



Sana Hussan ¹ Salma Kalim ²

Abstract: *This study investigates the efficacy of corpus-based activities in the teaching of English vocabulary. In contrast to traditional methods that depend on textbook exercises and dictionary usage, this research seeks to evaluate the effectiveness of corpus-based activities in vocabulary instruction. Additionally, the study aims to investigate students' attitudes towards corpus-based vocabulary exercises. A total of one hundred graduate-level degree program students from a government institution in Swabi participated in this study. The control group used a dictionary and textbook, whereas the experimental group worked with materials based on concordance. There were two intact groups. Quantitative research instruments were used in this quasi-experimental study. A pre-test, post-test, and online survey were employed as quantitative tools. The test results' statistical analysis showed that employing corpus-based vocabulary exercises excelled using a textbook. Additionally, the results of our study indicated that students felt positively about doing corpus-based vocabulary activities.*

Key Words: Corpus-based Activities, Pre-test, Post-test

Introduction

English has become a global language for people all over the world, and people learn it as a second language. In the current era, it does not serve only as a means of communication across the globe. Rather, it is becoming more and more important as a means of advancing economically and socially. Asian countries, including Pakistan, are giving special importance to enhancing the English Language skills of its young people. With the advent of Pakistan, academicians and officials have put great emphasis on improving the English language, but the results have not been satisfying. While real language input from corpus-based methodologies helps EFL students acquire vocabulary, integrating modern technologies like augmented reality (AR) can greatly enhance students' contextual understanding and engagement. For instance, AR offers a fresh way to learn that could substantially help EFL learners and possibly increase vocabulary retention and comprehension by offering immersive and interactive experiences through the superimposition of digital material onto the actual world.

Problem Statement

Outdated English language teaching methods pose a barrier to non-native Pakistani learners, but the grammar-translation method is still widely used in many educational institutions in the 21st century. Since most teachers concentrate more on getting students through exams than on actually developing their language skills, they frequently use translations into their mother tongue to teach English concepts. Students primarily focus on the completion of their syllabus and passing through exams rather than on genuinely developing their language skills (Channa, 2017). Furthermore, another issue is the lack of qualified English specialists, and the overall quality of English Education is far away from satisfactory, which means it is substandard. There is a significant gap in the adoption of innovative teaching techniques, such as the noticing hypothesis. Despite government initiatives like Language Teaching Reforms and train-

¹ PhD Scholar, Faculty of English, International Islamic University Islamabad, Pakistan.

² Assistant Professor, Faculty of English, International Islamic University Islamabad, Pakistan.

Email: salma.kalim@iiu.edu.pk



trainer programs, university students' proficiency is inadequate due to the traditional emphasis on grammatical structures over practical language skills. This focus on grammatical structures leads to a lack of language proficiency. Additionally, students also fail to recognize formulaic expressions like collocations and idioms, which adversely affect lexical development.

Research Questions

- 1- To what extent do the corpus-based materials influence the vocabulary development of EFL students?
- 2- How do EFL students perceive the effectiveness of the corpus-based materials approach for learning vocabulary?

Significance of the Study

Through the integration of corpus-based learning activities, this study will demonstrate how innovative EFL vocabulary learning strategies can have a positive impact within the Pakistani educational context. With regard to learner engagement, motivation, and academic results, the research is set to provide some useful insights into how technology integration can transform language learning. It aims to set the groundwork for future pedagogical approaches that use technology to get around conventional obstacles in language learning, providing a scalable model with wider applicability in EFL settings.

Literature Review

This section reviews the existing literature on corpus-based teaching methodologies in local and global contexts.

DDL, proposed by (Tim Johns in 1990) as a methodology in language learning, entails students working with corpora, attempting to find patterns, meanings, and usage themselves without direct guidance, as much as opposed to being told (Boulton & Vyatkina, 2021). Hence Tim Johns' idea of making language learners "language researchers" who would explore authentic data by looking for patterns inductively (Boulton, 2012) The main strength of DDL is in the area of learner autonomy because as a Learner becomes a data analyst, a student is allowed to find and analyze the meaning of language on his or her own. This approach is helpful in cultivating acutely analytic and enhanced conceptions of language application since learners are introduced to natural language usage examples in context. For vocabulary learning, DDL has been proposed to state that this method presents learners with the words in various real-life contexts so that they grasp not only the meaning of the word but also the word, as well as the context that the word occupies. It does help in retention and application because it is done contextually. It has been established that DDL facilitates improved comprehension of the sociocognitive context in which words are used and the recall of individual lexical items by training a richer, more embodied manner of language processing (Gilquin & Granger, 2010).

(Raza et al., 2024) examined how Pakistani undergraduate students studying English as a second language acquired vocabulary through corpus-driven language learning. The researchers chose two groups from a public institution as part of their experimental research design for this study. The participants for the study were selected from GC University Faisalabad, where the researchers were faculty members; they were 100 students in the 3rd semester of the English department and then were distributed into two groups. The chosen sample consisted of two groups that attended ELT lessons; the experimental group was given a specific treatment based on observing the hypothesis for the development of collocation knowledge, whilst the control group was given typical book-based instruction. The methodology used in the research is a pretest-posttest design to compare the traditional methods of language teaching with corpus-based teaching and their outcomes. Two types of corpora were used in this article; one is COCA, and the other is self-creation corpora from Dawn newspaper articles and other renowned journal articles to investigate several types of lexical collocations which were extracted from the corpora with the use of the corpus tool named as AntConc. This tool is very useful for teaching vocabulary to university students.

Tuson and Sofu (2023), in their quasi-experimental study, examined the effects of data-driven learning DDL for acquiring a vocabulary of English as a foreign language at a Turkish state university by using a pedagogical corpus. The corpora have contributed indirectly to language teaching by making

dictionaries and textbooks, as well as material development and designing syllabi. Here in this study, the corpora have been directly incorporated into language teaching through DDL, an approach introduced by Jhones (1991) for language teaching. The nature of the study is a comparison of traditional textbook instruction with DDL learning backed by computer-generated corpora for vocabulary learning. The goal of this study was to determine how much participants' vocabulary knowledge improved through DDL practice as compared to traditional textbook-based vocabulary instruction, as well as how the DDL group felt about learning vocabulary through DDL.

Fifty-eight lower-level Turkish students, aged between 15 and 22, were chosen as participants for this experiment. They were split into an experimental group ($n = 29$) and a control group ($n = 29$). Antconc was used as a tool to create corpora from textbooks. Data was collected through vocabulary tests, including pre-tests, post-tests, and delayed post-tests, questionnaires on students' perspectives of DDL, and semi-structured focused interviews. Data was analyzed through SPSS software. The results demonstrate that DDL has advantages for both short- and long-term vocabulary retention, as well as positive student attitudes about this technology-enhanced, participatory approach, as students in the DDL group outperformed their peers in traditional settings. The study emphasizes how pedagogical corpora can be integrated into language instruction, particularly for learners at a lower level. It also suggested a hybrid approach that combines traditional and corpus-based methods. The authors stress the importance of providing teachers with training in corpus utilization, arguing that a well-executed DDL program can greatly improve EFL vocabulary learning, with wider ramifications for language education teaching methodologies and curriculum design.

Szudarski (2022) outlined the impact of corpora on vocabulary studies within applied linguistics since the mid-1980s, emphasizing how corpus approaches are used in vocabulary research and instruction. The author applied a corpus to analyze vocabulary in various ways; these included frequency-based descriptions of vocabulary, contextual use of vocabulary, phrasal vocabulary, and formulaic language. Szudarski (2022) suggested that it is the frequency-based analysis that tells us about the essential role of commonly occurring words and phrases through corpus analysis. It also tells us about the high and low-frequency words. This study claimed that scholars have encouraged teachers, trainers, and textbook writers to pay special attention to these high-frequency words as they are the core element of linguistic competence. The author claimed that approximately 8000 of the most frequent words are required for the understanding of a specific text in L2. However, with the spoken discourse this requirement is lower to the 6000 to 7000 words level. The researcher explored further to identify technical and academic vocabulary types. In terms of academic vocabulary, he referred to Coxhead's (2000) Academic word list, which is influential in corpus studies in this area. Moreover, the focus is on data-driven learning (DDL), which is a pedagogical approach that uses corpus tools to improve vocabulary acquisition directly. It has been demonstrated that DDL considerably raises students' linguistic proficiency in a variety of learning environments. The use of computer-assisted language learning technologies (CALL) in teaching L2 learners complex vocabulary was highlighted by the author. The CALL incorporated vocabulary games and virtual environments, which helped the young children learn vocabulary more effectively. The chapter also specified the need for further research to develop awareness among language teachers about the efficacy of corpus-based tools in the teaching-learning process of languages.

Okataviani (2015) explored the effective use of corpus-driven techniques for teaching vocabulary in classroom activities. The author claimed that the use of corpus is not limited to language research. Rather, it can be used to enrich classroom activities. Vocabulary learning is a key aspect of learning to be proficient in the target language despite low competence in the grammar of the target language. In the traditional method of teaching, words were taught as individual units, but as we know, words do not exist in isolation; they always exist in a context that adds meaning to the words. Therefore, learning the nearby words is important to know the actual meaning of a word in the use of language. This paper discusses the application of the BNC as the largest corpora in teaching vocabulary, along with its shortcomings and benefits. She outlines how classroom activities such as examining the word frequency, collocation, and usage in real-life contexts are enhanced by the British National Corpus BNC, which consists of a large collection of real-life usage texts. In the methodology part, the author discussed the direct and indirect application of corpus-assisted techniques for teaching vocabulary, claiming that the corpus provides a large pool of authentic



language knowledge that surpasses the limited and artificial knowledge that traditional books comprise. She has used the wordsmith tool to sort out words at British National Corpus. Furthermore, Okataviani (2015) also addressed certain shortcomings in the implementation of DDL in classrooms, such as teachers' and learners' literacy in technology use, their preparedness, the complexity of corpus data, appropriateness for students' level, resource intensity, and pedagogical adjustments. I have found the methodology of this article very helpful for my research article.

Ma and Mei (2021) provided a detailed review of the potential of the corpus tools to enhance vocabulary teaching and learning, promoting a new approach to language learning known as corpus-based language pedagogy (CBLP). They emphasized that students can become autonomous language learners by using corpora to investigate real language data and find answers to their queries regarding appropriate language usage. This study focused on giving practical guidance for teachers who want to use corpus linguistics to enhance their classes by breaking down four fundamental design concepts into real-world examples, testing student knowledge, summarizing the language use pattern, hands-on corpus searches by students, and practice exercises. The strength of this study lies in the exploration of the corpus tools like lextutor, which could help language teachers in teaching vocabulary with authentic grounds, and COCA, which provides a large corpus of naturally occurring language patterns for language teaching and learning and an introduction of a new approach to corpus-based language teaching. Moreover, this article gives workable solutions to overcome the challenges that are faced by language teachers in incorporating the corpus tools and using technology-based teaching materials. However, I feel that the authors in this study have fully succeeded in demonstrating the effectiveness of the corpus tools in the language teaching and learning process, but they did not talk about how these corpus-assisted methods, compared to the traditional textbooks and dictionary-based vocabulary teaching methods, can help students retention and long-term learning outcomes. Despite this, the (Ma & Mei 2021) work filled a gap between corpus linguistics and language teaching by providing valuable insight and practical strategies for employing corpus tools in vocabulary teaching. aching-learning process of languages.

Paker and Ergul (2017) studied the effectiveness of corpus-based materials for teaching vocabulary in English at a graduate level. The participants involved in this study were 28 Turkish students of a private university aged between 17 and 20. The participants were divided into two groups: the experimental group and the control group. The experimental group was taught with corpus-based vocabulary teaching activities, while the control group was taught vocabulary exercises from textbooks and a dictionary. The methodology adopted for this study consisted of a proficiency test for the admission process in a program, a pretest at the start of a semester for language learning class, and a post-test at the end of a course in order to know the results of the various ways of teaching. The outcomes highlighted the usefulness of vocabulary acquisition through corpus-based materials over conventional textbook teaching methods. This study demonstrates that at the post-test level, the students showed a considerable improvement in students' vocabulary who were taught with corpus-based vocabulary material compared to those of the control group. Moreover, the study found a positive attitude toward corpus-based teaching, which was expressed by the students in their interview responses and in their personal diaries. The students favored the contextual uses of corpus examples that were offered in the concordance lines, which has made language structure and usage easier to understand. The study emphasizes the importance of using corpus-based materials for teaching English as a foreign language (EFL), especially when it comes to vocabulary development. It implies that these kinds of resources not only improve learning results but also have a favorable impact on learners' attitudes toward vocabulary study. The authors suggested conducting more studies to investigate the long-term efficacy of corpus-based vocabulary instruction over longer periods and with bigger and more diverse participant groups. However, I found this research study helpful in knowing about students' attitudes towards corpus-based learning. By offering empirical for the superiority of real-language data (corpora) over conventional techniques, this study advances the area of language instruction by promoting change in EFL settings to more realistic, context-rich learning situations.

A review of the scholarship reveals three main themes; (1) the importance of the effectiveness of corpus-based vocabulary teaching over the traditional methods, (2) corpus tools such as lextutor, wordsmith, and Antconc are recognized as essential resources for teaching vocabulary material for the

graduate students and (3) several studies addresses the need for further research to assess the long terms impact of using corpus-based material for teaching vocabulary. This study will help language learners and trainers use corpus-based techniques to teach vocabulary effectively and yield better results. This survey practically investigated the student's positive views regarding corpus-based learning for enriching vocabulary.

Methodology

This section explains the research methodology, the materials and instruments utilized, as well as the setting in which the study was carried out. The participants, sampling techniques, tools used for data collection, and specific procedures followed for data collecting and analysis are all described.

Research Design

Using a quasi-experimental design with pre- and post-tests, the learning impact of corpus-based vocabulary exercises on EFL university-level students was assessed. A quasi-experimental design is a type of research design more commonly used in experimentation that deals with estimating the impacts of an intervention or treatment in a study where it is impossible to make a random allocation between two groups. Compared to true experiments, quasi-experiments involve pre-existing groups or conditions such that participants cannot be randomly assigned to two or more groups; while this makes it less controlled, it affords a better fit to real-life situations. This design is often applied in educational studies when random allocation is not possible because of the restrictions of either ethical or practical nature. (Shadish et al., 2002)

In this present study, the quasi-experimental design was adopted because it is applicable in field conditions where subjects cannot be randomly assigned as normally expected. The students' sampling bias resembles real-life classroom configurations, which increases the generalizability of the results. The study can thus test the two groups with similar pre- and post-tests in order to establish the benefits of corpus-based vocabulary exercise, though it does not randomize the groups. This design will, therefore, offer a good test/research design balance as well as practical utility in an educational setting.

In other words, quasi-experimental designs are most beneficial in language learning studies because they allow for observation of interventions (like the corpus-based learning proposal in this study) on intact classroom groups and afford suggestions that are possessable and useful for teachers (Trevino, 2020). This approach allows the study to evaluate the real classroom environment, and the effects, goals, and objectives of the study are consistent with educational practitioners who endeavor to implement the findings in similar non-randomized environment settings.

Brown and Rodgers (2002) state that outlines the following standards for what constitutes a "real experimental study":

- a) The students are split into two groups at random.
- b) Two treatments based on experiments are given.
- c) Pre- and post-tests are given to both groups in order to assess the effects.

In order to quantify vocabulary acquisition, this study will employ a quasi-experimental design using pre- and post-tests. Through focus group talks and student diaries, qualitative insights into learner attitudes and perceptions will be gathered.

Instruments And Materials

Two groups participated in a non-random selection pre-test and post-test experiment designed to address the research problems. The instruments utilized to gather the data were an attitude questionnaire, a competency exam, and pre-and post-tests. The study's resources included a textbook and corpus-based exercises.

Participants

A sample of fifty bachelor-level EFL students from a Pakistani university is chosen, and they are split into two groups: one that will use a traditional dictionary or wordlist from a textbook memorization task for



learning vocabulary is given as a control group, and another used corpus-based activities by using Sketch Engine to learn the contextual use of words. Target words will be put on the sketch engine in order to find the concordance lines, and with these concordance lines, students will learn the correct use of the words by experiencing the correct use of words in the real language data.

Data Collection Procedure

The researcher developed the corpus-based tasks, tests, and attitude questionnaire that the experiment group would be taught before the experiment began. The University Preliminary English exam was utilized by the researcher as a competency exam to ascertain whether the experimental and control groups were homogeneous. The results of the proficiency test indicated that there was no appreciable difference between the experimental and control groups' language proficiency, indicating that the groups were comparable. For that reason, we started the experiment. The experimental and control groups took a pre-test from the researcher before the experiment started. The researcher used the same test for both the pre- and post-tests. The researcher and an instructor provided teaching to both the control and experimental groups.

In addition, a second instructor helped the researcher coach the experimental group in activities involving corpora. The researcher trained the other instructor on corpus-based activities for teaching, and she also taught the experimental group. The experimental group used corpus-based activities to learn the target words from the weekly lesson plans during a two-week period. The British National Corpus (BNC) was used to construct activities based on terminology derived from corpora. On the other hand, the control group learned the target terms using dictionary-based activities. The researcher gave the control group an hour of instruction on corpus-based vocabulary tasks because the pre- and post-tests contained questions focused on vocabulary.

Both groups completed a post-test after the treatment, in which the researcher included the target vocabulary. After the tests were scored, the data were imported into SPSS (Statistical Package for the Social Sciences) for analysis. Following the two-week trial, every participant in each group answered an attitude survey designed to find out what they thought about using corpus-based activities to teach vocabulary in English.

Data Analysis

The first research question of the study was addressed by comparing the performance of the control and experimental groups of students through quantitative analysis of their post-test results using the Statistical Package for Social Sciences (SPSS). First, the proficiency and pre-test data were analyzed to see if there had been a statistically significant difference between the two groups in terms of vocabulary and reading comprehension. To answer the second research question of the study, the data from the questionnaire were analyzed using the statistical package for Social Sciences (SPSS) 16.0 data editor. The means or frequencies of the items were examined, and data from learner diaries and interviews were gathered and descriptively examined in order to corroborate and clarify the questionnaire results.

Findings and Discussion

Pre-test Mean

Table 1

Group	N	Mean	Standard deviation	P	df
Experimental group	60	8.006	2.81	.19	98
Control group	40	6.38	1.80		

The table displays the pre-test mean and standard deviation for the experimental and control groups. The results showed that the experimental group's mean ($m=8.006$, $sd=2.81$) was higher than the control group's ($m=6.38$, $sd=1.80$). One-Sample Kolmogorov-Smirnov To ascertain whether the data were normal, a test was run. An independent samples t-test was used to assess the homogeneity of vocabulary knowledge between the control and experimental groups after it was determined that the results were

parametric. There was no statistically significant difference between the groups' vocabulary knowledge of English despite the fact that the experimental group's mean score was greater ($p = .19$ ($p > 0.05$)).

Research Question 1: Corpus-based materials influence the vocabulary development of EFL students To explore the initial research topic, which seeks to identify the implications of corpus-based exercises for Turkish intermediate performance level of EFL learners in terms of vocabulary acquisition, the post-test results for both the control group and the experimental group of pupils were examined to see if there were any notable differences between two scores for various groups.

Post-test Mean and Standard Deviation

Table 2

Group	N	Mean	Standard deviation	P	df
Experimental group	60	12.8	2.41	.001	98
Control group	40	9.3	2.61		

* $p < 0.05$

As observed in Table, the experimental group fared better in the post-test than the control group. The statistics showed that the experimental group's mean ($m=12.88$, $sd=2.41$) was higher than the control group's ($m=9.33$, $sd= 2.67$). Once again, the data were parametric. Thus, the independent samples t-test was used to see if there was a statistically significant difference between the two groups. The p-value of .001 ($p < 0.05$) indicates that there was a statistically significant difference between the experimental and control groups' post-test results. More specifically, the use of corpus-based activities in vocabulary instruction had a significant effect on the experimental group's students' performance.

Comparison of the Groups' Gain Scores

Table 2

Group	N	Mean	Std. Deviation	Std. Error Mean
Experimental group	60	8.06	2.816	0.364
Control group	40	6.83	1.801	0.285
After training				
Experimental group	60	12.88	2.419	0.312
Control group	40	9.33	2.674	0.423

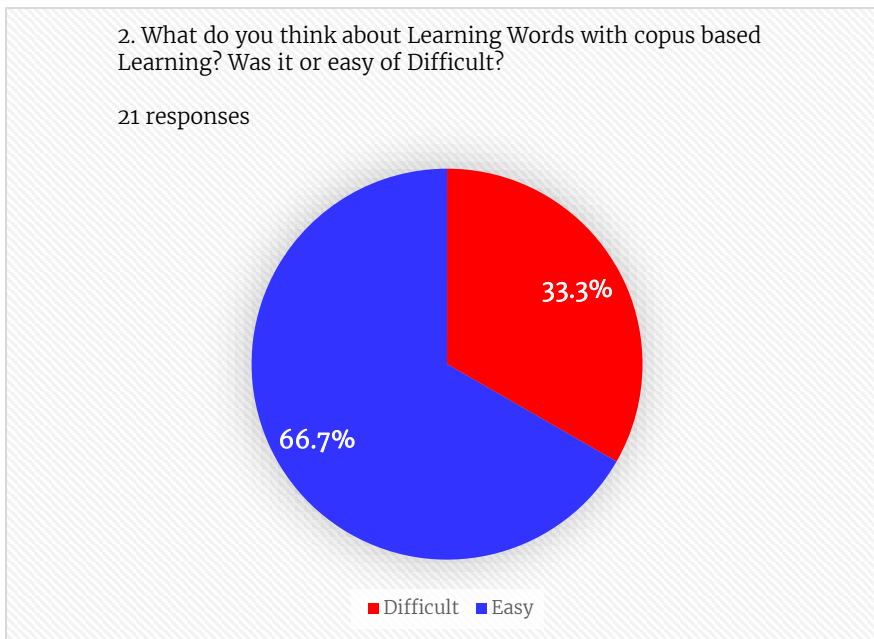
The comparison of the gain scores between groups shows that every student exhibited mastery of the target language. The students in the experimental group did acquire the target vocabulary, albeit not as much as those in the control group. Using corpus-based sources yielded results different from those found in textbook vocabulary exercises; it may, therefore, be inferred. Compared to pupils in the control group, those in the experimental group did better on the post-tests. It attested to the dependability and efficacy of teaching vocabulary through corpus-based methods.

Research Question 2: EFL student's perception of the effectiveness of the corpus-based materials approach for learning vocabulary

The second research question of the study, which attempts to find out how university-level competence level EFL learners feel about using corpus-based vocabulary activities, was also addressed by utilizing a Google Forms survey to examine the students' questionnaire replies. Descriptive statistics and the computation of questionnaire item means were also carried out in order to ascertain the students' attitudes regarding corpus-based vocabulary activities. The mean values of the students' thoughts about teaching vocabulary using corpus-based materials are displayed in the following table.



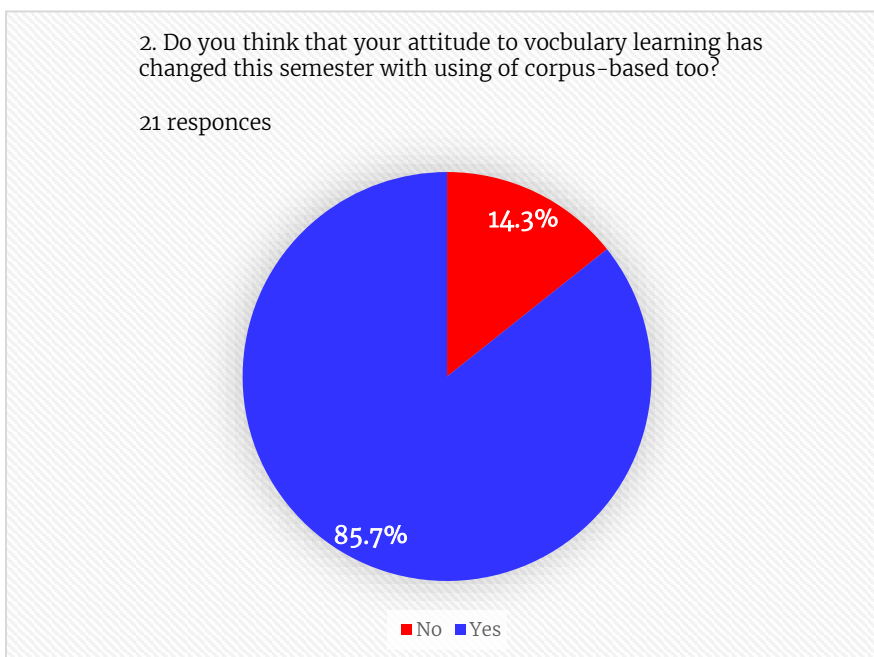
Graph 1



This pie Chart depicts insights into the perception of 21 respondents regarding the challenge of learning English words through corpus-based learning. The vast majority of respondents—two-thirds—thought that learning English words through corpus-based methods was simple. This implies that the majority of students can interact with corpus-based learning resources and derive some value from them. However, it was challenging for one-third of the responses. This suggests that although a significant minority

finds corpus-based learning to be beneficial, there are those for whom it presents difficulties. The majority's enthusiastic response is reassuring for teachers and implies that corpus-based learning strategies ought to be maintained and even extended across the curriculum. The response from the 33.3% of people who found it challenging indicates that more help is required. Helping these students engage with corpus-based learning more effectively could involve providing additional resources, increasing guided practice, or using other teaching techniques. Teachers want to think about carrying out more research to identify the precise corpus-based learning components that these kids struggle with and may find that additional scaffolding or customized instruction would be helpful.

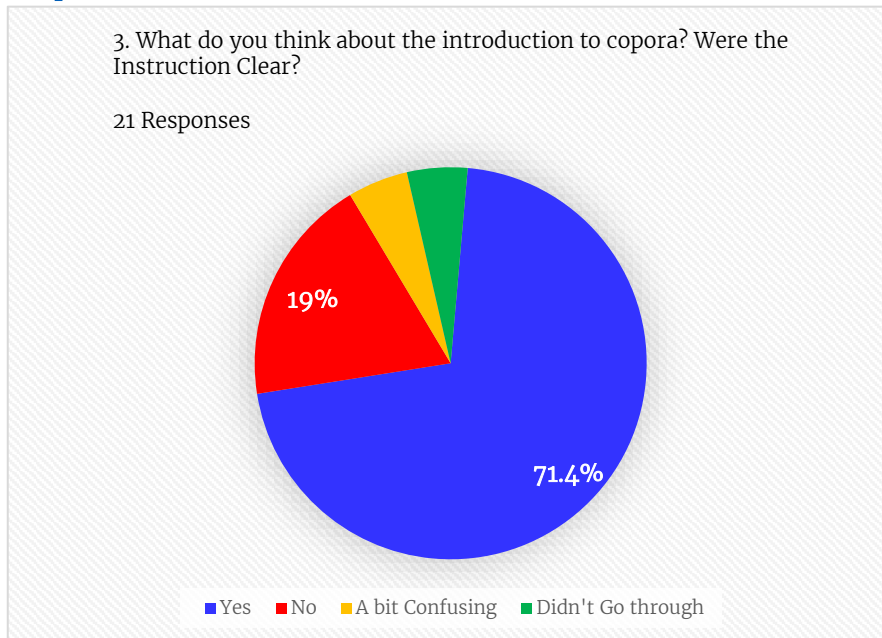
Graph 2



The result, according to the pie chart, claimed that the attitude of the students is positive towards vocabulary learning by using the corpus tool. Approximately 86% of the respondents stated that using a corpus-based tool over the semester had positively impacted their attitude toward vocabulary learning. This represents a significant majority of the respondents. This shows that their method or perspective of vocabulary learning has been improved or influenced by corpus-based technology. On the other hand, a tiny minority—roughly 14%—said they

experienced no change in attitude, suggesting that either they had previously established a good vocabulary-learning strategy that didn't need to be modified or that the tool had no effect on them. The high proportion of favorable responses might encourage teachers to use corpus-based resources more widely in their lesson plans, particularly when teaching vocabulary. Additionally, it implies that students are likely to find corpus-based tools beneficial and that they can improve their learning process.

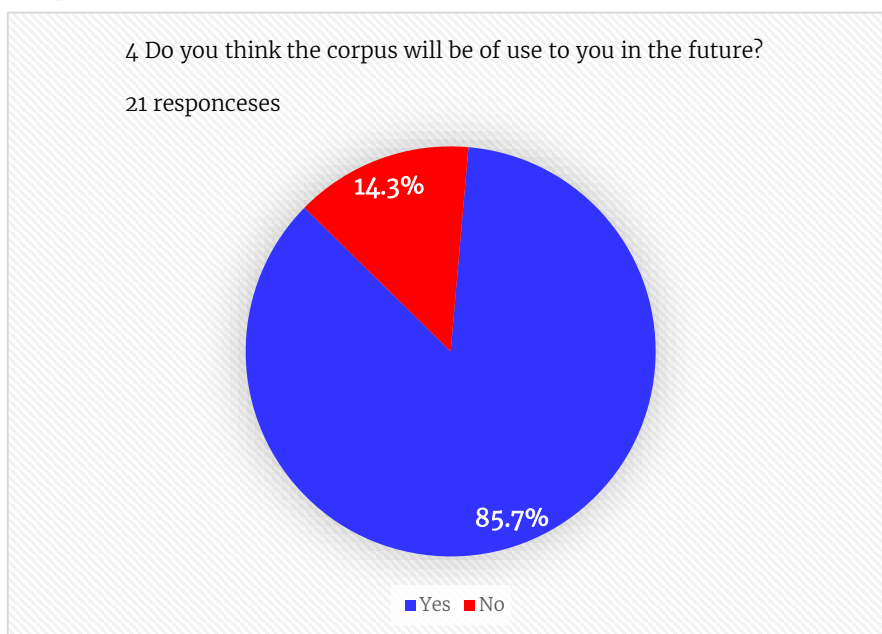
Graph 3



The introduction to corpora was effectively presented for the majority of respondents, as seen by the vast majority (71.4%) who thought the instructions were clear. This high percentage indicates that the majority of the introduction's materials were clear and easy to understand. On the other hand, 9.5% of respondents said the directions were completely unclear, and 19% believed they were a little confused. This total of 28.5% indicates that there is potential for enhancement in the organization or presentation of the instructions. Although the

majority's favorable response is encouraging, the educational approach should think about changing the methods or teaching materials to better meet the needs of the roughly 30% of students who felt that the directions were unclear or confusing. To guarantee clarity, this can entail making difficult terminology simpler, supplying additional examples, or facilitating a more engaged tutorial session.

Graph 4

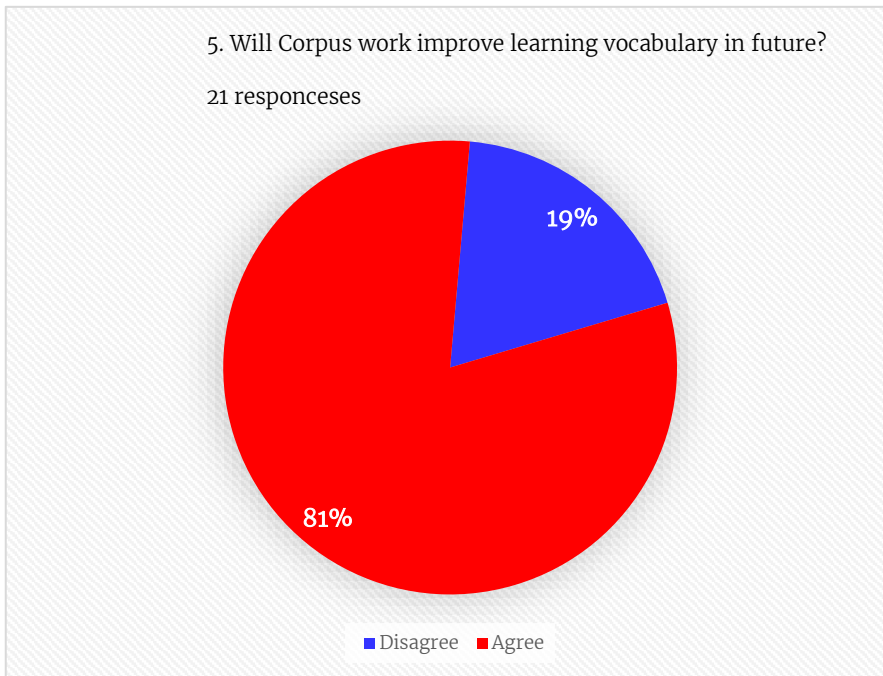


More than 85% of the respondents think they will benefit in the future from the corpus they are using. This suggests that they have a very favorable opinion of the value and applicability of corpus-based technologies for their continuing education or work. 14.3% of respondents believe the corpus will not be useful in the future. This may indicate that the respondents did not think the current corpus-based activities were relevant or useful, or they did not foresee needing such tools for their professional or academic

endeavors in the future. Given the extremely positive feedback, it is recommended that educators and developers keep incorporating corpus-based tools into educational programs and consider expanding their use, especially in linguistics and language learning-related fields. Teachers could think about offering more training or context to show how these techniques can be used successfully in a variety of academic or professional contexts for the minority of students who did not find it useful. To learn more about the reasons behind the small fraction of respondents who do not find the corpus useful, more research may be helpful. The specific comments provided by these respondents may highlight curricular gaps that need to be filled in or limitations in the way it is currently being implemented. Regardless of their present opinions, a follow-up survey or qualitative interviews could assist in obtaining extensive insights into how all students perceive the integration of corpus-based tools in their learning environment and how these tools could be improved for greater long-term utility.



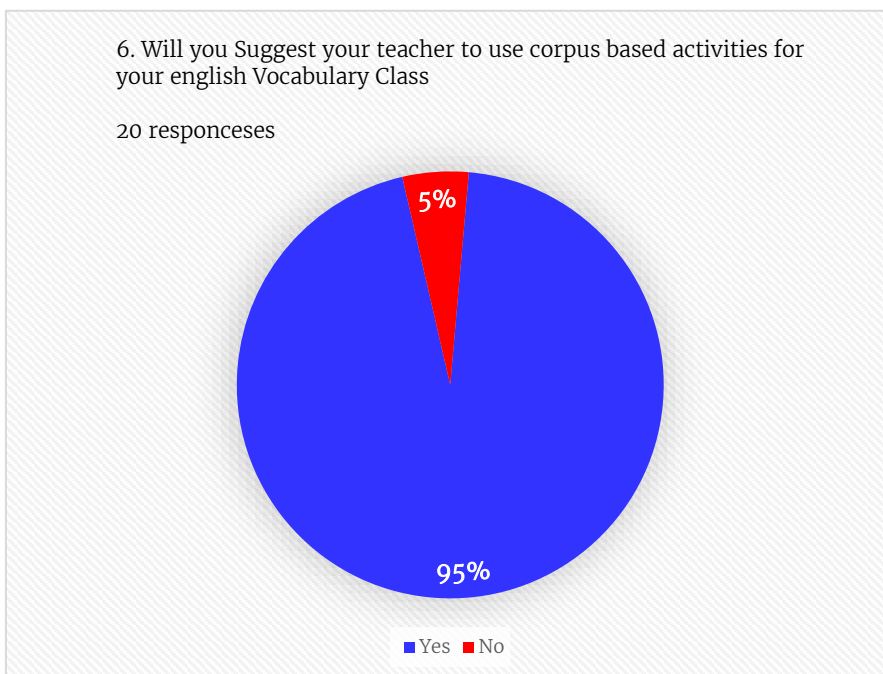
Graph 5



The pie chart presents the views of 21 participants regarding the potential for corpus work to enhance vocabulary learning in the future. A large majority of 81% express agreement, indicating a strong belief in the effectiveness of corpus-based approaches to promote vocabulary acquisition. A sizable minority, 19%, disagree, suggesting some doubt about these tools' efficacy or a preference for different teaching methods. This difference in viewpoints emphasizes how crucial it is for instructors and curriculum designers to continue incorporating corpus-

based learning into classroom environments while also taking into account the preferences and worries of people who are less certain. Knowing the particular causes of the conflicts may facilitate the improvement of these approaches to guarantee that they satisfy the varied requirements of every student, thus raising the possibility of their acceptance and efficacy.

Graph 6



The responses received from 21 participants about whether or not they would suggest corpus-based activities for English vocabulary classes are displayed in a pie chart. Ninety-five percent of those surveyed said they would recommend corpus-based activities to their teacher. This resounding affirmation is indicative of the activities' favorable reception and general perceived efficacy in improving vocabulary acquisition. On the other hand, just 5% of respondents said they would not suggest using these techniques, which could mean they were unhappy with the strategy or

preferred alternative teaching approaches. According to the research, most students seem to enjoy corpus-based activities, maybe because they can offer real-world context and usage examples that enhance and personalize learning. For teachers, this can entail expanding or incorporating corpus-based approaches into their lesson plans in order to accommodate the preferences of the majority. Concurrently, comprehending the causes of the small group's resistance may aid in modifying the exercises to accommodate a greater variety of learning styles and requirements, therefore enhancing their efficacy as a whole.

Findings and Discussion

As opposed to conventional textbook and dictionary-based methods, the study's findings clearly show that corpus-based activities are more effective at improving vocabulary acquisition. Context-rich learning materials and real-language data have a considerable positive impact on post-test scores for the experimental group, as demonstrated by this. Students seem to find corpus-based learning to be both effective and enjoyable based on their favorable evaluations of it. This finding is well-aligned with the existing scholarship. As Yilmaz and Soruc (2015) also found, teaching vocabulary with the help of corpus tools yields the best results in learning the new words of the target language. Most students expressed gratitude for the instructions' clarity and the possible long-term advantages of utilizing corpus-based technologies. The effective application of such cutting-edge teaching techniques depends on this favorable perception.

In this case, therefore, it is clear that the subjects in the experimental group benefited from the corpus-based materials by increasing their vocabulary. It was also analyzed that there was a significant difference between the two groups as the experimental group's post-test mean score of 12.88 (SD=2.41) was significantly higher as compared to the control group, where the mean score was 9.33 (SD=2.67). The p-value was found to be 0.001, which was less than the system-generated $p < 0.05$. There is a precedent for the observed improvement in vocabulary retention and collocation comprehension in Pakistani students as a result of the corpus-driven approaches identified by (Raza et al., 2024; Paker & Ergul, 2017). Turkish students dramatically improved their language learning while using the AntConc and Sketch Engine in the present study, which provided students with real language data and helped them to work with words in context. Hence, the findings of this study are well-aligned with the findings of Al-Mahbashi et al. (2015) and (Raza et al., 2024).

Furthermore, the students in the experimental group were introduced to terms in many different contexts through the use of self-composed corpora and corpus BNC. This improves their vocabulary and the way they comprehend the usage of the words they have learned. This is in agreement with Okataviani (2015), who noted that, especially compared to traditional approaches, CBL offers broader contextualized language experience that leads to longer-lasting vocabulary acquisition.

The attitudes toward corpus-based learning revealed by students through the attitude survey were generally positive. Eighty-six percent of the participants stated that their attitudes towards the acquisition of the vocabulary enhanced when the corpus-based tools were used; most of them noted that the exercises were practical and could be applied in real-life situations. It is confirmed by the conclusion made by Tuson and Sofu (2023), who noted that similar to the students in our DDL-based study, students also expressed rather positive attitudes towards learning vocabulary through concordancing activities, citing higher engagement and improved self-confidence in their vocabulary use.

Despite the fact that the majority of the students noted corpus-based approaches as effective, 86% of the students' attitudes toward LSP did not change at all. This indicates that perhaps corpus-based techniques are not suitable for all learners. This is in tune with the concern raised by Szudarski (2022), who acknowledged that to enjoy the full advantages of learning based on the corpora, lower-level students or those who are not very inclined toward technology may find it hard to comprehend the intricacies of the corpus technology may require and may need more guidance or other forms.

Conclusion

The purpose of this study was to evaluate the effectiveness of vocabulary education using corpus-based exercises. An additional purpose of the research was to look at students' perspectives on vocabulary exercises based on corpora. A statistically significant difference in scores was observed between the experimental and control groups in the post-test findings, with the experimental group scoring higher. Students' attitudes toward using corpus-based vocabulary activities were good, according to the results of the attitude questionnaire. Additionally, the findings of the online poll showed that students felt favorably about vocabulary education that made use of corpus-based resources. The results of the questionnaire and post-test showed that students benefited from vocabulary instruction using corpus-based activities. Thus,



it may be argued that reading dictionaries or textbooks was not as beneficial as using corpus-based vocabulary resources.

Limitation of the Study

In the first study, one hundred graduate-level EFL students enrolled in a program took part. Subsequent research endeavors may explore multiple proficiency levels (basic, intermediate, and advanced) to enhance the overall applicability of the findings and assess the effectiveness of corpus-based resources across a broader language spectrum. In addition, extending the study's duration beyond a single semester would enable an analysis of these learning approaches' long-term impacts. Further insights into the factors influencing vocabulary learning and program efficacy may be obtained by combining data on student attrition, attendance rates, and access to online resources. With these thorough modifications, the study framework would be improved, and more reliable findings would be provided that could be applied to the creation of language curricula in higher education.

Future Research Recommendation

Future research needs to examine how vocabulary retention and understanding are affected over time by corpus-based learning. To strengthen the generalizability of the findings and validate them, research should also be done with bigger and more varied participant groups and across a range of competency levels. The integration of information on student attrition, attendance rates, and online resource accessibility may yield additional perspectives on the variables impacting language acquisition and program effectiveness.

References

- Al-Mahbashi, A., Noor, N. M., & Amir, Z. (2015). The effect of data-driven learning on receptive vocabulary knowledge of Yemeni University learners. *3L, Language, Linguistics, Literature*, 21(3), 13–24. <http://journalarticle.ukm.my/9065/1/9511-27888-1-PB.pdf>
- Boulton, A. (2012). What data is used for data-driven learning? *EuroCALL Review*, 20(1), 23–27. <http://dx.doi.org/10.4995/eurocall.2012.16038>
- Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology*, 25(3), 66–89. <http://hdl.handle.net/10125/73450>
- Brown, J. D., & Rodgers, T. S. (2002). *Doing second language research: An introduction to the theory and practice of second language research for graduate/master's students in TESOL and applied linguistics and others*. Oxford University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213. <https://doi.org/10.2307/3587951>
- Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching? In *The Routledge Handbook of corpus linguistics* (pp. 359–370). Routledge.
- Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177–190. <https://doi.org/10.1515/jccall-2021-2008>
- Oktavianti, I. N. (2015). Data-driven Learning in the Classroom: The Use of British National Corpus in Teaching Vocabulary - Repository Universitas Ahmad Dahlan. *Uad.ac.id*. http://eprints.uad.ac.id/2521/1/data-driven_learning_in_classroom.pdf
- Paker, T., & Ergül Özcan, Y. (2017). *The effectiveness of using corpus-based materials in vocabulary teaching*. [master Thesis]. <https://hdl.handle.net/11499/2177>
- Panezai, S. G., & Channa, L. A. (2017). Pakistani government primary school teachers and the English textbooks of Grades 1–5: A mixed methods teachers'-led evaluation. *Cogent Education*, 4(1), 1269712. <https://doi.org/10.1080/2331186x.2016.1269712>
- Raza, A., Shah, M. I., & Imran, S. F. (2024). The Innovative Pedagogy and Foreign Language Learning: Exploring the Effect of Corpus-Driven Language Learning on Lexical Knowledge of Pakistani Undergraduates. *Kurdish Studies*, 12(2), 5975–5992. <http://dx.doi.org/10.58262/ks.v12i2.444>
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasiexperimental designs for Generalized causal inference*. Boston, MA: Houghton Mifflin.

- Szudarski, P. (2022). Corpora and teaching vocabulary and phraseology. In *The Routledge Handbook of Corpora and English Language Teaching and Learning* (pp. 41-55). Routledge.
- Tosun, S., & Sofu, H. (2023). The Effectiveness Data-Driven Vocabulary Learning: Hands-on Concordancing through a Pedagogical Corpus. *Journal of Language and Education*, 9(3), 176–190. <https://doi.org/10.17323/jle.2023.12426>
- Yılmaz, E., & Soruç, A. (2015). The use of Concordance for Teaching Vocabulary: A Data-driven Learning Approach. *Procedia - Social and Behavioral Sciences*, 191, 2626–2630. <https://doi.org/10.1016/j.sbspro.2015.04.400>



Appendix A

Proficiency Test

Name: _____

Roll No: _____

Instructions: Choose the most appropriate word or phrase that fits all the blanks in each set of lines.

- The team had to _____ their efforts to meet the deadline. They decided to _____ more resources into the project. Despite the challenges, they managed to _____ through and complete the task on time.
 - Adjust
 - Allocate
 - Push
 - Follow
- The scientist's _____ on climate change was groundbreaking. His _____ was well-received by the academic community. He aimed to make a significant _____ in the field of environmental science.
 - Speech
 - Hypothesis
 - Work
 - Idea
- The company aims to _____ its market share by 10% this year. To _____ this goal, they will need to _____ their product line and improve customer service.
 - Maintain
 - Achieve
 - Develop
 - Hold
- The new policy will _____ the way businesses operate in the region. It is expected to _____ economic growth and _____ investment.
 - Transform
 - Boost
 - Hinder
 - Limit
- She has a _____ passion for helping others. Her _____ commitment to her work is admirable. Everyone appreciates her _____ dedication.
 - Profound
 - Limited
 - Lukewarm
 - Superficial
- The researchers used a _____ method to gather data. They _____ the participants thoroughly. The _____ results provided valuable insights.
 - Comprehensive
 - Partial
 - Simple
 - Quick
- The artist's latest work _____ the essence of modern society. It _____ different aspects of human behavior. The _____ piece was widely acclaimed.
 - Reflects
 - Shows
 - Hides
 - Disrupts

8. To _____ the project, we need to _____ clear objectives. It is important to _____ progress regularly.
 - a) Finalize
 - b) Set
 - c) Track
 - d) Support

9. The manager needs to _____ the team's performance. He should _____ any issues early. Effective _____ can lead to better outcomes.
 - a) Observe
 - b) Manage
 - c) Address
 - d) Monitor

10. The conference will _____ experts from various fields. It aims to _____ discussions on innovative solutions. The event is expected to _____ new partnerships.
 - a) Attract
 - b) Hinder
 - c) Limit
 - d) Prevent

11. The new software will _____ productivity. It helps to _____ time spent on repetitive tasks. The _____ efficiency will benefit the entire organization.
 - a) Decrease
 - b) Boost
 - c) Limit
 - d) Hinder

12. The book's _____ is complex and engaging. It _____ multiple storylines. The _____ is filled with unexpected twists.
 - a) Theme
 - b) Plot
 - c) Cover
 - d) Title

13. To _____ a successful negotiation, one must _____ effective communication skills. It is important to _____ the other party's needs.
 - a) Ensure
 - b) Develop
 - c) Ignore
 - d) Compromise

14. The organization aims to _____ awareness about environmental issues. They _____ various campaigns to educate the public. The _____ effort has had a significant impact.
 - a) Raise
 - b) Lower
 - c) Ignore
 - d) Limit

15. The professor's lecture was _____ and engaging. He _____ complex concepts with ease. The students _____ the clarity of his explanations.
 - a) Confusing
 - b) Clear
 - c) Dull
 - d) Boring



Appendix B

	Details	Left context	KWIC	Right context
1	<input type="checkbox"/> <input type="radio"/> Written books a...	>On the recommendation of the Public Accounts Committee, the report was compiled for the Commission by academics who analyzed the cost-benefi		
2	<input type="checkbox"/> <input type="radio"/> Written books a...	unwarranted and psychologically traumatic.</s><s>One of those involved in compiling the report, Terence Lee of the University of Surrey, argued that r		
3	<input type="checkbox"/> <input type="radio"/> Written books a...	ry lorries because of serious road damage, according to a confidential study compiled by the Transport and Road Research Laboratory, details of whic		
4	<input type="checkbox"/> <input type="radio"/> Written books a...	ir</s><s>Acid rain damage in Britain spreading</s><s>Unpublished figures compiled as part of research by the Department of the Environment revea		
5	<input type="checkbox"/> <input type="radio"/> Written books a...	UK, according to a study commissioned by Greenpeace.</s><s>The report, compiled by health specialist Cathv Read, warned that traffic exhaust may		
6	<input type="checkbox"/> <input type="radio"/> Written books a...		▼	d to be" a very
7	<input type="checkbox"/> <input type="radio"/> Written books a...		***	November</s><
8	<input type="checkbox"/> <input type="radio"/> Spoken context...	Forestry Commission review </s><s> A review conducted on the part of the Forestry Commission assessing" the returns to the		ard, the South /
9	<input type="checkbox"/> <input type="radio"/> Written miscell...	economy and to society from investments in forestry", has concluded that conifer plantations are economically justifiable. </s><s>		IMB).</s><s>A
10	<input type="checkbox"/> <input type="radio"/> Written miscell...	On the recommendation of the Public Accounts Committee, the report was compiled for the Commission by academics who		where the leve
11	<input type="checkbox"/> <input type="radio"/> Written miscell...	analyzed the cost-benefits of particular types of forestry. </s><s> In environmental terms, it concluded that conifer planting can		text critical app
12	<input type="checkbox"/> <input type="radio"/> Written miscell...	help alleviate global warming because of the rate at which these trees take up, and store, carbon. </s><s> Conifers can thrive on		earch purposes
13	<input type="checkbox"/> <input type="radio"/> Written miscell...	upland		ations.</s><s>
14	<input type="checkbox"/> <input type="radio"/> Written miscell...		***	lar data needs
15	<input type="checkbox"/> <input type="radio"/> Written miscell...			ious geographi
16	<input type="checkbox"/> <input type="radio"/> Written miscell...			se not been 20

1. Compile

2. What is category of the word compile?
3. How many inflections are there for the word compile?
4. Which word category is used before and after the word compile?

2. Comprehension

	Details	Left context	KWIC	Right context
1	<input type="checkbox"/> <input type="radio"/> Written miscell...	us for leaving them a vast amount of dross which makes the location and comprehension of useful material more difficult.</s><s>There will be those who argue tha		
2	<input type="checkbox"/> <input type="radio"/> Written books a...	it social formative function, such functions are" achieved despite design's comprehension ").</s><s>But worse, at this point design can" no longer be aware of itself		
3	<input type="checkbox"/> <input type="radio"/> Written books a...	been since the prerequisites of a truly rational model of design activity – comprehension of all of design's moments as a prelude to modelling what whole they con		
4	<input type="checkbox"/> <input type="radio"/> Written books a...	skins begins: I shall take your mind on a journey.</s><s>It is a journey of comprehension , taking us to the edge of space, time, and understanding.</s><s>On it I s		
5	<input type="checkbox"/> <input type="radio"/> Written books a...	in instead of making crvotic comments that taxed her stunned powers of comprehension .</s><s>He laughed deeolv, tracing a finer delicately over the hot swolle		
6	<input type="checkbox"/> <input type="radio"/> Written books a...		▼	of a settlement".</s><s>Th
7	<input type="checkbox"/> <input type="radio"/> Written books a...		***	Although the position is n
8	<input type="checkbox"/> <input type="radio"/> Written books a...	, the income arising in the company did not arise from property comprised in the settlement within the meaning of TA 1988,		shings) aspects of reading.<
9	<input type="checkbox"/> <input type="radio"/> Written books a...	s672(1). </s><s> Lord Romer said at p335" but I must confess that, with all respect to the Commissioners, it passes my		y to enable him to give bindi
10	<input type="checkbox"/> <input type="radio"/> Written books a...	comprehension how the structure of a company can form part of a settlement". </s><s> The court took a very simple view: the		see a client as often as pos
11	<input type="checkbox"/> <input type="radio"/> Written books a...	shares in the relevant company were the assets comprised in the settlement. </s><s> If dividends were paid then the dividends		><s>"Or did it suit your plai
12	<input type="checkbox"/> <input type="radio"/> Written books a...	would derive from assets comprised in the		to keep him occupied.</s><
13	<input type="checkbox"/> <input type="radio"/> Written books a...		***	'surely you didn't imagine I
14	<input type="checkbox"/> <input type="radio"/> Written books a...			ora hastily, moving further :
15	<input type="checkbox"/> <input type="radio"/> Written books a...			anguages – four in all.</s><
16	<input type="checkbox"/> <input type="radio"/> Written books a...			J're not so strange,' she att
17	<input type="checkbox"/> <input type="radio"/> Written books a...			is the one suspected hu Ry

1. Is 'Comprehension' a(n) verb, noun or adjective?
2. What does 'Comprehension' mean?
3. Can you write one sentence using 'Comprehension'?

3. Handsom

	Details	Left context	KWIC	Right context
1	<input type="checkbox"/> <input type="radio"/> Written books a...	ry was a 'rendezvous of fashion and society: Aikin in 1795 said that it was a ' handsom , well-built town... rendered gay by assemblies and other places of amuseme		
			▼	

		. </s><s> Cotton mills and Victorian working-class housing, Preston, Lancashire in 1920. </s><s> Another photograph of an		
		industrial and urban landscape that no longer exists. </s><s> Preston in the eighteenth century was a 'rendezvous of fashion and		
		society: Aikin in 1795 said that it was a ' handsom , well-built town... rendered gay by assemblies and other places of		
		amusement, suited to the genteel style of the inhabitants'. </s><s> The Earl of Derby had a town house here. </s><s> When		
		cotton spinning was introduced in 1777, there were 6000 inhabitants. </s><s> A hundred years		

1. Is 'handsom' a(n) verb, noun or adjective?
2. What does 'handsom' mean?
3. Can you write one sentence using 'handsom'?

4. Hinder

<input type="checkbox"/> Details	Left context	KWIC	Right context
<input type="checkbox"/> Written books a...	duce energy efficiency labelling of electrical goods and to repeal regulations	hindering	reductions of greenhouse gases.
<input type="checkbox"/> Written miscell...	aterials to the historian's of the future may prove to be a significant problem	hindering	the study of history.
<input type="checkbox"/> Written books a...	with the mineral surface eventually becomes saturated with solutes, thereby	hindering	further dissolution.
<input type="checkbox"/> Spoken context...	n then er the policies of the rich countries in the West are erm a actively erm	hindering	the development prospects of developing countries
<input type="checkbox"/> Written books a...	the attitudes of both the social elite and the labor movement have served to	hinder	economic growth.
<input type="checkbox"/> Written books a...	▼		
<input type="checkbox"/> Written books a...	...		
<input type="checkbox"/> Written books a...	six headings: Europe, The World, Land Use/Planning, Pollution Control, Research/Monitoring and Government.		The main
<input type="checkbox"/> Written books a...	oints are as follows: *Europe.		The European Community will be urged to introduce energy efficiency labelling of
<input type="checkbox"/> Written books a...	electrical goods and to repeal regulations	hindering	reductions of greenhouse gases.
<input type="checkbox"/> Written books a...	system is suggested.		The government will also urge the EC to bring forward the phasing out of chlorofluorocarbons
<input type="checkbox"/> Spoken context...	(CFCs) to 1997.		Financial and technical support to aid developing countries in their efforts to
<input type="checkbox"/> Written books a...	...		

1. Is 'hinder' a(n) verb, noun or adjective?
2. What does 'hinder' mean?
3. Can you write one sentence using 'hinder'?

5. Hypothesis

<input type="checkbox"/> Details	Left context	KWIC	Right context
<input type="checkbox"/> Written books a...	1) unless there is correlation between the amplitudes c1 and c2.	hypothesis	, such a correlation is induced by H" (Rabi-flopp
<input type="checkbox"/> Written books a...	ard Riemann (1826–66), whose university teacher's thesis of 1854 'On the	hypotheses	which underlie geometry' (published 1868) can
<input type="checkbox"/> Written books a...	id puzzled confrontation which occurs when there is a clash, not of different	hypotheses	, but of different ways of looking at the same pr
<input type="checkbox"/> Written miscell...	most forms of quantitative research which use such data to test predefined	hypotheses	. I will leave to one side any consideratic
<input type="checkbox"/> Written miscell...	gel and Elton 1983).	hypotheses	about the nature of slavery in the ante-bellum S
<input type="checkbox"/> Written miscell...	▼		
<input type="checkbox"/> Written books a...	...		
<input type="checkbox"/> Written books a...	original can be interrogated in an open-ended way in the light of unforeseen enquiry.		[6] If we wished to test a new theory
<input type="checkbox"/> Written books a...	about Napoleon's allergy to snuff, say, it would not make sense to examine look-alikes of Napoleon's clothing.		Prior to the
<input type="checkbox"/> Written books a...	snuff-allergy	hypothesis	, snuff-content would not be a consideration in the making of a garment replicas.
<input type="checkbox"/> Written books a...	original artefact with authenticated provenance would suffice for this forensic purpose.		However, logical replication as
<input type="checkbox"/> Written books a...	distinct from physical replication seems to offer more.		Capturing the operational persona of an early
<input type="checkbox"/> Written books a...	...		
<input type="checkbox"/> Written books a...	...		

1. Is 'hypothesis' a(n) verb, noun or adjective?
2. What does 'hypothesis' mean?
3. Can you write one sentence using 'hypothesis'?