
Developing EFL Vocabulary to improve Academic writing through a frequency list at the Secondary level in Sahiwal, A Pakistani English Perspective

Ayesha Wakeel

Research Scholar, Department of English, University of Sahiwal, Sahiwal, Pakistan

ORCID:

<https://orcid.org/0009-0000-2537-3514>
ayeshawakeelwakeel@gmail.com

Abstract

Vocabulary maintenance in long-term memory among EFL learners is a sustained challenge, often relying on rote memorization and random lists of English terms that do not improve academic writing. This study is based on a tagged corpus of Dawn opinion articles for vocabulary development and writing improvement at the secondary level. The main objective of this study is to provide participants with contextual information about lexical items through a frequency list of common nouns. A mixed-method approach was used, and opinion articles comprising 1,481,863 tokens, published by various authors from 2013 to 2026, were selected as the data sources. Twenty beginner-level EFL learners from Sahiwal participated in a pre-test and post-test with the help of a questionnaire and data were analyzed in SPSS software. After conducting tests with an experimental group using corpus-based vocabulary learning (new method) and a control group using Dawn vocabulary learning (traditional method), the control group also completed another post-test after learning the vocabulary from the corpus-based frequency list of nouns. The findings indicated significant improvement from pre-test to post-test, increased average scores, and high learner vocabulary satisfaction with the help of the contextualized noun list. Moreover, the frequency list effectively strengthened vocabulary retention through the use of AntConc and TagAnt and provided engaging learning experiences in the EFL context. Future research should involve more diverse and larger participant groups, extend the duration of teaching, and include other grammatical aspects to enable learners to write essays or paragraphs and develop proficiency in English writing.

Keywords: Corpus-based learning, Vocabulary Development, Academic Writing, EFL learners, Long-Term Retention, Secondary-level learners

Introduction:

Based on the Common European Framework of Reference (CEFR), 88% of secondary school learners in Pakistan are the basic users (A1–A2) of English, while only 12% reach the (B1) proficiency level, indicating an overall low level of English proficiency. In Pakistan, people mainly use their native languages; however, English language is necessary in education, business, social interaction, and International meetings. In EFL education the four language skills are the areas in which the learners are engaged: listening, speaking, reading and writing. Today, successful communication is essential in every profession and in academic writing because it provides clarity and confidence and according to Vygotsky's sociocultural theory which states that a second language develops through social interaction. Therefore, Vygotsky describes writing as a social process instead of cognitive act. (A teacher guides a pupil through the writing process in educational institutions within (ZPD) the Zone of Proximal Development). As a result, strong writing skills promote effective communication in educational and professional institutions. Therefore, learners need to develop strong writing skills to communicate effectively in written form within their academic disciplines (Birhan, 2018). Recognizing the importance of writing, many recent studies have focused on investigating this skill (Crosthwaite, Storch, & Derakhshan, 2020; Lee, 2020; Yu, Xu, Jiang, & Chan). Although learners face many challenges in developing foreign-language writing skills, newspapers provide an effective resource for improving these skills. Historically, Europe introduced the printing press in the fifteenth century, and Germany published the first official weekly newspapers in 1605. The industrial revolution and the penny press spread newspapers to other countries. Therefore, in 17th-century England, men discussed local and European news in coffeehouses. By the early nineteenth century, newspapers expanded to the American colonies, Asia (Hicky's Bengal Gazette, 1780), and the Middle East. In 1941, Muhammad Ali Jinnah, the founder of Pakistan, founded the first English newspaper Dawn in Pakistan, a country in South Asia. The difficulty that students encounter in language development is one of the formal academic writing abilities, since they don't have the ability to use contextually appropriate vocabulary for specific topics. Due to the formal, serious, and accurate word choices of the Dawn newspaper, it is very popular among English learners, who use it with AntConc to improve their English vocabulary. Research conducted by Siddiq et al. (2021) and Sinturat et al. (2022) suggests that corpus-based tools make significant contributions in vocabulary acquisition in terms of understanding word meaning in various contexts and retaining vocabulary in long term memory. By traditional method of learning lexis from Dawn articles, students face difficulty in mastering a native-like English vocabulary. Therefore, learners need a modern method to learn English by developing a corpus, which means a collection of digitally stored texts. For this purpose, this study selects digital texts from Dawn opinion columns published on the Dawn website from 2013 to 2026. These data show how writers use words in different contexts across multiple Dawn articles and provide a better understanding of vocabulary development because the learners study a frequency list of nouns rather than less frequent

words. Moreover, this method reduces learner's time and costs compared with purchasing newspapers and reading the articles manually. Furthermore, this approach helps non-native speakers become proficient in their target language. Despite its importance, the Dawn newspaper deserves detailed academic investigation.

Statement of the Problem:

Newspapers not only highlight societal and political issues but also provide vocabulary knowledge for academic writing. At the secondary level in Sahiwal, students face difficulties because they are unable to comprehend lexical items, and this limited vocabulary hinders the development of English language proficiency. The primary concern of this research is to determine how learners develop vocabulary from the Dawn newspaper through a frequency list of nouns using "AntConc and TagAnt." These innovative approaches enhance learners' writing skills and enrich their vocabulary by introducing the words that Dawn writers use in their articles.

Questions:

1. How can a corpus provide better opportunities for English language learning than the traditional method of learning vocabulary from newspapers?
2. How does the noun frequency list of the Dawn newspaper help Pakistani EFL learners develop native-speaker-like English?
3. What is the impact of tagged contextualized vocabulary on improving the vocabulary of secondary-level learners?

Objectives:

1. To identify how different writers contextualize words in the Dawn newspaper's opinion section.
2. To examine the use and effectiveness of high-frequency lexical items in developing EFL learners' vocabulary.
3. To investigate the improvement in academic writing among secondary-level learners through the implementation of a high-frequency word list.

Literature Review:

Many researchers have argued that vocabulary knowledge is significant for assessing the quality of second language writing (Csomay & Prades, 2018), reading (Qian, 1999), listening (Stæhr, 2009; Zhang & Graham, 2020), and speaking (Alharthi, 2020; Lailiyah & Setiyaningsih, 2020). To address this challenge, several researchers' publishers published corpus-based studies for EFL learners because the traditional method of teaching vocabulary through words meanings and translation from texts posed a major challenge for non-native learners. Firstly, Fauzi and Surandi (2018) proposed a Computer-Based Corpus Approach to teach tourism vocabulary to fourth semester Diploma 3 students instead of the traditional

method of learning vocabulary from textbooks. Secondly, there are four languages that have academic vocabulary lists developed by Ahsanddin, Hanafi, Basthomi, Taufiqurrahman, A. Bukhori, Samodra, Widiati, and H. Wijayati (2022). They identified the five sub-corpora in this research, which are Indonesian, English, German, Arabic and art and design. The researchers primarily aimed to develop a field-related and high-frequency vocabulary. Similar to the previous study, this research also showed positive results on learners' vocabulary development. Thirdly, UÇAR and YÜKSELİR (2015) analyzed the corpus-based activities which affected verb-noun collocations in EFL classes. The researchers took the main course book Speak Out of Intermediate and COCA for the experimental group of the preparatory class students of school of foreign languages, and used the same textbook without a corpus as the control group. This study also states that learners' vocabulary improved. The above mentioned studies involved a limited number of participants and they collected data from text books that provide fixed expressions. By using these fixed expressions, learners cannot learn vocabulary in real-life contexts which hinders their progress in writing. To resolve this problem, researchers apply the corpus approach to newspapers, where word variation emerges daily.

To begin with, Bibi, Fatima and Qureshi (2024) investigated the impact of a frequency list of vocabulary items on EFL learners. They found, from using higher secondary textbooks of Punjab as a source of Pakistani English, that the experimental group's vocabulary was significantly higher than the control group (traditional group) that learned the English language in the traditional way. Subsequently, Akhtar, Anwar and Qureshi (2019) compiled a Pakistani English corpus from multiple sources at the secondary level. The researchers compiled the dataset from Dawn, Daily Times, The News, and business recordings. Over a one-month period in 2018, the researchers assembled these data from different domains, including legal, educational, administrative, judicial and religious fields. The findings indicated that learners achieved better language proficiency through the corpus-based approach. Next, Qureshi and Haidov (2020) collected Pakistani and English newspapers and developed various corpus-based collocations that helped learners use language in real-life settings. Such language learning helped EFL learners improve their creative writing. Finally, Sabiri, Iqbal, Batool, Abubaker, and Zafar (2022) studied a corpus of prepositional phrases and collected data from Dawn newspaper editorials. Notably, AntConc identified combinations of prepositional phrases that formed complete sentences both semantically and syntactically. Overall, all these studies illustrate the importance of vocabulary; however, these studies focused on developing academic writing, there is still a need for research that helps learners develop both vocabulary and writing skills. This research focuses on opinion sections in which Pakistani writers express their own views on educational, economical, environmental, technological, and political issues. In the Dawn newspaper, opinion articles cover a wider range of topics than editorials, and many students also learn from this newspaper because its writers use formal academic language that facilitates learners preparation for competitive examinations. One of the major written

competitive examinations is the Central Superior Services (CSS) carried out by the Federal Public Service Commission (FPSC). Through this exam, successful candidates become Grade-17 officers in Pakistan and serve the public. Because many researchers and teachers highly recommend the Dawn newspaper for writing, corpus-based research using a noun-frequency list is more beneficial as it assists participants to learn high-frequency lexical items that writers frequently use in daily life and in formal writing rather than low-frequency words which are difficult to memorize. Thus, this study aims to facilitate secondary-level learners in Pakistan because students have limited familiarity with lexical items during primary education. At the secondary level students can learn and use vocabulary more effectively than primary-level students. Primary-level learners usually focus on basic tenses and spellings, so they cannot use lexical bundles in context. In contrast, secondary-level learners develop their vocabulary from informal to formal diction.

Theoretical Framework:

The theoretical framework is based on the receptive and productive categories that Paul Nation (2001) proposes. These two categories are interrelated and support language proficiency. Through these two categories, learners can comprehend vocabulary (receptive skills) and express themselves in writing. Nation also proposes four strands of language learning: meaning-focused input, which includes reading and listening; meaning-focused output, in which learners produce language from what they have learned through writing and speaking; language-focused learning, which includes learners' concentration on grammar, vocabulary, spelling, and pronunciation; and fluency development, in which learners practice the language fluently. He also proposes that learners can achieve 80 percent understanding of a language after they acquire 2,000 high-frequency words. This theory is relevant to the present study because vocabulary comprehension (receptive) through newspapers and the application of vocabulary (productive use) in writing activities are closely connected. In addition, all four strands of language learning provide an appropriate foundation for vocabulary development.

Methodology

Research Design:

This research utilized a mixed-method design to analyze the linguistic choices in the Dawn newspaper because it combined the strengths of both qualitative and quantitative research. The research used pre-tests and post-tests to collect the quantitative data for measuring learners' vocabulary improvement whereas the qualitative data provided a deeper interpretation of newspaper language in its context. The research design consisted of two groups: a control and an experimental group. This research conducted the pre-test to assess the existing English language knowledge of both groups. Correspondingly, this research conducted the post-test for both groups to examine the difference in learning lexical bundles. By including a sample of 20 students, the control group obtained instructions through the conventional method of learning vocabulary of Dawn opinion-section whereas the

experimental group obtained instructions through the innovative method of corpus-based vocabulary approach at Ali City High School, Sahiwal District. Baker (2023) argued that the corpus-based approach allowed researchers to identify textual patterns and treat language as self-contained objects. To develop the corpus, this research selected multiple opinion articles from Dawn archive and tagged them with TagAnt. TagAnt provided parts-of-speech tags for each tokens in the articles. The study used a list of 500 nouns from the corpus through the key word in context (KWIC) feature to improve the noun vocabulary of the experimental group.

Ethical Consideration:

After tests, it was unethical to leave the control group without providing technological learning. Therefore, this research taught the same list of nouns to the control group used with the experimental group, as presented below in figure 1:

File	Left Context	Hit	Right Context
Babar Sattar.txt	NNP -HYPH_NNP called_VBN_NNP political_JJ_NN arm_	NN	NNP of_IN_NNP the_DT_NNP Lashkar_NNP_
Hajrah Mumtaz.txt	to_TO_NNP slow_VB_NNP the_DT_NNP pace_	NN	NNP of_IN_NNP violence_NN_NN at_IN_
Huma Yusuf.txt	leader_NN_NNP at_IN_NNP the_DT_NNP helm_	NN	NNP of_IN_NNP the_DT_NNP Afghan_NNP_
Arif Azad.txt	NNP_NNP Kurd_NNP_NNP , , _NNP , , _ the_DT_NNP head_	NN	NNP of_IN_NNP LJ_NNP_NNP ' ' s_POS_
Jawed Naqvi.txt	time_NN_NNP following_VBG_NNP the_DT_NNP signing_	NN	NNP of_IN_NNP the_DT_NNP nuclear_JJ_
Mahir Ali.txt	being_VBG_NNP the_DT_NNP staging_NN_NNP ground_	NN	NNP of_IN_NNP proxy_JJ_NNP sectarian_JJ_
Rafia Zakaria.txt	to_TO_NNP counter_VB_NNP a_DT_NNP sense_	NN	NNP of_IN_NNP empowerment_NN_NNP among_IN_
Zahid Hussain.txt	found_VBD_NNP that_IN_NN only_JJ_NN 50pc_	NN	NNP of_IN_NNP Pakistanis_NNP_NNPS accept_VBP_
Faisal Siddiqui.txt	poll_NN_NN revealed_VBD_NN that_IN_NN 66pc_	NN	NNP of_IN_NNP Pakistanis_NNP_NNP believe_VBP_
Hammad Siddiqui.txt	point_VBP_NNP to_IN_NNP the_DT_NNP resilience_	NN	NNP of_IN_NNP sectarianism_NN_NN _ _NNP This_
I.A. Rehman.txt	and_CC_FW enjoying_VBG_CC the_DT_NNP reputation_	NN	NNP of_IN_NNP being_VBG_NNP good_JJ_
Mozam Hussain.txt	Education_NNP_SYM Commission_NNP_NN ' ' s_POS_VBD criterion_	NN	NNP of_IN_NNP the_DT_NNP research_NN_
Asim Sajjad Akhtar.txt	based_VBN_NNP on_IN_NNP the_DT_NNP number_	NN	NNP of_IN_NNP times_NNS_NNP the_DT_
Babar Sattar.txt	NNP _ _NNP) -RRB- _ _NFP -RRB- _ADD The_DT_NNP ineligibility_	NN	NNP of_IN_NNP these_DT_VBN men_NNS_
Shaikat Hameed Khan.txt	soon_RB_NNP given_VBN_NNP the_DT_NNP number_	NN	NNP of_IN_NNP researchers_NNS_NNP that_WDT_
Zarrar Khuhro.txt	and_CC_FW the_DT_NNP low_JJ_NN interest_	NN	NNP of_IN_NNP the_DT_NNP world_NN_
Zohra Yusuf.txt	West_NNP_NNP at_IN_RB the_DT_NNP expense_	NN	NNP of_IN_NNP _ _NNP _ _ for_IN_NNP example_
Dr Niaz Murtaza.txt			
Jawed Naqvi.txt			
Muhammad Ali Siddiqui.txt			
Tariq Khosa.txt			
Mahir Ali.txt			
Rafia Zakaria.txt			

Figure 1: The list of contextual nouns

The above data is extracted from AntConc after being part-of-speech tagged by TagAnt. It explained the context of nouns in KWIC feature with the name of authors (in the file column as a source) that helped in finding the particular article of the author from which the noun selected among participants. The display presented the selected nouns, as well as the linguistic context on the left and right in the articles from Dawn opinion. The hit column was followed by the NN words in the form of 'tagged nouns' that were used to indicate the position. Therefore, the contextual sentences facilitated learners to explore the usage of nouns in their authentic sentences, and helped them to gain a clearer understanding of the meaning and usage of the grammatical position of nouns in academic writing.

Data Collection:

From 2013 to 2026, this research collected multiple articles written by different authors and pasted in notepad to save them in text format. To avoid bias, the study placed each text in a separate folder according to the author's name and publication year. Furthermore, two manual checks verified all files to avoid duplication and remove unnecessary information. The corpus incorporated, in figure 2, all the collected data and comprised 1481863 tokens, which were sufficient for learning corpus-based vocabulary at secondary level.

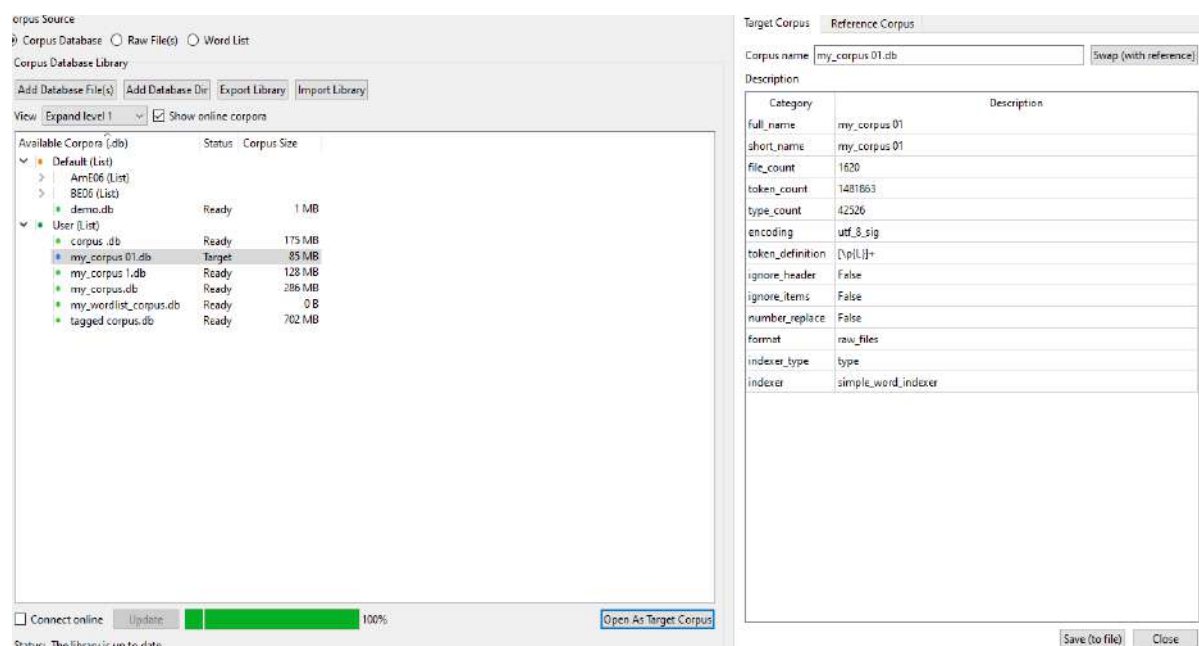


Figure 2: Total tokens of this research corpus

Although small-scale corpora for pilot studies often ranged from 2,000 to 5,000 words (Evans, 2007), studies on lexical bundles and corpus linguistics generally used corpora containing at least 100,000 tokens to conduct meaningful frequency analysis (Biber et al., 1999; McEnery & Hardie, 2011). Firstly, the data as given below in figure 3 used that was barrier for learners to comprehend vocabulary because the AntConc data collected without pos-tagging from TagAnt.

File Edit Settings Help

Target Corpus
 Name: my_corpus 01
 Files: 1620
 Tokens: 1481863

File	Left Context	Hit	Right Context
1 Faisal Siddiqi.txt	policy to the military establishment. The 'war against terror', the '	fight	against the Baloch insurgency' and their role in fighting
2 Jawed Naqvi.txt	decimation of the Maoists in Chhattisgarh even ahead of their	fight	against the BJP's communal fascism. Maoism is an
3 Jawed Naqvi.txt	latter's party agrees to join the Congress in the	fight	against the challenge of fascism. It would all add
4 Najmuddin A. ...	as the Afghan corps in Helmand, are handicapped in the	fight	against the insurgents because the police are non-functional
5 Zarrar Khuhr.txt	and their land". These are the ones who will even	fight	against the Israeli army when it comes (once in
6 Abbas Nasir.txt	body, the National Counter-Terrorism Authority (Nacta), would reinvigorate the	fight	against the lawless hordes holding society hostage as it
7 Ahmer Bilal ...	Thus far, government representatives have repeatedly affirmed the decision to	fight	against the proliferation of terrorism and the threat of
8 Sehar Tariq.txt	opposition of public policy schemes. While we should continue to	fight	against the sexism embedded in our society, we should
9 Abbas Nasir.txt	of these soldiers are in the front line of the	fight	against the Tehreek-i-Taliban Pakistan and some of
10 Afzal Ali Shigri.txt	continues to prod the reluctant governments into some action. The	fight	against terrorism seems to be low on the government'
11 Afzal Ali Shigri.txt	heavy odds, amidst relentless casualties and lack of resources. The	fight	against terrorism seems to be low on the government'
12 Latha Jishnu.txt	grievously from terrorist acts, and Modi has made the global	fight	against terrorism a key plank of his foreign policy —
13 Nargis Sethi.txt	on its development agenda and from effectively carrying out the	fight	against terrorism and militancy. The electricity problem has four
14 Arsla Jawaid.txt	susceptible to extremism. There is no scenario where Pakistan's	fight	against terrorism can be won solely through military means.
15 Asma ...	but in a manner that weakened the democratic transition. The	fight	against terrorism in Pakistan has taken a toll on
16 Abbas Nasir.txt	to see a distraction removed from the scene as the	fight	against terrorism needs focus. I know many friends will
17 Irfan Husain.txt	despite thousands of our security personnel being killed in the	fight	against Islamist militancy. But never before has an American

Search Query ☒ Words ☐ Case ☐ Regex Results Set All hits Context Size 10 token(s)
 fight Start ☐ Adv Search

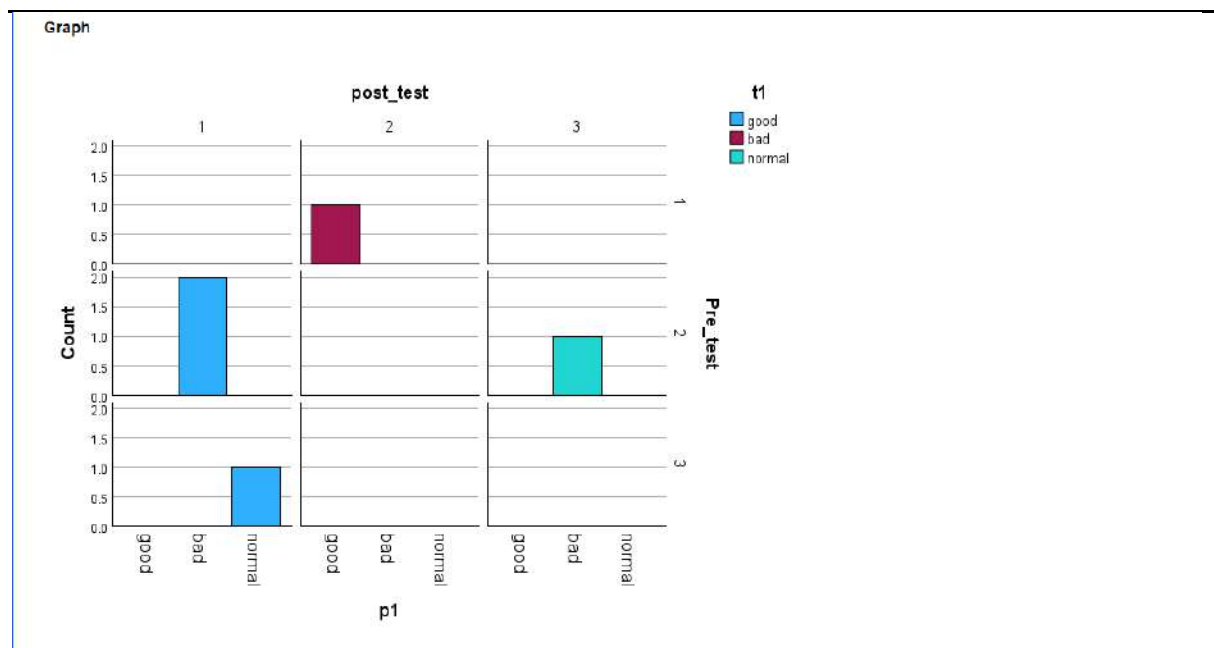
Figure 3: Data without pos-tagging

In the above figure the word "Fight" in KWIC was a great challenge to use in an academic writing. To facilitate participants' understanding of the texts, TagAnt provided tagged data because, without tagging, respondents found it difficult to identify nouns in newspaper articles.

After the vocabulary instruction, the current study examined to understand learners' experiences with the corpus-based vocabulary teaching approach using the newspaper. To verify participants' satisfaction and learning from corpus-based vocabulary approach, Google Forms served as a platform for the online questionnaire, which consisted of 15 questions with 5 response options: strongly agree, agree, neutral, disagree, and strongly disagree.

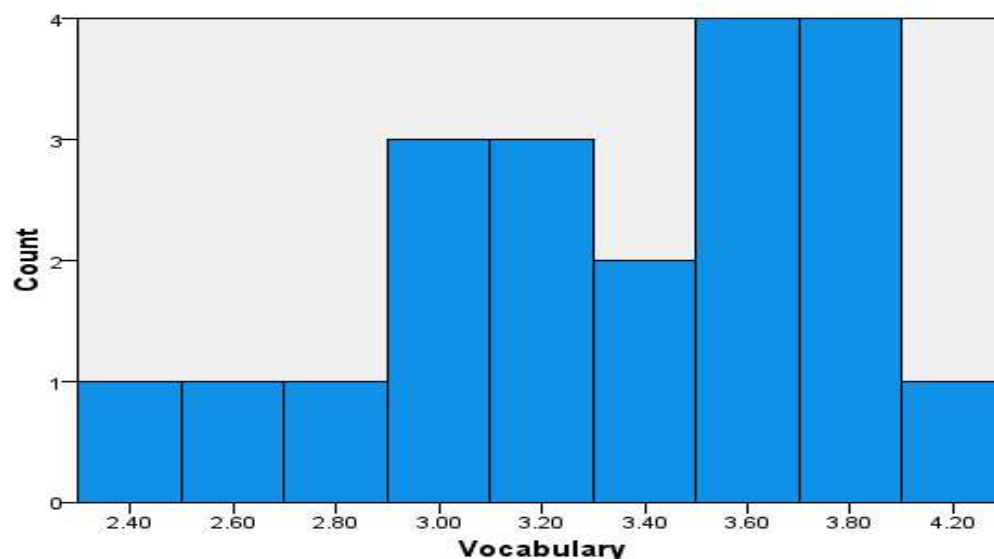
Data Analysis:

SPSS software analyzed the data and demonstrated learners' responses before and after learning through the corpus. The graph 1 indicated that the post-test improved results than the pre-test results. Correspondingly, the good, normal, and bad categories in the graph were based on high, average, and low scores out of 20.



Graph 1: Improvement in a post-test result

After the post-test, the study shared an online questionnaire link in the WhatsApp group to collect students' feedback of corpus-based vocabulary learning and teaching. To check the responses of learners' vocabulary, an histogram graph, graph 2, generated from the SPSS software.



Graph 2: An histogram graph

In the histogram, the X-axis ranged from approximately 2.4 to 4.4 demonstrated the learners' responses to the noun frequency list. Most students responded between 3.5 and 3.8, evidencing that many participants learned greater vocabulary from the tagged corpus. Only a few students scored below 3.0, which illustrated the overall vocabulary improvement among EFL learners. Moreover, using descriptive statistics, this study involved measuring the mean and the standard deviation, mentioned below in table 1, of the corpus-based vocabulary as well as participants' vocabulary satisfaction based on the questionnaire responses of 20 students.

→ **Correlations**

Descriptive Statistics			
	Mean	Std. Deviation	N
Vocabulary_satisfaction	3.4400	.39256	20
Vocabulary	3.3500	.45826	20

Correlations			
		Vocabulary_satisfaction	Vocabulary
Vocabulary_satisfaction	Pearson Correlation	1	.620**
	Sig. (2-tailed)		.004
	N	20	20
Vocabulary	Pearson Correlation	.620**	1
	Sig. (2-tailed)	.004	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

Table 1: Mean and Standard deviations

The mean value of the Vocabulary Satisfaction was 3.44 (SD = 0.39), whereas the mean value of Vocabulary was 3.35 (SD = 0.46). On the five-point Likert scale, both mean values were above 3.00, indicating that students gave from neutral to positive regarding vocabulary learning and vocabulary satisfaction. The standard deviations for vocabulary and vocabulary satisfaction were 0.45826 and 0.39256, respectively, indicating high consistency and low variability of respondents' responses.

Discussion:

The results of this study reveal that the collected corpus refined EFL learning, and the Dawn noun-frequency list is effective for the experimental group in developing a native-speaker-like English. The control group's results from the traditional vocabulary learning method show that the tagged contextualized vocabulary is essential for vocabulary development at the secondary level.

This study has some limitations. First, this study focuses only on a list of nouns while it excludes other parts of speech because participants concentrate more on school examination than extra curriculum activities. Second, this study applies the intervention to a specific sample. Therefore, the results are not applicable to the entire population. Third, the intervention period is too short for EFL learners to make them enable to write an essay.

However, the SPSS analysis shows that the results are stable. This is because the authenticity and validity of the Pakistani English Corpus strengthens the findings.

A major strength of this study is that it collects data from an authentic language source, a Pakistani English newspaper, instead of relying on a separate list of English words to which a corpus-based approach is not applicable. Another valuable finding is that the noun frequency list proves meaningful for academic Writing, as the Vocabulary Satisfaction analysis highlights. Additionally, the descriptive analysis of the cross tabulation, in table 2, shows that all students participated in the questionnaire.

➔ **Crosstabs**

Case Processing Summary						
	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
Vocabulary*	20	100.0%	0	0.0%	20	100.0%
Vocabulary_satisfaction						

Vocabulary * Vocabulary_satisfaction Crosstabulation										
Count		Vocabulary_satisfaction								Total
		2.60	2.80	3.20	3.40	3.60	3.80	4.00	4.20	
Vocabulary	2.40	0	0	1	0	0	0	0	0	1
	2.60	0	1	0	0	0	0	0	0	1
	2.80	0	0	0	1	0	0	0	0	1
	3.00	1	0	1	0	0	1	0	0	3
	3.20	0	0	2	1	0	0	0	0	3
	3.40	0	0	0	1	0	0	1	0	2
	3.60	0	0	1	2	0	1	0	0	4
	3.80	0	0	0	1	2	0	1	0	4
	4.20	0	0	0	0	0	0	0	1	1
Total		1	1	5	6	2	2	2	1	20

Table 2: Cross tabulation

The rows represent vocabulary scores, while the columns represent vocabulary satisfaction scores. Most participants obtained vocabulary scores ranging from moderate to high (3.20 to 3.80), with 4 participants (20%) in each score category, while 3.00 and 3.20 were reported by 3 participants (15%). Similarly, the most frequent vocabulary satisfaction score was 3.40, reported by 6 participants (30%), followed by a score of 3.20, reported by 5 participants (25%). Participants with higher vocabulary scores generally reported higher vocabulary satisfaction, whereas those with low vocabulary scores showed low vocabulary satisfaction. Overall, the positive responses of participants demonstrate that corpus-based vocabulary learning is helpful in developing proficiency in English among non-native speakers in Pakistan.

This study indicates that the common and frequent use of words has a significant impact on EFL learners' English proficiency. And this study is consistent with previous studies including Bibi, Fatima, and Qureshi (2024) that showed the impact of frequency list of vocabulary items on EFL learners. Research by Durrant (2022) highlights that inadequate genre awareness, syntactic competence, lexical knowledge, and the underutilization of modern technology remain key obstacles to effective academic writing instruction. Therefore, previous studies supports corpus-based learning in educational institutions.

The findings suggest that English teachers can use the Dawn newspaper to improve participants' English writing skills, and the curriculum developers can include this corpus-based vocabulary in textbooks. Moreover, future studies should include essay writing tasks based on the noun frequency list and involve larger samples. Additionally, teachers should teach learners during their summer or winter vacations for longer periods to strengthen participants' English language.

Conclusion:

This corpus-based vocabulary analysis examines the Dawn newspaper opinion articles published from 2013 to 2026. The central focus of this present research is to find out the impact of the tagged corpus that was used in the development of the noun frequency list in the the Dawn newspaper for the EFL learners. The study also aims to examine the participants' understanding of contextualized words in the noun frequency list and to determine the contribution of the corpus to academic writing. The findings reveal that the corpus of 1,481,863 tokens promotes the contextual understanding of vocabulary used in authentic contexts rather than simple memorization through the Keyword in Context in (KWIC) feature. The findings suggest that the corpus-informed method is more effective in the improvement of vocabulary than those the traditional method of learning vocabulary from the Dawn newspaper. The analysis of vocabulary satisfaction and post-tests indicate that participants are able to write better than they did before the study. Furthermore, the study emphasizes the importance of learning vocabulary and lexical items in context rather than as isolated words. It emphasizes the need to implement common and high-frequency vocabulary among secondary- level learners in Sahiwal. This study focuses on the tagged corpus-based nouns in vocabulary building. Thus, this conclusion points out the necessity of greater involvement in corpus-based research and highlights the importance of adjectives, phrasal verbs, adverbs, collocations, prepositions, and other grammatical features, particularly for beginner EFL learners developing English proficiency.

References:

Ahsanuddin, M., Hanafi, Y., Basthomi, Y., Taufiqurrahman, F., Bukhori, H. A., Samodra, J., Widiati, U., & Wijayati, P. H. (2022). Building a corpus-based academic vocabulary list of four languages. *Pegem Journal of Education and Instruction*, 12(1), 159–167.
<https://doi.org/10.47750/pegegog.12.01.15>

- Akhter, S., Anwar, B., & Qureshi, A. H. (2019). Breadth of EFL vocabulary building at secondary level in Pakistan: A corpus-based perspective. *Global Regional Review*, 4(4), 105–112. [https://doi.org/10.31703/grr.2019\(IV-IV\).12](https://doi.org/10.31703/grr.2019(IV-IV).12)
- Alene, S., Amogne, D., & Mohammed, S. (2025). Effects of corpus-based instruction on students' academic writing development and their perceptions toward learning with corpora: A focus on self-compiled corpora. *ETHIOINQUIRY Journal of Humanities and Social Sciences*, 4(2), 15–31. <https://doi.org/10.20372/9vv7xg68>
- Amjadiparvar, A., & Shafie, A. (2021). The contribution of collaborative corpus-based practice to writing improvement. *Journal of Language Teaching and Learning*, 11(2), 98–113. <https://doi.org/10.66887/jltl.v11i2.317>
- Anwar, B., & Qureshi, A. H. (2019). British or American English: Emerging trends in Pakistani English newspapers. *Online Journal of Communication and Media Technologies*, 9(4), Article e201918. <https://doi.org/10.29333/ojcm/5863>
- Bibi, A., Fatima, A., & Qureshi, A. H. (2024). Developing EFL academic writing at higher secondary level with frequency word-list: Pakistani English corpus perspective. *Pakistan Social Sciences Review*, 8(2), 830–837. [https://doi.org/10.35484/pssr.2024\(8-II\)66](https://doi.org/10.35484/pssr.2024(8-II)66)
- Birhan, A. T. (2021). Effects of teaching lexical bundles on EFL students' abstract genre academic writing skills improvement: Corpus-based research design. *International Journal of Language Education*, 5(1), 585–597. <https://doi.org/10.26858/ijole.v5i1.15785>
- Fauzi, A. R., & Suradi, S. (2018). Building the students' English vocabulary for tourism through computer-based corpus approach. *Indonesian Journal of Integrated English Language Teaching*, 4(2), 133–148. <https://doi.org/10.24014/ijielt.v4i2.6450>
- Khanda, G., & Kalim, S. (2025). Narrative dynamic in Dawn newspaper on Israel-Palestine conflict: A corpus-based discourse analysis. *Global Political Review*, 10(1), 1–15. [https://doi.org/10.31703/gpr.2025\(X-I\).01](https://doi.org/10.31703/gpr.2025(X-I).01)
- Lailiyah, M., & Setiyaningsih, L. A. (2021). Academic vocabulary of EFL students' writing: A corpus-based study. In *Prosiding Seminar Nasional Riset Bahasa dan Pengajaran Bahasa (SENARILIP V)* (Vol. 5, pp. 28–32). Politeknik Negeri Bandung.
- Metang, P., & Narathakoon, A. (2025). A corpus-based study of lexical bundles of keywords found in online news articles. *THAITESOL Journal*, 38(1), 1–21. <https://doi.org/10.55766/thaitesol.v38i1.1042>
- Qureshi, A. H., & Haidov, R. (2020). Developing EFL creative writing with the help of collocations: An English newspapers corpus perspective. *International Journal of Management*, 11(11), 912–922. <https://doi.org/10.34218/IJM.11.11.2020.086>
- Sabiri, M. S., Iqbal, R. H., Batool, R., Abubaker, M., & Zafar, S. (2022). A corpus-based study of prepositional phrases in Dawn newspaper editorials. *VFAST Transactions on Education and Social Sciences*, 10(4), 69–76. <https://doi.org/10.21015/vtess.v10i4.1280>
- Ucar, S., & Yükselir, C. (2015). The effect of corpus-based activities on verb-noun collocations in EFL classes. *Turkish Online Journal of Educational Technology*, 14(2), 195–205.
- Waealee, N., & Waemusa, Z. (2025). Exploring the integration of corpus-based approach in vocabulary instruction in Thai EFL school context: A quantitative report. *Turkish Online Journal of Educational Technology*, 24(4), 91–100.