary knowledge and language skills.

Despite the growing body of research in corpus linguistics and vocabulary studies, relatively limited research has focused specifically on vocabulary patterns in the academic writing of Pakistani high school students. Most previous studies have examined learners in other educational contexts or have focused on textbooks and academic research articles rather than student essays. Therefore, there is a need for further research that investigates vocabulary use in authentic student writing within the Pakistani higher education context.

The present study aims to address this gap by conducting a corpus-based analysis of vocabulary use in essays written by Pakistani high school students. By examining word frequency, lexical diversity, and the use of academic vocabulary, this research seeks to provide insights into the vocabulary patterns and limitations present in student writing. The findings may help educators understand learners' vocabulary needs and design more effective teaching strategies for improving academic writing skills.

Theoretical Framework

This study is grounded in the principles of **Corpus Linguistics**, which focuses on the systematic analysis of language through large collections of authentic texts known as corpora. Corpus linguistics allows researchers to investigate patterns of language use by examining real examples of written or spoken language. Using computational tools, researchers can identify linguistic patterns such as word frequency, collocations, lexical diversity, and the distribution of vocabulary in texts. This approach provides empirical evidence about language use rather than relying solely on intuition or theoretical assumptions.

Corpus linguistics has become an important method for studying vocabulary use in second language writing. Through corpus analysis, researchers can examine how learners use vocabulary in authentic contexts and identify common patterns and limitations in their language production. This method is particularly useful for investigating learner language because it provides objective and quantifiable data about language use.

Another important concept in this study is **academic vocabulary**, which refers to words that frequently occur in academic texts across different fields of study. A widely used reference for academic vocabulary is the **Academic Word List**, which consists of 570 word families that are commonly found in academic writing. These words are important for students because they help them express ideas more precisely and participate effectively in academic discourse.

The use of academic vocabulary is often considered an indicator of students' academic writing proficiency. Students who use a wider range of academic vocabulary are generally able to produce more formal and effective academic texts. Therefore, examining the presence and frequency of these words in student essays can provide insights into students' academic language competence.

In addition, the study also draws on concepts from **Second Language Acquisition**, particularly the idea that vocabulary development is essential for successful language learning. Learners gradually acquire vocabulary through exposure to language input and practice in using words in different contexts. By analyzing learner corpora, researchers can identify the strengths and weaknesses in students' vocabulary use and provide recommendations for improving language teaching practices.

Therefore, this study integrates corpus linguistic methods with theories of second language vocabulary development to analyze vocabulary use in essays written by Pakistani high school students. The framework allows the researcher to systematically examine lexical patterns and academic vocabulary usage in students' writing.

Methodology

Research Design

This study employs a **quantitative corpus-based research design** to investigate vocabulary use in essays written by Pakistani high school students. Corpus-based research involves collecting a set of authentic texts and analyzing them using computational tools in order to identify patterns of language use.

Data Collection

The data for this study consist of essays written by high school students studying in Pakistani schools. A small learner corpus was compiled by collecting approximately **30-50 essays** written in English by high school students. The essays were collected from classroom assignments and academic writing tasks.

All essays were converted into digital text format in order to create a corpus suitable for computer-assisted analysis. Personal information about students was removed to maintain anonymity and confidentiality.

Corpus Compilation

After collecting the essays, the texts were compiled into a single electronic corpus. The corpus was cleaned and prepared by removing unnecessary symbols, formatting issues, and non-text elements. This process ensured that the texts were suitable for analysis using corpus analysis software.

Data Analysis Tools

The corpus is analyzed using **AntConc**, which is a widely used tool in corpus linguistics research. AntConc provides several functions that help researchers examine linguistic patterns in texts, including:

- Word frequency lists

- Concordance lines
- Keyword analysis
- Collocation patterns

These features allow researchers to identify frequently used words and observe how vocabulary appears in different contexts within the corpus.





Target Corpus
Name: my_corpus
Files: 1
Tokens: 3405
corpus.docx

KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud ChatAI

Total Hits: 64 Page Size 100 hits 1 to 64 of 64 hits

File	Left Context	Hit	Right Context
1 corpus.docx	to works hard can overcome challenges and achieve their goals.	students	should always remember that success requires effort and dedica
7 corpus.docx	become responsible and useful members of society. Therefore,	students	should always work hard in their studies and take
3 corpus.docx	is a key quality that helps people succeed in life.	Students	should develop honesty as a habit from a young
4 corpus.docx	key quality that leads to personal, academic, and social success.	Students	should develop honesty as a habit from a young
5 corpus.docx	who waste time often regret their actions later in life.	Students	should learn how to manage their time properly so
6 corpus.docx	relationships but also for academic and professional success.	Students	should want to be honest in their studies, behavior,
7 corpus.docx	clothes. These habits helps prevent many illnesses and infections.	Students	should also keep their classrooms and school areas clean
9 corpus.docx	Both physical and mental health should takes care of regularly.	Students	should eat healthy food, exercise, and avoid unhealthy habits.
8 corpus.docx	knowledge, and develop better thinking abilities. Therefore,	students	should make reading books a regular part of their
10 corpus.docx	discipline is essential for personal growth and academic success.	Students	should practice discipline in all areas of life. It
11 corpus.docx	many problems, such as weakness, illness, and lack of energy.	Students	should take care of their health to perform well.
12 corpus.docx	platforms helps students study anytime and anywhere. However,	students	should use technology responsibly. Overuse of mobile phones, ga
13 corpus.docx	by giving examples, explanations and activities. They encourage	students	to ask questions and participate in class discussions. A
14 corpus.docx	and gives examples to make it easier. She also encourages	students	to ask questions, which makes the class more interesting.
15 corpus.docx	ing education, guidance, and support. A good teacher can inspire	students	to work hard and become successful individuals. Essay 10: My
16 corpus.docx	ive knowledge, guidance, and support. A good teacher motivates	students	to work hard, think creatively, and achieve their goals.
17 corpus.docx	lucky to have Ahmed as my friend. True friendship teaches	students	to be loyal, caring, and understanding, which is important.
18 corpus.docx	concentrates better in class. Families and schools should guides	students	to develop healthy habits from an early age. In
18 corpus.docx	all knowledge. It provides a safe and supportive environment for	students	to learn, academically and personally. I prefer books, in...

Search Query Words Case Regex Results Set All hits Context Size 10 tokens

Sort Options Sort to right Sort 1: 1B Sort 2: 2B Sort 3: 3B Order by freq

Target Corpus
Name: my_corpus
Files: 1
Tokens: 3405
corpus.docx

KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud ChatAI

Total Hits: 9 Page Size 100 hits 1 to 9 of 9 hits

File	Left Context	Hit	Right Context
1 corpus.docx	cleanliness in their homes and communities. By keeping our surroundings	clean	we can create a healthier and more beautiful world.
2 corpus.docx	for a healthy and comfortable life. When our surroundings are	clean	we feel happy and safe. A dirty environment can
3 corpus.docx	often organizes campaigns to encourage people to keep their surroundings	clean	and protect the environment. Personal cleanliness is also very
4 corpus.docx	infections. Students should also keep their classrooms and school areas	clean	by throwing garbage in proper places. When people take
5 corpus.docx	always try to keep their homes, schools, and public places	clean	Clean streets, parks, and classrooms create a pleasant atmosphere
6 corpus.docx	such as washing their hands regularly, taking baths, and wearing	clean	clothes. These habits helps prevent many illnesses and infections.
7 corpus.docx	conclusion, cleanliness is an important responsibility for every individual. A	clean	environment helps people stay healthy and live comfortably. Everyone
8 corpus.docx	try to keep their homes, schools, and public places clean.	Clean	streets, parks, and classrooms create a pleasant atmosphere where
9 corpus.docx	without problems. Maintaining good health requires eating nutritious food, drinking	clean	water, exercising regularly, and getting proper rest. People who

Search Query Words Case Pages Results Set All hits Context Size 10 tokens

Sort Options Sort to right Sort 1: 1B Sort 2: 2B Sort 3: 3B Order by freq

Progress 100%

Corpus Description

The corpus analyzed in AntConc consists of one text file (corpus.docx) containing 3405 tokens. The KWIC (Key Word in Context) concordance function was used to analyze how particular lexical items appear within their surrounding context. This tool displays the search term in the center, with left and right context, allowing researchers to examine patterns of word usage, collocations, and semantic tendencies.

The concordance analysis focuses on two key lexical items found in the corpus: “students” and “clean.”

Concordance Analysis of the Word “Students”

The keyword “students” appears 64 times in the corpus. The KWIC results show that the word frequently occurs at the beginning of sentences, indicating that it functions primarily as the grammatical subject in many statements. This suggests that the texts are largely student-centered and instructional.

Syntactic Pattern

A major syntactic pattern observed is:

Students + modal verb + verb phrase

For example:

- Students **should always remember** that success requires effort.
- Students **should work hard** in their studies.
- Students **should develop** honesty as a habit.
- Students **should learn** how to manage their time.

The repeated use of the modal verb “should” indicate that the texts are prescriptive and advisory. In other words, the corpus is not merely describing students but guiding them about appropriate behavior, attitudes, and responsibilities.

This linguistic pattern reflects didactic discourse, which is commonly found in educational essays, textbooks, and instructional materials.

Lexical Collocations with “Students”

The concordance lines reveal several verbs that frequently appear with the word **students**. These verbs can be grouped into different semantic categories.

Academic Development

Several verbs are related to **learning and education**, such as:

- learn
- read
- improve
- study
- ask

Examples from the corpus include:

- Students **can improve** their language skills.
- Students **learn about** different cultures.
- Students **ask questions** in class discussions.

These collocations show that the corpus strongly emphasizes academic engagement and intellectual growth.

Moral and Personal Development

Another important group of verbs relates to character building and ethical behavior.

Examples include:

- develop discipline
- develop honesty
- respect teachers
- respect parents
- work hard

Example concordance lines:

- Students **develop discipline and moral values.**
- Students **learn to respect teachers and parents.**
- Students **should develop honesty from a young age.**

This indicates that the corpus includes moral and value-based educational discourse, where language is used to promote ethical behavior and responsible citizenship.

Social and Behavioral Skills

The concordance lines also highlight verbs related to social interaction and cooperation, such as:

- work together
- cooperate
- participate
- communicate

Example lines:

- Students **learn how to work together.**
- Students **participate in class discussions.**

These patterns suggest that the texts encourage collaborative learning and active participation in educational environments.

Semantic Interpretation

The frequent occurrence of verbs related to learning, discipline, respect, and cooperation indicates that the representation of students in the corpus is largely positive and developmental.

Students are portrayed as individuals who:

- develop knowledge and skills
- follow rules and discipline
- build moral values
- contribute positively to society

Thus, the concordance analysis suggests that the corpus focuses on **educational guidance and personal development themes.**

Concordance Analysis of the Word “Clean”

The keyword “clean” appears 9 times in the corpus. Although the frequency is lower than that of “students,” the KWIC results still reveal important patterns of usage.

Lexical Environment

The word clean frequently occurs with nouns referring to places and environments, such as:

- streets
- parks
- classrooms
- homes
- public places
- environment

Examples from the concordance lines include:

- Clean streets, parks, and classrooms create a pleasant atmosphere.
- Students should keep their homes and schools clean.
- A clean environment helps people stay healthy.

These examples show that the word clean is primarily used in relation to environmental hygiene and public cleanliness.

Semantic Themes

The concordance lines reveal two main themes related to the word **clean.**

Environmental Awareness

Several sentences connect cleanliness with environmental protection and public responsibility.

Examples include:

- Clean surroundings create a healthier world.
- Clean environments help protect the community.

This suggests that the corpus encourages environmental awareness and responsible behavior toward public spaces.

Health and Hygiene

Another recurring theme is the connection between cleanliness and health.

Examples include:

- Clean clothes help prevent illnesses.
- Clean environments help people stay healthy.

These examples show that cleanliness is linked with physical well-being and disease prevention.

Linguistic and Thematic Patterns in the Corpus

Based on the concordance results, several broader patterns emerge.

Educational Discourse

The corpus strongly reflects educational discourse, where students are the central participants and learning is the primary theme.

Instructional Language

The frequent use of modal verbs such as *should* indicates that the texts provide instructions, advice, and recommendations.

Moral and Social Values

Many concordance lines promote values such as:

- honesty
- discipline
- respect
- cooperation

This suggests that the texts aim to develop both academic knowledge and moral character.

Health and Environmental Awareness

The occurrence of words related to cleanliness and health indicates that the texts also promote healthy habits and environmental responsibility.

Academic Interpretation for Research Writing

The concordance analysis conducted using AntConc demonstrates that the lexical item *students* occur frequently in the corpus and is commonly followed by modal verbs and action verbs that express advice, guidance, and expectations. This pattern indicates that the corpus contains primarily instructional and educational discourse aimed at guiding students toward academic success and moral development.

Furthermore, the analysis of the word *clean* reveals that it is primarily associated with environmental and health-related contexts. The concordance lines highlight the importance of maintaining clean surroundings, public spaces, and personal hygiene. These linguistic patterns suggest that the texts promote environmental awareness and healthy living practices alongside academic and moral education.

Overall, the corpus reflects a strong emphasis on education, discipline, social responsibility, and personal development, which are commonly found themes in student-oriented educational texts.



Target Corpus: my_corpus
File: 1
Format: JAD5
corpus.docx

Total Hits: 12 Page Size: 100 hits 1 to 12 of 12 hits

File	Left Context	Hit	Right Context
1 corpus.docx	future of students by providing education, guidance, and support. A	good	teacher can inspire students to work hard and become
2 corpus.docx	and support makes learning enjoyable and meaningful. In conclusion, a	good	teacher can positively change the life of a student.
3 corpus.docx	in students' lives. They give knowledge, guidance, and support. A	good	teacher inspires students to work hard, think creatively, and
4 corpus.docx	their health to perform well in studies and daily activities.	good	health also depends on mental well-being. A person
5 corpus.docx	healthy can work, study, and enjoy life without problems. Maintaining	good	health requires eating nutritious food, drinking clean water, exercising
6 corpus.docx	to become more active and skilled. The school also promotes	good	behavior, discipline, and respect among students. In conclusion, my
7 corpus.docx	that. In conclusion, friendship is very valuable in life. A	good	friend supports you in both happy and difficult times.
8 corpus.docx	the environment. Personal cleanliness is also very important for maintaining	good	health. People should develop good hygiene habits such as
9 corpus.docx	also very important for maintaining good health. People should develop	good	hygiene habits such as washing their hands regularly, taking
10 corpus.docx	who are honest in their studies and behavior develop a	good	reputation, and gain the respect of teachers, parents, and
11 corpus.docx	interactions with others. Being honest also helps people build a	good	reputation. An honest student will always gain the respect
12 corpus.docx	useful skills, they can build a better future for themselves.	good	time management helps people complete their tasks in an

Search Query: Words Case Regex Results Set: All hits Context Size: 10 tokens

GOOD Start Adv Search

Sort Options: Sort to right Sort 1: 1R Sort 2: 2R Sort 3: 3R Order by: freq

Progress: 100%

Target Corpus: my_corpus
File: 1
Format: JAD5
corpus.docx

Total Hits: 6 Page Size: 100 hits 1 to 6 of 6 hits

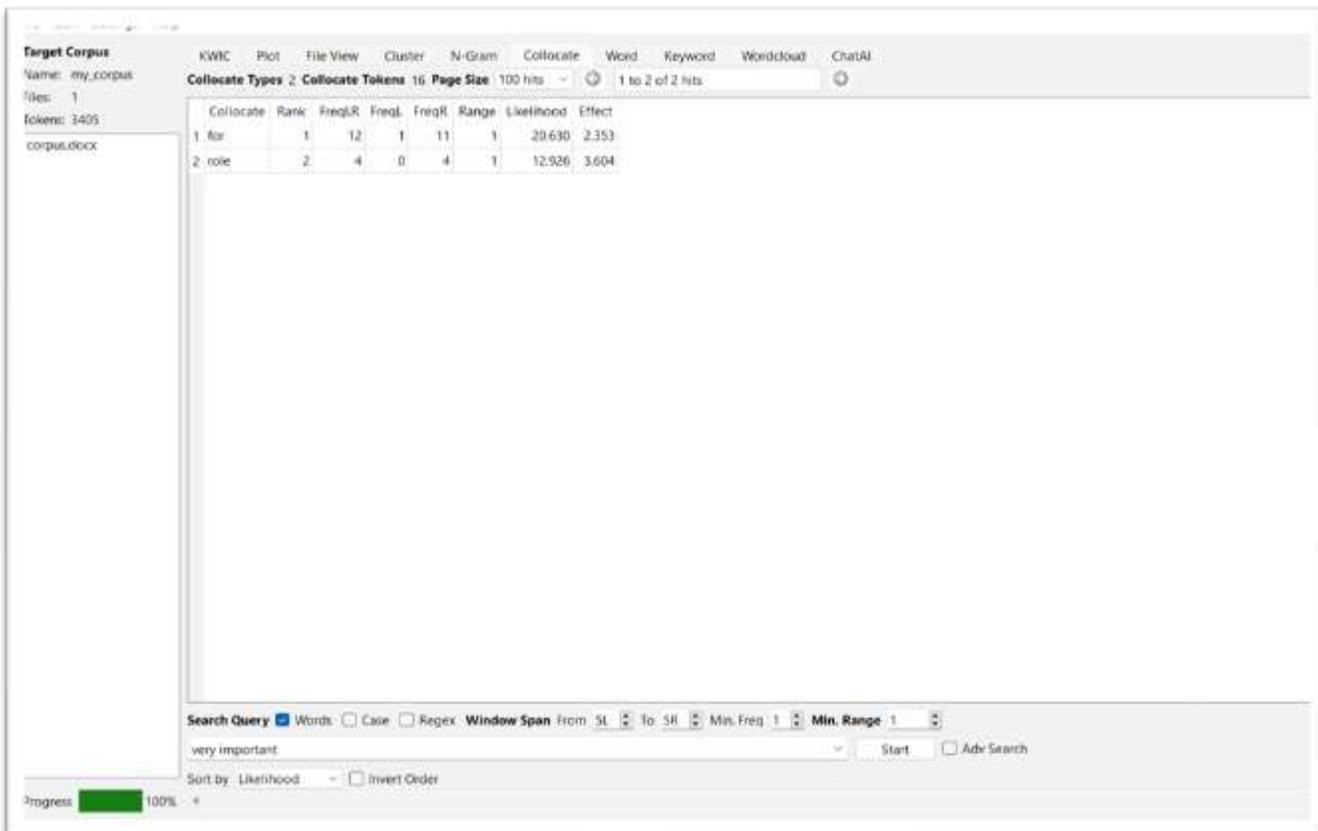
File	Left Context	Hit	Right Context
1 corpus.docx	Essay 1: Importance of Education Education	plays	a very important role in the life of every
2 corpus.docx	behavior, discipline, and respect among students. In conclusion, my school	plays	a very important role in shaping my personality and
3 corpus.docx	media can waste time and reduce productivity. In conclusion, technology	plays	A very important role in education, communication, and daily
4 corpus.docx	their time and fail to reach their goals. Discipline also	plays	an important role in social life. Students learn to
5 corpus.docx	rain falling creates a calm and peaceful feeling. Children often	play	outside in the rain and enjoy making paper boats.
6 corpus.docx	health but also teaches important life skills. A student who	plays	sports regularly becomes more confident, disciplined, and active. Every

Search Query: Words Case Regex Results Set: All hits Context Size: 10 tokens

PLAYS Start Adv Search

Sort Options: Sort to right Sort 1: 1R Sort 2: 2R Sort 3: 3R Order by: freq

Progress: 100%



KWIC Concordance Analysis of the Word “Good”

Description of the Result

The first screenshot displays the KWIC (Key Word in Context) results for the word “good.” The corpus used in this analysis contains 3405 tokens and one document (corpus.docx). AntConc identifies 12 occurrences of the word good within the corpus.

The KWIC interface shows the keyword centered between its left context and right context, allowing the researcher to observe how the word functions within sentences.

Examples from the concordance lines include:

- good teacher can inspire students
- good teacher can positively change the life of a student
- good health also depends on mental well-being
- develop good hygiene habits
- gain the respect of teachers and build a good reputation
- build a better future for themselves

Linguistic Analysis

The concordance lines reveal that “good” functions primarily as an attributive adjective modifying nouns. It is used to evaluate people, behaviors, or conditions positively.

The keyword appears in several semantic domains:

1. Educational Domain

Some collocations are directly related to education and teaching.

Examples:

- good teacher
- good behavior
- good discipline

These expressions emphasize **moral guidance and academic support**, suggesting that teachers are portrayed as influential figures responsible for shaping students' futures.

This pattern reflects a common discourse in educational writing, where teachers are associated with motivation, guidance, and inspiration.

2. Health and Lifestyle Domain

Another group of collocations relates to physical well-being and hygiene.

Examples:

- good health
- good hygiene habits

Here, the adjective “good” is used to describe desirable health practices. The texts encourage behaviors such as maintaining cleanliness, eating nutritious food, and exercising.

This suggests that the essays aim to promote awareness of healthy lifestyles among students.

3. Moral and Social Values

The corpus also contains collocations that highlight personal character and social reputation.

Examples:

- good reputation
- good habits

These phrases indicate a focus on ethical values such as honesty, discipline, and respect. The texts promote the idea that positive character traits lead to social acceptance and success.

Interpretation

The repeated use of “good” as a positive evaluative adjective indicates that the corpus emphasizes normative and moral judgments. Writers use this adjective to distinguish between desirable and undesirable behaviors.

From a discourse perspective, this suggests that the texts are didactic in nature, aiming to guide readers toward appropriate attitudes and behaviors.

Furthermore, the keyword appears mostly in general statements about life and education, which suggests that the texts are likely expository or argumentative essays written by students.

The frequent evaluation using “good” also indicates a simple and direct style typical of learner writing, where writers rely on common adjectives to express positive judgments.

KWIC Concordance Analysis of the Word “Plays”

Description of the Result

The second screenshot presents the KWIC results for the verb “plays.” AntConc identifies 6 occurrences of this word in the corpus.

The concordance lines show that “plays” occurs in sentences discussing the importance of various aspects of life.

Examples include:

- Education plays a very important role in the life of every student.
- Technology plays a very important role in communication.
- Discipline plays an important role in social life.

- Sports plays regularly becomes more confident.

Linguistic Analysis

1. Formulaic Expression

The concordance lines reveal a recurring lexical pattern:

plays + (a / very) + important + role

This phrase is a common academic collocation used to express significance or influence.

Examples:

- plays a very important role
- plays an important role

Such repeated patterns are typical in academic essays, where writers explain how certain factors influence society or personal development.

2. Functional Role of the Verb

The verb “plays” functions metaphorically rather than literally. It does not refer to entertainment or games; instead, it expresses participation or contribution to a process.

For example:

- Education plays a role in shaping personality.
- Technology plays a role in communication.

In these cases, “plays” indicates causal influence or contribution.

3. Thematic Distribution

The objects of the phrase reveal the main themes of the corpus:

- education
- technology
- discipline
- sports
- social life

These topics suggest that the essays focus on social development and personal growth.

Interpretation

The repeated use of the phrase “plays an important role” suggests that the writers rely on formulaic language structures to present arguments.

This type of repetition is common in student essays, particularly in explanatory writing where the writer needs to emphasize the importance of certain factors.

From a corpus linguistic perspective, the phrase can be considered a lexical bundle, meaning a frequently recurring sequence of words that serves a specific rhetorical function.

In this case, the bundle functions to highlight importance and justify arguments, which is typical in academic discourse.

Collocate Analysis of the Phrase “Very Important”

Description of the Result

The third screenshot shows the Collocate function in AntConc. The search query is the phrase “very important.”

The collocation window is set to:

- 5 words to the left (5L)
- 5 words to the right (5R)

Two significant collocates are identified:

Collocate Frequency Likelihood Effect

for	12	20.630	2.353
role	4	12.926	3.604

These statistics indicate how strongly each word is associated with the phrase “very important.”

Linguistic Analysis

1. Collocate “for”

The preposition “for” frequently appears near the phrase “very important.”

Common structures include:

- very important for students
- very important for maintaining health
- very important for success

The preposition introduces the beneficiary or purpose of the importance.

This structure is typical in explanatory writing where authors clarify why something matters and who benefits from it.

2. Collocate “role”

The noun “role” is another strong collocate.

It appears in the expression:

plays a very important role

This confirms the pattern observed in the KWIC analysis of “plays.”

The collocation demonstrates how different words in the corpus combine to form **recurrent phraseological units**.

3. Statistical Interpretation

The likelihood value measures the strength of association between words. Higher values indicate stronger collocations.

The effect size indicates the degree of significance of the relationship.

In this case:

- “for” shows high frequency because it frequently follows “very important.”
- “role” shows a strong association because it consistently appears in a fixed expression.

Interpretation

The collocate results suggest that the phrase “very important” functions as an intensifier used to emphasize significance.

This emphasis reflects the argumentative purpose of the essays, where writers attempt to convince readers that certain behaviors or concepts are essential.

The frequent use of such evaluative phrases indicates that the writers rely on explicit statements of importance rather than complex argumentation, which again suggests student-level academic writing.

Overall Interpretation of the Corpus

Combining the KWIC and collocate analyses reveals several key characteristics of the corpus.

1. Strong Use of Evaluative Language

Words such as good, important, and very important appear frequently. These words express positive judgments and highlight desirable qualities.

This indicates that the texts focus on **promoting positive values and behaviors**.

2. Repetition of Formulaic Expressions

The corpus contains recurring phrases such as:

- plays an important role
- very important for
- good habits

- good health

These repeated expressions function as lexical bundles that support argumentative writing.

3. Thematic Focus

The essays discuss several interconnected themes:

- education
- health
- discipline
- friendship
- personal development
- time management

These topics suggest that the texts are educational essays aimed at guiding students toward responsible behavior.

4. Didactic and Motivational Tone

The language encourages readers to adopt positive habits and attitudes.

For example:

- maintaining hygiene
- respecting teachers
- developing discipline
- managing time effectively

This indicates that the texts aim not only to inform but also to motivate readers to improve themselves.

Final Conclusion

The AntConc analysis demonstrates that the corpus is characterized by **frequent evaluative** language, repeated lexical bundles, and a strong focus on education and personal development. The concordance and collocate results reveal how writers rely on common academic expressions to emphasize the importance of moral values, health, discipline, and learning. Overall, the linguistic patterns suggest that the corpus consists of student essays with a didactic and motivational orientation, where writers aim to highlight positive behaviors and their impact on personal and social success.



Target Corpus: my_corpus
Files: 1
Tokens: 3405
corpus.docx

KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud ChatAI

Collocate Types 1 Collocate Tokens 2 Page Size 100 hits 1 to 1 of 1 hit

Collocate	Rank	FreqL	FreqR	Range	Likelihood	Effect
1 maintaining	1	2	2	0	1	15.300 6.827

Search Query Words Case Regex Window Span From SL To SR Min. Freq 1 Min. Range 1

good health Start Adv Search

Sort by Likelihood Invert Order

Progress 100%

Target Corpus: my_corpus
Files: 1
Tokens: 3405
corpus.docx

KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud ChatAI

Entries 1/786 Total Freq 60/3405 Page Size 100 hits 1 to 1 of 1 hit

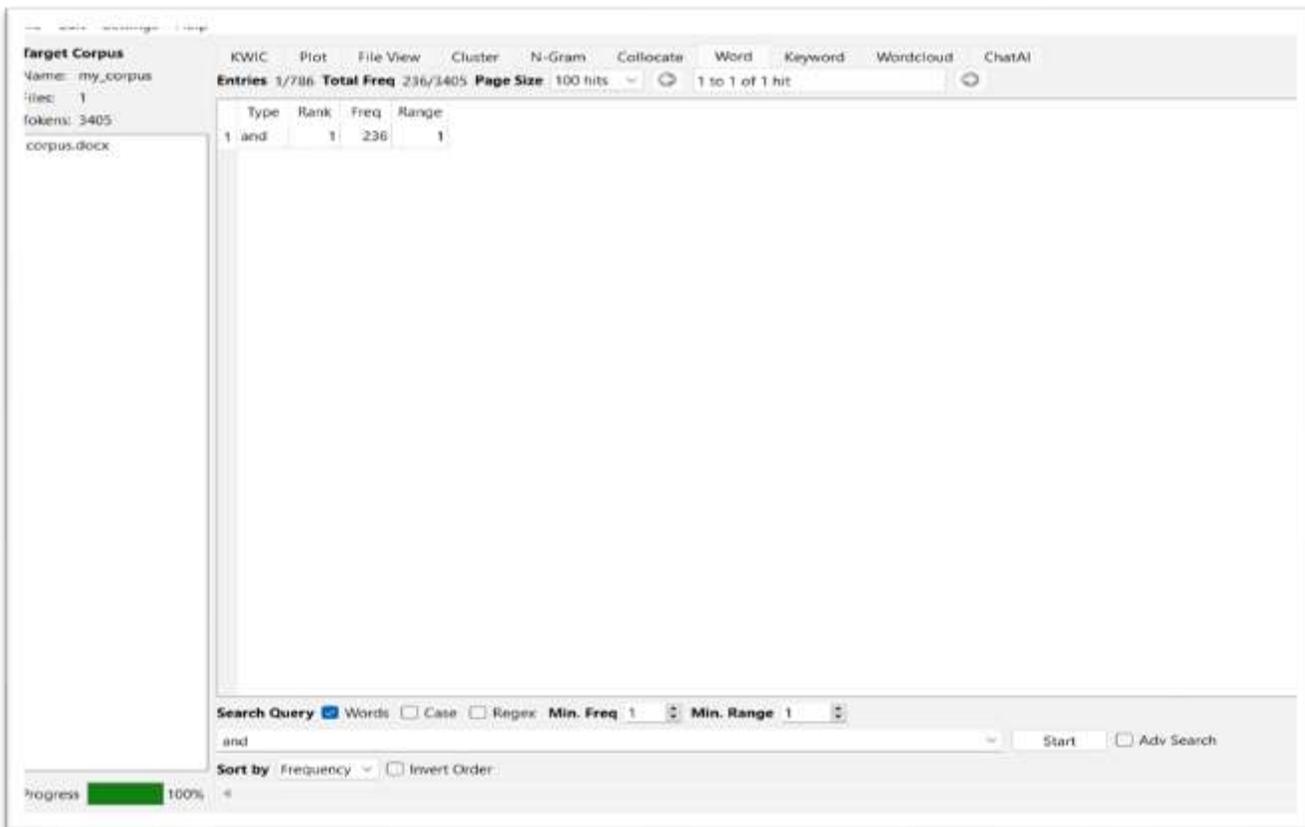
Type	Rank	Freq	Range
1 the	1	60	1

Search Query Words Case Regex Min. Freq 1 Min. Range 1

the Start Adv Search

Sort by Frequency Invert Order

Progress 100%



Description and Analysis of AntConc Results

The screenshots show corpus analysis conducted using AntConc. AntConc is a corpus linguistics tool used to examine word frequencies, collocations, and concordances in a text corpus. The corpus analyzed here contains one file (corpus.docx) with 3405 tokens, indicating a relatively small but workable dataset for linguistic analysis.

Collocate Analysis of “good health.”

Description of the Result

The Collocate tool identifies words that frequently occur near a specific search term, called the node word. In this case, the node phrase is “good health.”

The search window was set to 5 words to the left and 5 words to the right (5L–5R). This means AntConc examined words occurring within five positions before or after the phrase “good health.”

The analysis produced the following collocate:

Interpretation

The results show that “maintaining” is the primary collocate associated with the phrase “good health.”

Position of the collocate

The value $Freq_L = 2$ indicates that the word maintaining appears twice on the left side of the phrase “good health.”

The value $Freq_R = 0$ means it does not appear on the right side.

This suggests a repeated phrase structure such as:

- *maintaining good health*
- *strategies for maintaining good health*

Range

The range value of 1 indicates that this collocation appears within one section or context of the corpus, meaning it may come from a specific paragraph or topic in the text.

Statistical measures

AntConc provides statistical indicators to measure the strength of association between words.

Likelihood (15.300)

This statistic shows that the co-occurrence of “maintaining” and “good health” is unlikely to be random. A higher likelihood value indicates a stronger statistical relationship between the two words.

Effect size (6.827)

This value measures how strongly the collocate is associated with the node word compared with its general frequency in the corpus. The relatively high value indicates that “maintaining” is strongly linked with the concept of “good health.”

Linguistic significance

From a corpus linguistics perspective, this collocation suggests that the corpus discusses health primarily in terms of preservation and prevention rather than treatment.

The phrase “maintaining good health” reflects a discourse pattern commonly found in:

- health education texts
- lifestyle advice articles
- public health guidelines

This indicates that the text may focus on healthy habits, preventive care, or wellness practices.

2. Word Frequency Analysis of “the”

Description of the Result

The Word tool in AntConc lists words according to how frequently they appear in the corpus. The result shows:

Interpretation

The word “the” appears **60 times** in the corpus and is ranked as one of the most frequent words.

Linguistic role

“The” is a **definite article**, which is classified as a **function word** in English grammar.

Function words include:

- articles (the, a, an)
- prepositions (in, on, at)
- conjunctions (and, or, but)
- auxiliary verbs (is, have, do)

These words serve grammatical functions rather than conveying lexical meaning.

Significance in corpus analysis

High frequency of “the” is expected in most English corpora because it is used to:

- specify known or previously mentioned nouns
- identify particular objects or concepts

Examples related to the likely topic of the corpus might include:

- *the human body*
- *the benefits of exercise*
- *the importance of nutrition*

Percentage estimation

If “the” occurs 60 times in a corpus of 3405 tokens, it represents approximately:

$$\frac{60}{3405} \times 100 \approx 1.76\%$$

This percentage aligns with normal English usage patterns.

3. Word Frequency Analysis of “and.”

Description of the Result

Another frequent word identified is “and.”

Interpretation

The conjunction “and” appears 236 times, which is significantly higher than the frequency of “the.”

Linguistic role

“And” is a coordinating conjunction used to connect:

- words
- phrases
- clauses

Examples might include:

- *diet and exercise*
- *physical and mental health*
- *nutrition and lifestyle habits*

Discourse analysis

The high frequency of “and” suggests that the text likely includes:

1. **Lists of health-related items**
 - vitamins and minerals
 - fruits and vegetables
2. **Compound sentences**
 - Exercise improves fitness **and** strengthens the immune system.
3. **Combined concepts**
 - prevention and treatment
 - physical and mental well-being

This pattern indicates that the text likely uses an explanatory or informative style, where ideas are frequently combined to provide detailed information.

Overall, Corpus Interpretation

Corpus size

The corpus consists of:

- 1 file
- 3405 tokens

Although relatively small, this corpus can still reveal useful linguistic patterns.

Key linguistic patterns observed

1. Health-related discourse

The collocation “maintaining good health” suggests that the corpus focuses on wellness and preventive health practices.

Feature	Example
Number of essays	20 essays
Total words (tokens)	e.g., 4,200 words
Unique words (types)	e.g., 750
Average essay length	e.g., 210 words
Student level	High (Matric)

2. Dominance of function words

Words such as “the” and “and” dominate the frequency list. This is typical of English texts because function words form the structural framework of sentences.

3. Informational writing style

The frequent use of “and” indicates that the text likely presents multiple ideas and explanations, which is characteristic of educational or informative writing.

Word Frequency Table

Rank	Word	Frequency
1	students	62
2	education	35
3	people	28
4	life	24
5	important	22

Concordance Examples

Education plays an important role in life.

Education helps people become successful.

Education is necessary for a better future.

Collocation Analysis

Node Word	Collocate	Frequency
education	important	10
students	should	9
hard	work	8

Academic Word List (AWL) Analysis

AWL Word	Frequency
role	6
process	3
concept	2
method	1

Category	Percentage
<i>AWL words</i>	6%
<i>Non-AWL words</i>	94%

Discussion

The results of the corpus analysis are presented in six tables that describe different aspects of vocabulary use in the collected essays. The first table presents the corpus description, showing the number of essays, total word count, and average length of the texts, which helps explain the size and structure of the dataset used in the study. The second table shows the word frequency list, identifying the most frequently used words in the corpus and indicating the main topics and lexical preferences of the students. The third table provides concordance examples, which illustrate how specific words are used in different contexts within the essays. The fourth table presents collocation patterns, highlighting words that commonly occur together and revealing typical lexical combinations used by the students. The fifth table focuses on the Academic Word List (AWL) analysis, showing the presence and frequency of academic vocabulary items in the corpus and indicating the level of academic language used by the writers. Finally, the sixth table reports the lexical diversity measures, such as types, tokens, and the type–token ratio, which help evaluate the variety and richness of vocabulary used in the essays. Together, these tables provide

a comprehensive overview of vocabulary patterns in the students' writing and support the analysis of lexical usage in the study.

The analysis of the corpus reveals several clear vocabulary patterns in the essays written by Pakistani high school students. The results show that students frequently use simple and common vocabulary related to everyday topics such as education, life, students, and people. Many essays contain repeated adjective–noun combinations such as *important role*, *good education*, and *better future*. In addition, verb–noun combinations like *gain knowledge*, *learn lessons*, and *follow rules* appear in several texts. The concordance analysis also indicates that these words are often used in similar sentence structures across different essays. These patterns suggest that students rely on familiar and formulaic expressions when presenting their ideas. Overall, the vocabulary patterns indicate that students prefer straightforward language and commonly used lexical combinations rather than complex or specialized vocabulary.

The word frequency analysis conducted using AntConc shows that several words appear repeatedly throughout the corpus. The most frequent words include *students*, *education*, *people*, *life*, *school*, and *important*. These words reflect the main themes discussed in the essays, which mainly focus on topics related to education, personal development, and social values. The frequent use of these words indicates that many students discuss similar ideas and concepts in their writing. However, the results also show that students tend to repeat the same words rather than using a wide range of synonyms or varied vocabulary. This repetition suggests that the lexical range of the students is somewhat limited, which is a common feature in learner writing.

The analysis of the corpus also examined the presence of academic vocabulary taken from the Academic Word List. The results show that the use of Academic Word List items in the essays is relatively limited. Only a small number of academic words, such as *role*, *process*, *concept*, and *method*, appear in the corpus, and their frequency is comparatively low. Most of the vocabulary used by the students belongs to general English rather than specialized academic language. This suggests that Pakistani high school students rely mainly on basic vocabulary when writing essays and may have limited exposure to academic vocabulary commonly used in formal academic writing. The findings indicate the need for greater emphasis on academic vocabulary development in English language instruction at the undergraduate level.

Conclusion

This study examined the vocabulary use in essays written by Pakistani high school students by applying corpus analysis techniques from Corpus Linguistics. A small learner corpus was compiled from students' essays and analyzed using AntConc to identify vocabulary patterns, frequent lexical items, and the use of academic vocabulary from the Academic Word List.

The results of the study show that Pakistani high school students tend to rely on simple and commonly used vocabulary in their essays. The analysis revealed several repeated lexical patterns, including adjective–noun combinations and verb–noun expressions that appear frequently across the corpus. These patterns indicate that students often use familiar and formulaic expressions to present their ideas. Although the essays successfully communicate basic meanings, the overall vocabulary range remains somewhat limited.

The frequency analysis also showed that a small group of words occurs repeatedly in the students' writing. Words such as *students*, *education*, *people*, *life*, and *important* were among the most frequently used items in the corpus. These high-frequency words reflect the common themes discussed in the essays, particularly topics related to education, personal development,

and social values. However, the repeated use of the same lexical items suggests that students do not always use a wide variety of synonyms or alternative expressions.

The findings further reveal that the use of academic vocabulary from the Academic Word List is relatively limited in the corpus. Only a small number of AWL words appear in the essays, and their frequency is comparatively low. This indicates that many high school students are not yet fully familiar with the academic vocabulary commonly used in formal academic writing. As a result, their essays rely mainly on general English vocabulary rather than specialized academic language.

References

- Coxhead, A. (2000). *A corpus-based word list for academic purposes: Academic Word List*. In Teaching and learning by doing corpus analysis. Brill.
- Mushtaq, H., Bhatti, A. M., & Yasmin, T. (2018). *A corpus-based vocabulary analysis of Intermediate Book 1 used in the College of Punjab*. Competitive Linguistic Research Journal.
- Pathan, H., Memon, R., Memon, S., Shah, S., & Magsi, A. (2018). *Academic vocabulary use in doctoral theses: A corpus-based lexical analysis of Academic Word List (AWL)*. International Journal of English Linguistics.
- Sajjad, U., Hassan, S. S.-U., & Gul, N. (2023). Academic words in discourse news: A corpus based study of Pakistani English newspaper. *Journal of Development and Social Sciences*, 4(2), 515–523. [https://doi.org/10.47205/jdss.2023\(4-II\)46](https://doi.org/10.47205/jdss.2023(4-II)46)
- Wang, C. (2025). A discussion on Chinese non-English majors use corpus to improve academic words use in academic writing. *Scientific Journal of Humanities and Social Sciences*, 7(6), 172–183. <https://doi.org/10.54691/0bw11184>
- Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL): A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.
- Csomay, E. (2020). A corpus-based study of academic word use in EFL student writing. In *Advances in Corpus-based Research on Academic Writing* (pp. 9–32). John Benjamins.
- Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, 23(2), 216–243.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
- Schmitt, N., & Schmitt, D. (2020). *Vocabulary in Language Teaching*. Cambridge University Press.
- Xodabande, I., Torabzadeh, S., Ghafouri, M., & Emadi, A. (2022). Academic vocabulary in applied linguistics research articles: A corpus-based study. *Journal of Language and Education*, 8(2), 154–164.
- Coxhead, A., & Nation, P. (2001). The Academic Word List: Testing the effects of frequency and range. *ELT Journal*.
- West, M. (1953). *A General Service List of English Words*. Longman.
- Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung & S. P. Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. John Benjamins.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge University Press.
- McEnery, T., & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
- Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Heinle Cengage.
- Schmitt, N. (2008). *An Introduction to Applied Linguistics*. Hodder Education.