

# Using Corpus Linguistics as a Valuable Teaching Tool to Teach English Language to BS Students at a Government University

<sup>1</sup>Muhammad Anwar, <sup>2</sup>Arif Khan Masood, <sup>3</sup>Saqib Abbas

<sup>1</sup>Senior Lecturer, Humanities & Social Sciences, Bahria University Karachi

Campus,

<sup>2</sup>Senior Lecturer, Humanities & Social Sciences, Bahria University Karachi Campus,

<sup>3</sup>Assistant Professor, Department of English, Benazir Bhutto University, Lyari  
Karachi

**Corresponding author:**

Muhammad Anwar

[muhammadanwar.bukc@bahria.edu.pk](mailto:muhammadanwar.bukc@bahria.edu.pk)

## Abstract

This study discusses the use of corpus linguistics as a valuable teaching tool to teach English compulsory to BS students and certain issues which arise when using corpus linguistics as a teaching tool to teach English language to BS students at a government university in Karachi. Corpus linguistics is an important area within applied linguistics, but it is also a valuable tool to teach and learn a language through computers and books. This study shows the effectiveness of corpus linguistics for teaching English language at BS level. It was quantitative research in which an experimental design was used. For this purpose, the data was collected from 54 students at a government university studying at BS 1st Semester. The findings of this study suggest that using corpus linguistics as a tool plays a significant role in making students know about the usage of words with different prepositions and multiple meanings, their frequency and about the natural way of using English language (collocation) which are very important for the learners who are learning English as a second or foreign language.

**Keywords:** *BS students, corpus linguistics, a government university, dictionaries, target language*

## 1. Introduction

This chapter explains how this study is conceived as a result of my observations as an English teacher at a government University. I have been teaching various subjects of English linguistics and English literature to BS students for more than 14 years. I have found many resources for English language teaching- one of them is corpus linguistics which is a very important resource or tool to teach English as a second language at a BS level.

Let's first of all, discuss what corpus linguistics is. Corpus (Thornbury, 2002) is a huge collection of written or spoken language which is used to study any language, whereas

linguistics is a scientific study of any language; it is also a study of its structure, grammar and history. In corpus linguistics, the word corpus means a body of text which may be made up of different examples of spoken or written language or of both. The word ‘corpora’ is the plural form of corpus that may be based on simple, short texts on narrow topics, or they may span millions of words, such as the British National Corpus, a 100-million-word corpus of British English. A corpus is plain text consisting of a single word or phrase. Alternatively, corpora may be tagged for grammar or other functions. A simple search may find out the frequency of various words and constructions (Thornbury, 2002). Corpus is also found in dictionaries like Oxford Advanced Learner’s Dictionary, Cambridge Advanced Learner’s Dictionary, Longman Advanced Learner’s Dictionary, Fine online Dictionary, Webster Dictionary, Collins Dictionary, etc.

### **1.1 Aims and Objectives of the Study**

The aim of this study is to identify the practices of using corpus linguistics as a tool or resource for teaching English language effectively to BS students at a government university. The subject of English compulsory has been added to their program to enable them to speak and write English well. In other words, this course has been designed and added to BS program to enhance and improve their speaking and writing skills so that they do not face any problems related to English language.

### **1.2 Research Objective**

To check the effectiveness of using corpus linguistics as a teaching tool for teaching English language to the students of BS at a government university

### **1.3 Research Question**

How effective is it to use corpus linguistics as a teaching tool to teach English language to the students of BS at a government university?

### **1.4 Significance of the Study**

English language is vitally important in Pakistan, India, Sri Lanka, Bangladesh and many other countries. Nobody can achieve success in any career without English. Since indigenous languages are not equipped for these roles and English has been provided for a convenient vocabulary or the use of this language is deemed to be powerful and prestigious, English has become the language of the legal system, regional administrative network, higher education, sciences and technology, trade and commerce. (Anwar, 2018). Corpus linguistics can play a very important role in providing a tool to teach learners about the usage (collocation), frequency and the multiple meanings of words in English language.

## **2. Review of Literature**

### **2.1 Nature of Corpus Linguistics**

Corpus linguistics is a rapidly expanding technique within the field of linguistics that allows for the study of language patterns. (Gries, 2009). A language corpus, according to Biber, Conrad, and Reppen (1998), is a systematic collection of written and spoken texts that is available for qualitative and quantitative study. Aijmer (2002) emphasizes that corpora denote actual language use and how language is used in different registers. The information we have in the Collins Corpus serves as the foundation for all COBUILD dictionaries. There are 4.5 billion words in the whole corpus. The Bank of English™ is a subgroup of 650 million words. These words have been chosen carefully from different sources. These data give an accurate detail of usage of modern English. Lexicographers usually use the Bank of English for writing dictionaries. The data provides information on word usage, meaning, word combinations, and frequency of use. This data on frequency helps in selecting the words for the COBUILD dictionaries.

Corpus linguistics and empirical linguistics are used interchangeably (Sampson 2001: 6). However, it has been claimed that corpus linguistics is more than merely a methodology, though it is not a linguistic theory in itself; rather, it is “a new research enterprise” and “a new philosophical approach to the subject” (Leech 1991: 106). Hunston (2002: 20) makes clear that computers serve two distinct purposes: they make it easier to gather and store vast amounts of linguistic data and they make it possible to develop the software needed to access and analyze the corpus data.

## 2.2 Spoken and Written Corpora

We have access to a wide range of techniques through corpus linguistics to study language, its patterns and structure. Even a quick analysis of the area shows that there are a great many written and spoken corpora (O’Keeffe, McCarthy & Carter 2007). The British National Corpus (BNC), which has 100 million words of both written and spoken English, is one of the most often cited corpora. Newspapers, novels, letters, and essays make up 90% of the corpus’ textual material (in written form), while business meetings, phone-ins, and radio broadcasts account for the remaining 10% (in spoken form) (Thornbury, S. 2002).

Moreover, there are more than 410 million words in the Corpus of Contemporary American English (COCA) which is a free available corpus. Between 1990 and 2010, it was developed by 20 million words per year from spoken and written English (academic works, novels, periodicals, newspapers, and spoken English). The Cambridge English Corpus, formerly known as the Cambridge International Corpus, is a corpus of more than one billion words in the English language (with both text corpus and spoken corpus data). Written and spoken data in both British and American English are included in the Cambridge English Corpus (CEC). The Cambridge Learner Corpus, a 40-million-word corpus composed of English exam responses produced by English language learners, is also a part of the CEC.

According to O’keeffe, A., McCarthy, M., & Carter, R. (2007), corpora may be utilized as reference tools by ELT professionals. Teachers and students of English can look up the usage of certain words and phrases in corpora when they are not sure whether they are correct in terms of grammar and lexis. This is crucial for non-native English speakers who frequently have students ask them why specific English words are used "the way they are." Despite the fact that English is their second language, and they are good at it, non-native teachers’ intuitions may not be accurate or reliable. Therefore, teachers may check corpus data if they are unsure to make sure that the language, they are teaching is accurate.

Data from the corpus can also be used directly in the classroom through data-driven learning (DDL). Johns (1991: 2) introduced this paradigm initially, claiming that the language learner "is also a researcher whose learning must be guided by access to

linguistic knowledge". Learners uncover common patterns in grammar and lexis by examining authentic language material, whereas the teacher just helps the entire process. This inductive technique helps students identify which linguistic forms are employed in communicative contexts and increases their knowledge of the function of language in everyday situations.

### **2.3 Corpora**

A corpus is a collection of texts that serves as the primary resource for corpus linguistics. According to Thornbury, S. (2002), corpora come in varied sizes, are assembled for diverse reasons, and are made up of numerous sorts of texts. To a certain extent, all corpora are homogenous; they are made up of texts from one language, one variant of a language, one register, and so on. They are also different to a certain extent because they are made up of a variety of distinct texts. Besides the texts which comprise them, most corpora contain information on the texts, part-of-speech for each word, and parsing metadata.

### **2.4 Information about Word lists and Frequency**

A word list is a list of all the words in a particular corpus, generally sorted alphabetically or in frequency order, together with the information on how many times a particular word appears in the corpus. Simple word lists use 'words' as a string of letters, therefore the number of instances of the word 'run' is provided without the difference between noun and verb, and the details of its family like the words 'runs', 'running', and 'ran' are provided separately. More advanced lists discriminate between the word run's noun and verb occurrences and provide summary statistics for the entire lemma, such as 'run', 'runs', 'running', and 'ran', all of which appear as verbs (Leech et al., 2001). The word lists which differentiate between meanings (for example, between run [move in swift motion] and run [manage an event or organization]) are far more difficult, and indeed not publicly available.

### **2.5 Phraseology**

Phraseology is the study of phrases, which is an important component of corpus linguistics: Sinclair (1991) discovered that the meaning of a word is discovered through a number of words, or phrases. It takes in the study of collocations, lexical bundles, and language that occurs in preferred sequences. Collocation is the most common method of studying phrases.


## **2.6 Collocations**

The word 'collocation' is the combination of words in a language that occurs more frequently than by coincidence. There are students who have good ideas, but they usually lose their marks because they do not know some very important collocations of a key word which are vitally important for their writing (Hill 1999:5). As a result, they develop longer, more wordy ways of characterizing or addressing the problem, raising the likelihood of more mistakes. He gives the following example: "His disability will continue until he dies" rather than "He has a permanent disability." (2000:49-50) There is no secret solution for fixing these errors. Collocation must be learnt through both direct study and a vast amount of quality input. Learners frequently struggle with the notion of collocations. The fundamentally straightforward concept that word choice is severely constrained by what comes before and after it "is perhaps the single most elusive aspect of the lexical system and, thus, the most difficult for learners to acquire" (Thornbury 2002:7).

**Table 1. Types of Collocation**

<b>Verb + noun</b>	give a party / have responsibility
<b>Adjective + noun</b>	large package / complete solution
<b>V. + adj. + noun</b>	take regular exercise / achieve excellent progress
<b>Adverb + verb</b>	completely destroy / deeply regret
<b>Adverb + adjective</b>	absolutely amazed / entirely useless
<b>Adverb + adjective + noun</b>	completely unacceptable behavior
<b>Adjective + preposition</b>	smart for / blamed for / good at
<b>Noun + noun *</b>	court case / arm chair / tea cup

move fast on foot

1 ★  [intransitive] to move using your legs, going faster than when you walk

- Can you run as fast as Mike?
- They turned and ran when they saw us coming.
- She came running to meet us.
- I had to run to catch the bus.
- + adv./prep. The dogs ran off as soon as we appeared.
- He ran home in tears to his mother.
- The children spent the morning running around in the park.
- She can run really fast.

In spoken English **run** can be used with **and** plus another verb, instead of with **to** and the infinitive, especially to tell somebody to hurry and do something: *Run and get your swimsuits, kids.* • *I ran and knocked on the nearest door.*

**Figure 1. Showing collocation of the verb ‘run’**

## 2.6 Register

Much of the comparative work that uses corpora has contrasted the appearance of a language like English in diverse situations. Such settings have been established in accordance with the linguistic theory (for example, in Matthiessen, 2006, where register is defined in accordance with systems theory) or in accordance with a less academic, 'commonsense' understanding of where clear difference may exist. For example, Biber et al. (1999) consider the broad 'register' categories of conversation, news reporting, fiction, and academic prose. Others have made finer distinctions: the CANCODE corpus of spoken English, for instance, differentiates between 'transactional,' 'professional,' 'socializing,' and 'intimate' contexts (Carter, 2004); Hyland Corpus Linguistics 237 (2000) differentiates academic writings such as research articles, book reviews, abstracts, and textbooks from academic disciplines.

Most variations in frequency have been explained by changes in communicative function or rhetorical aim.

## **2.7 Corpus Linguistics in the Classroom**

Corpus linguistics may be used in a variety of ways in the classroom by both teachers and students. A teacher may gather a set of student assignments and use a concordance tool to analyze instances of their language in order to identify common mistake patterns. Errors in learners' writing that are systematic might be exploited to construct instructional resources. Alternatively, a teacher might search known corpora of texts written by native speakers of the target language for examples of patterns and then he/she may provide them to students as examples. Learners can train themselves as to how to use a concordancing program (St. John, 2001) and corpora, and then they can become researchers; they may look for examples and develop their own rules for grammatical structures, general usage and idioms, such as they may investigate the differences between the verbs 'look' and 'see'. This method is known as Data Driven Learning (DDL).

## **3. Methodology**

### **3.1 Research Approach**

This chapter explains the nature of the research methodology which was used for the collection of data. It also tells us about the research question, participants, aims and objectives of the study. It was a quantitative approach in which experimental design was used to collect data from 54 students at a government university in Karachi. The participants were from 3 different departments of BS 1<sup>st</sup>.semester of a government university. The students were regular and hardworking. 18 students were selected from each class by using simple random sampling. In the classroom, students were given 3 different activities to work out. The first activity consisted of some verbs and adjectives for which the correct prepositions had to be found. The second activity consisted of some phrases; the frequency in everyday conversation of the phrases had to be found. The third activity consisted of collocation i.e. there were some words similar in meanings for which the correct usage had to be found.



### 3.2 Participants

For this study, I selected 54 students from 3 departments offering BS program at a government university in Karachi. There were 18 participants each from 3 departments. The students were studying at BS 1<sup>st</sup> semester. Though there were approximately more than 150 students, but 15 to 20 students were coming regularly in each class. I, therefore, selected 54 students who were also very enthusiastic, energetic, and studious. I obtained the consent of these 54 students after mentioning my purpose of study. There were 30 female and 24 male students.

**Table 2. Participants' Information**

<b>Age</b>	18-22	18-22	18-22
<b>Semester</b>	BS 1 <sup>st</sup>	BS 1 <sup>st</sup>	BS 1 <sup>st</sup>
<b>Departments</b>	D1 (Zoology)	D2 (CS)	D3 (Pharm D)
<b>T No. of students</b>	18	18	18
<b>Gender (males)</b>	05	13	06
<b>Gender (females)</b>	13	05	12

### 3.3. Sampling

In this research, simple random sampling was used. 54 participants were selected from 3 different classes of BS 1st semester at a government university. According to (McCombes, 2023), simple random sampling is a form of probability sampling in which a subset of participants from the population is chosen randomly. Each person in the population has an equal probability of being chosen. Then the data is collected from as big a population as possible. Since I had to know the efficacy of the activities designed with the help of corpus linguistics, I had no problem choosing the participants.

### 3.4 Research question

1. How effective is it to use corpus linguistics as a teaching tool to teach English language to BS students at a government university?

Apply, communicate, write, divide, obligatory, speak, move, die, put, throw, call, shout, responsible, send, dispel, enhance, devise, fabricate, pretty, handsome

**Figure 1. Verbs given to all 3 groups.**

Look, fix, way, wide, scold, perceive, cite, dump, hurry, mute, cling, shout, freedom, liberty

**Figure 2. Words given to all 3 groups.**

Beautiful/pretty, high/tall, small/short, large/big, stroll/walk, read/recite, good-looking/handsome,

**Figure 3. Synonyms given to all 3 groups.**

### 3.6 Procedure

For this purpose, I compiled a list of a large number of corpora containing verbs and phrases with their usage and gave them to the students. I also recommended different dictionaries including Oxford Advanced Learner's Dictionary, Longman Advanced Learner's Dictionary, Collins Dictionary and Cambridge Advanced Learner's Dictionary which are corpus-based containing a great collection of examples so that they could perform their activities in an easy and proper way. These dictionaries were online dictionaries which students used with the help of their mobile phone Internet.

## 4. Findings & Analysis

The findings show that all three groups did their activities by using corpus linguistics. In the table below, a search on these words with different prepositions from BNC and corpus-based dictionaries reveal how the meaning of a word changes only due to the use of different prepositions such as 'apply for' has a different meaning from 'apply to', 'die of' has a different meaning from 'die for'. Students also learnt about the frequency of these words in British National Corpus. In this way, students understood the meanings of one word with different uses of prepositions and they also increased their vocabulary. 42 students (78%) did the first activity very well. They found the

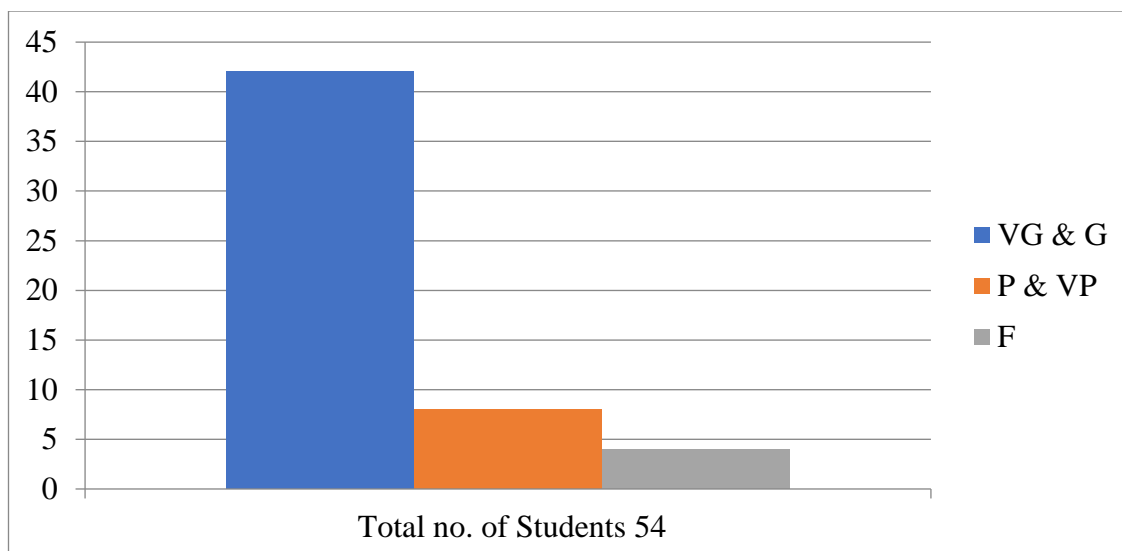
prepositions to a great extent which are used with the verbs given to them. 8 students (14%) could not do the activity as was expected from them and 4 students (7%) did neither good nor bad.

**Table 3. Use of prepositions with the words that change the meaning: Word frequency is also shown from BNC.**

Words	Frequency	Preposition used in bold
apply for	954	I am <b>applying for</b> the post of lecturer.
apply to	2357	Do the same rules <b>apply to</b> part-time workers?
communicate to	1497	She will <b>communicate</b> it <b>to</b> you.
communicate with		People use more than words when they <b>communicate with</b> each other.
write to	1650	I have <b>written to</b> my uncle.
write about	399	O'Brien often <b>writes about</b> her native Ireland.
write for	144	Maureen Dowd <b>writes for</b> 'The New York Times'.
write down	474	He <b>wrote down</b> everything she said.
divide into	62	Here, the river <b>divides into</b> three channels.
divide	03	The money will be <b>divided</b> equally <b>among</b> the charities.
among	87	If you <b>divide 21 by</b> 3, you get 7.
divide by		
obligatory for	21	It is <b>obligatory for</b> companies to provide details.
speak to	2001	I <b>spoke to</b> her last Wednesday.
speak of	837	It was the first time she had ever <b>spoken of</b> marriage.
speak at	129	John <b>spoke at</b> the teachers' annual convention.
die of	241	The animals <b>died of</b> starvation in the snow.
die from	98	Many patients are <b>dying from</b> cancer.
die for	125	He is <b>dying for</b> a cup of tea.
die on	59	The mower just <b>died on</b> me.
die in	373	During the night she <b>died in</b> sleep.
throw at	46	Someone <b>threw</b> a stone <b>at</b> the car. He <b>threw</b> his shirt <b>to</b> someone in
throw to	11	the crowd.
shout at	143	I wish you would stop <b>shouting at</b> the children.
shout for	30	We could hear them <b>shouting for</b> help.
shout in	16	She <b>shouted in</b> pain.
responsible for	6654	Police believe that the same man is <b>responsible for</b> three other murders in the area.
send to	178	I will <b>send</b> a copy <b>to</b> you.
send by	07	Monday is the last day to <b>send</b> cards <b>by</b> post.

**Table 4. Students' performance percentage-wise**

Students' Performance in activity 1				
Total students: 54				
Very good	Good	Fair	Poor	Very poor
24	16	04	07	03
44%	30%	7%	13%	6%
Total: 74%		7%	Total: 19%	



**Figure 4. Students' Performance**

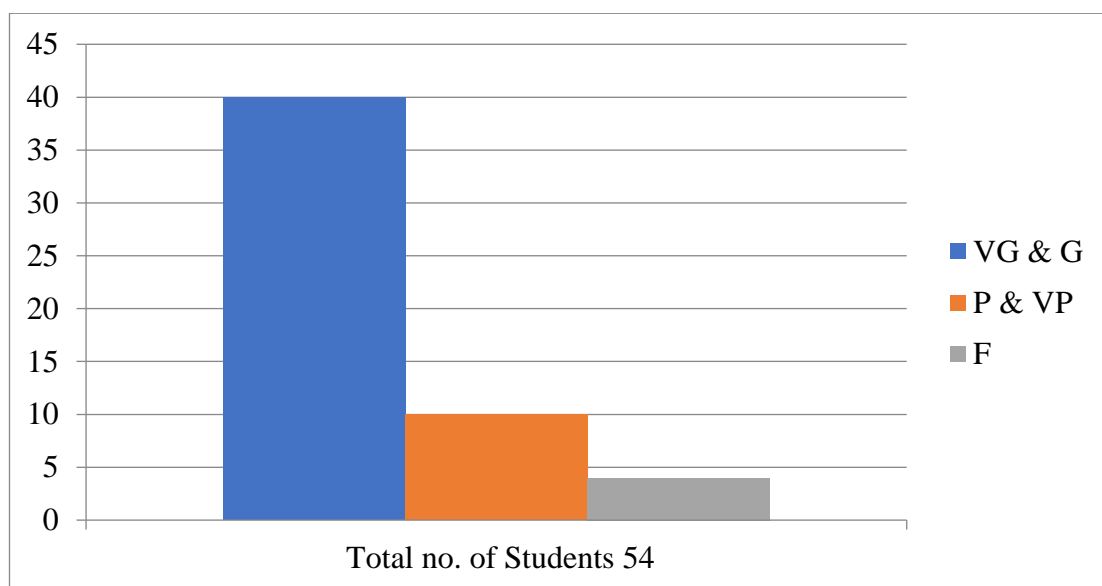
Though the following activity took much time and for this activity either smart phone or laptop along with internet was required, students enjoyed it a lot because they were able to find some very useful information about the number of uses of certain words. 40 students (74%) did the second activity well that is about the frequency of words from **British National Corpus (BNC)** [100 million words of text texts from a wide range of genres, such as from spoken, fiction, magazines, newspapers, and academic] along with its common use from Collins Dictionary. 10 students (18%) could not do it well whereas 4 students (7%) were neither good nor bad.

**Table 5. Showing how commonly a word is used in Collins COBUILD**

Words	W. Frequency per 100 M	Years	Words commonly used in corpus-based Collins dictionary
<b>Look</b>	51549	1980 - 1990	Used extremely commonly. There are 1000 most commonly used words in the Collins COBUILD and the word ' <b>look</b> ' is one of them.
<b>Fix</b>	1392	1980 – 1990	Used commonly. The word ' <b>Fix</b> ' is one of the 10000 most commonly used words in the Collins COBUILD.
<b>Scold</b>	57	1980 – 1990	Used occasionally. The word ' <b>scold</b> ' is one of the 30000 most commonly used words in the Collins COBUILD
<b>Perceive</b>	889	1980 – 1990	Used very commonly. The word ' <b>perceive</b> ' is one of the 4000 most commonly used words in the Collins COBUILD.
<b>Cling</b>	390	1980 – 1990	Used commonly. The word ' <b>cling</b> ' is one of the 10000 most commonly used words in the Collins COBUILD.
<b>Mute</b>	242	1980 – 1990	Used commonly. The word ' <b>mute</b> ' is one of the 10000 most commonly used words in the Collins COBUILD.
<b>Freedom</b>	5832	1980 – 1990	Used very commonly. The word ' <b>freedom</b> ' is one of the 4000 most commonly used words in the Collins COBUILD.
<b>Liberty</b>	1359	1980 – 1990	Used very commonly. The word ' <b>liberty</b> ' is one of the 4000 most commonly used words in the Collins COBUILD.
<b>Wide</b>	11775	1980 – 1990	Used extremely commonly. The word ' <b>wide</b> ' is one of the 1000 most commonly used words in the Collins COBUILD.

**Table 6. Showing students' performance percentage-wise**

Students' Performance in activity 2 Total students: 54				
Very good	Good	Fair	Poor	Very poor
19	21	04	06	04
35%	39%	7%	11%	8%
Total: 74%		7%	Total: 19%	



**Figure 5 Students' Performance**

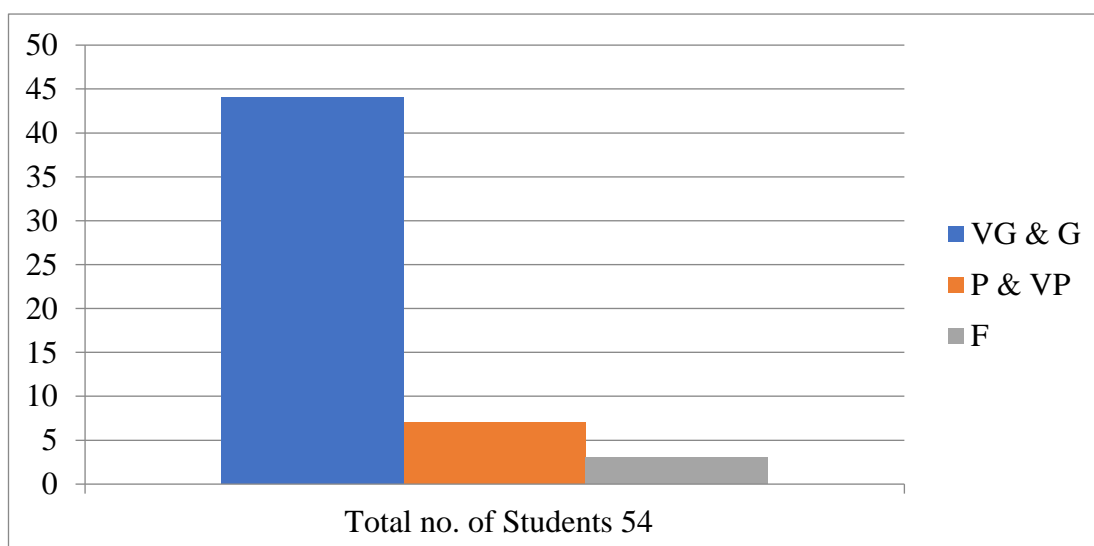
In the third activity, the students learnt about collocation which is a very important part of any language. Without knowing about collocation, no one can speak English as a second language like a native speaker. 44 students (81%) enjoyed the activity and did it well and only 7 students (13%) could not do it as they should have done it, whereas 3 students (6%) were neither good nor bad.

**Table 7. Correct usage of some of the words similar in meanings**

	Words	Collocation: Correct usage
01	Beautiful/pretty	John is a handsome young man. She looks pretty with her hair down.
02	High/tall	Jasmine is getting very tall. This is the highest mountain in Japan.
03	Small/short	The sweater comes in three sizes – small, medium and large. Winter is coming and the days are getting shorter.
04	Large/big	The T-shirt comes in Small, Medium and Large. There has been a big increase in crime
05	Read/recite	He recited a poem he learnt few days ago. She is reading a letter that came yesterday.

**Table 8. Showing students’ performance percentage-wise**

Students’ Performance in activity 3				
Total students: 54				
Very good	Good	Fair	Poor	Very poor
26	18	03	04	03
48%	33%	6%	7%	6%
Total: 81%		6%	Total: 13%	



**Figure 6 Students’ Performance**

## 5. Discussion

In the first activity, students were given some verbs and adjectives to find out the prepositions which are used with them. 78 % of students found them successfully with the help of the corpora. In this way, they were able to know multiple meanings of a word with different prepositions and they were also able to know the use of these words with appropriate prepositions. 74% of students did the second activity successfully and they were able to know about the frequent use of the phrases. This

activity helped them know the worth of those phrases. 81% of students did the third activity successfully. With the help of this activity, they were able to learn how to use words correctly, especially words with similar meanings.

## **6. Conclusion**

From the above discussion, it has been found that corpus linguistics is a very useful tool for teaching English language in the classroom. It can play a significant role in teaching English language both spoken and written. Teachers must use corpus data in the classroom so that they can give different interesting activities to their students for the improvement of even their vocabulary and collocation along with their writing and spoken skills. It is indeed a very effective tool for teaching English at the upper intermediate level.

### **6.1 Recommendations**

- Corpus linguistics as a tool may be used in upper-intermediate and BS level.
- For using this tool computers, internet and certain corpus-based dictionaries are necessary. Therefore, these things should be made available, or a teacher should compile his/her own lists from these resources.
- Corpus linguistics can be used to identify the correct usage of any word.
- It can also be used to check or see the use of structures and the frequency of words – whether the word is common or uncommon.

### **6.2 Limitations**

- Corpus linguistics should not be used as a tool for teaching English language at elementary or intermediate level.
- Corpus linguistics as a tool should be avoided at the universities or educational institutes in which modern technology such as computers, internet, etc. are not available.
- Corpus linguistics as a tool is not very effective to improve speaking skills.
- Teachers with no computer skills or little skills cannot benefit from corpus linguistics.

## References

- Aijmer, K. (2002). *English Discourse Particles: Evidence from a Corpus*. Amsterdam: Benjamins.
- Anwar, M. (2018). Practices of English Language Teaching at Government Institutes. *Corpus approaches to grammaticalization in English, 13*, 69-79.
- Biber D, Johansson S, Leech G, Conrad S & Finegan E (1999). *Longman grammar of Spoken and written English*. London: Longman.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating* Carter R (2004).
- Collins COBUILD English Language Dictionary (editor in chief, J. Sinclair). 1987. *Compass, 3(5)*, 1225–1241.
- Gries, S. T. (2009). *What is Corpus Linguistics? Language and Linguistics* <https://www.oxfordlearnersdictionaries.com/>
- Hundt, M. (2004). The passival and the progressive passive A case study of layering in Hunston, S. (2022). *Corpora in applied linguistics*. Cambridge University Press.
- Hunston, S., & Thompson, G. (2006). *System and corpus: Exploring connections*.
- Hyland, K. (2000). Disciplinary discourses: Social interactions in academic writing.
- Johns, T. (1991). *Should you be persuaded: Two samples of data-driven learning language structure and use*. Cambridge University Press.
- Leech, G., Rayson, P. and Wilson, A. (2001) *Word Frequencies in Written and* London: Collins. Longman.
- Matthiessen, C. M. I. M. (2006). *Frequency profiles of some basic grammatical systems: use and language teaching*. Cambridge University Press.
- O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language. Spoken English: Based on the British National Corpus*. Longman, London.
- Sinclair, J. (Ed.) (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Thornbury, S. (2002). *How to teach vocabulary*. Essex: Pearson Education Limited. University Press.