



Breadth of EFL Vocabulary Building at Secondary Level in Pakistan: A Corpus-based Perspective

Vol. IV, No. IV (Fall 2019) | Page: 105 – 112 | DOI: 10.31703/grr.2019(IV-IV).12

p- ISSN: 2616-955X | e-ISSN: 2663-7030 | ISSN-L: 2616-955X

Shamim Akhter*

Behzad Anwar†

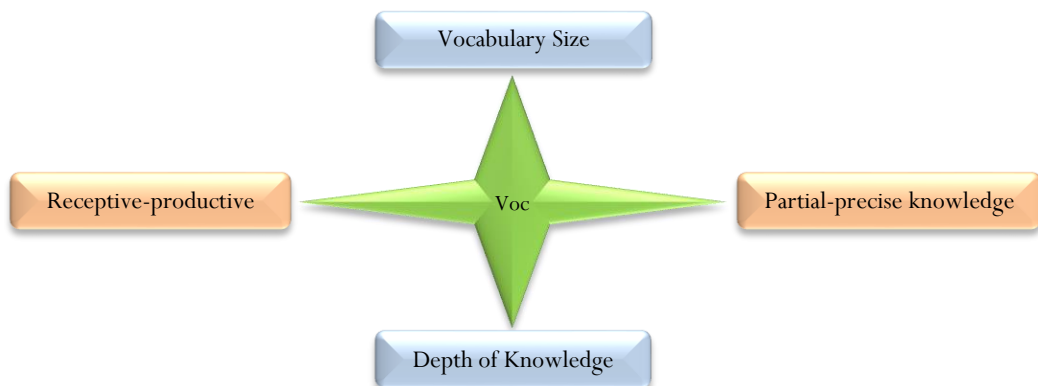
Abrar Hussain Qureshi‡

Abstract *Vocabulary learning is one of the challenging tasks in a foreign language classroom. In order to make this process motivating and friendly, one of the useful techniques is to unburden the teacher and to provide the learners with maximum autonomy. Graves (2007) asserts that EFL learners should have the vocabulary size of approximately 3,000 to 6,000 English words. The undertaken study is an attempt in this regard to determine the breadth of the EFL vocabulary at secondary level. A corpus of Pakistani English has been compiled in this regard. In order to process data, Wordsmith Tools, a statistical method for analyzing corpus data, has been used to retrieve the frequency list. Afterwards, the results have been analysed with SPSS. The findings establish the fact that the determining the breadth of vocabulary with the help of corpus is very significant in developing EFL vocabulary at secondary level.*

Key Words: Corpus; frequency list; EFL; learning strategies; vocabulary.

Introduction

Vocabulary plays a core part in language learning. Vocabulary learning is most valuable and integral part of whole language learning process (Mukoroli, 2010). It is an important factor in academic language (Ajayi, 2005). Moore and Calvert (2000) communicate that vocabulary can be learned in a collection of various courses consolidating with visual targets and strategies. Vocabulary Learning Strategies which are developed by Schmitt (1997) show five type of strategies used in vocabulary which are named as Determination Strategy (DET), Cognitive Strategy (COG), Memory Strategy (MEM), Metacognitive Strategy (MET) and Social Strategy; complicated strategies like DET and COG are used by students mostly whereas, after the advent of computers and its use in different domains including EFL has changed the whole scene altogether. Qian (2002) has expanded its dimensions and divided into four following connected aspects:



*Researcher, Department of English, Institute of Southern Punjab, Punjab, Pakistan. Email: misschudry96@gmail.com.

† Assistant Professor, Department of English, University of Gujrat, Punjab, Pakistan.

‡ Associate Professor, Department of English, Institute of Southern Punjab, Punjab, Pakistan.

The use of corpus has made a significant impact in shaping the modern trends in English as a foreign language. These new techniques are very helpful particularly in vocabulary building and exploring *new trends in vocabulary teaching* (Hoey, 2000). In this regard, corpus frequency list strategies are proving very useful to enhance vocabulary skills of the learners.

Qian (2002) classifies memory and cognitive strategies in handling vocabulary of English language. He states that verbal, written repetition, taking notes, using word lists, flash cards, vocabulary sections in a text book, audio visual aids, placing vocabulary on physical objects and keeping vocabulary notebooks can be very useful in the comprehension of English as a foreign language. He also stresses on cognitive strategies to be the crucial factor in learning EFL vocabulary.

Read (2000) makes a distinction between those who structure their own learning of EFL vocabulary and those who do not make any serious self-efforts apart from the classroom instruction and do not go through self-practice. The unstructured EFL learners only depend on classroom instructions and make no independent efforts. While structured learners make their own learning by creating innovations in learning strategies, record and review new vocabulary items consistently and even use that new EFL vocabulary outside the classroom.

Kumaravadivelu (2003) states that key word method to learn EFL vocabulary is very potential to face the challenges of foreign language vocabulary. However, the results of his study were not motivating since a post test, arranged only a week after, indicated that EFL learners forgot the meanings and use of the new words up to 52 percent who were instructed only with the help of key words method while the EFL learners lost up to 76 percent who were instructed with combine context condition method. However, the comparative impact of word list method led to further experiments in EFL learning strategies. He also believes that L2 vocabulary depth is also a key indicator in this regard.

Vocabulary is of paramount importance in language learning and communication, though it has not been highlighted properly. Especially, the nature of the words and their role in building the overall meaning of the utterance has not been emphasized in EFL context. Since the last three decades, the significance of potential vocabulary has been brought to light by language researchers and lexicologists. This trend of vocabulary building in a foreign language has emerged due to its established role in improving the overall skills of the foreign language, especially in achieving fluency and accuracy. Traditionally, the issue of vocabulary was handled with the help of memorization and laborious drills of new words. There is a considerable shift from linguistics to corpus linguistics. The later has revolutionized the art and craft of foreign language learning. As the EFL learner lacks the contextual meanings of the most of the words, it creates pressures even for the language teachers. The use of corpus data helps the learners to develop vocabulary with their true semantic dimensions simultaneously. Moreover, the advantage of corpus data has also merits as it has been collected from all natural contexts.

Statement of the Problem

Conventional lexicons frequently give hazy, constrained, and manufactured cases for every definition, which is lacking for completely understanding a new word. Also, understudies cannot manufacture vocabulary in regular process in light of the fact that in our organizations, primary language is utilized for direction. Even for the deliberate efforts for the construction of vocabulary building, they take help from mother tongue. One of the many problems in this regard, is the bulk of the vocabulary, the students are supposed to learn at each level. Similarly, a number of techniques may be applied through instructing the learners to apply various vocabulary learning strategies as efficiently as possible. But in spite of all these user friendly techniques of developing vocabulary, an issue keeps on pressurizing the learners as well as the teacher that what should be the healthy breadth of vocabulary at a certain level as EFL learners at various stages of language learning will need different types of vocabulary in different number. This study is an attempt to solve the vocabulary problem at secondary level with the help of corpus and it also determines the breadth of the vocabulary for the under investigation level.

Research Questions

1. What type of difficulties are faced by learners while learning EFL Vocabulary?
2. How much effective is the corpus of Pakistani English in learning English language vocabulary?

3. What will be the suitable breadth of EFL vocabulary at secondary level for EFL learners?

Research Design

With the invention of computers, an empirical approach to linguistic analysis has emerged. It represents natural samples of target language stores as an electronic database. Such studies primarily rely on computer software for the quantitative linguistic analysis. Though Firth (1957) is said to be the precursor in the art and craft of corpus linguistics yet the true innovations were brought by Sinclair (1992) with the inclusion of the computers and software. Firthian approach was not corpus based while Sinclair approach was purely corpus based that produced more accurate results while handling a large data. Sometimes, the quantitative analysis is supplemented with qualitative linguistic analysis to make the results more critical.

Corpus linguistics has been intensely debated as a method for linguistic analysis. Some believe it to be only a source of accurate linguistic analysis that provides a pile of information about various dimensions of target language while according to some linguists, it merely provides a methodology for linguistic analysis or it is a bag of resource tools or techniques that help the researchers in understanding the phenomenon of language. Similarly, there is a significant shift in vocabulary research since the last few decades with the emergence of corpus linguistics. Recent studies on vocabulary have proved the utility of corpus in EFL context. It has proved to be a great learning tool.

Quantitative approach has been carried out for the undertaken study. The study has been conducted with a number of 60 learners of secondary level from Government institutions for the academic year 2019. Participants have been divided into two groups. In Group 'A' which has been exposed to frequency list of EFL vocabulary (named as experimental group), the participants have been motivated to use the list of most frequent words for vocabulary building. Other Group 'B' (the control group) is taught in the conventional way of learning vocabulary, the 30 participants have been taught the vocabulary items in class, by means of reading, dictionary using, translating, comparing and following traditional patterns. The same pre- and post-test based on vocabulary recognition was given to both groups. To examine the impact of frequency list instructional methodological program on students' vocabulary achievements, the research has been designed, experimenting through a pre-test- post-test evolution and control group framework. The test comprises of multiple-choice questions on vocabulary items. The participants' previous knowledge is assessed by the pre-test taken by both groups (A and B) while similar post-test was administered at the end of the study session to examine the students' achievement on the concerned topic, i.e. learning of the new lexical items. The post-test is used to assess the difference between both methods, instruction by frequency list and teaching through traditional instruction on students' achievement.

EFL Vocabulary and Corpus

Linguistics in Corpus is a relatively new way of studying linguistics as it developed rapidly since 1980s with the development of computer science, which offers dynamic technical support. Its advantages are almost incomparable in delivering huge amounts of real, efficient and powerful ability to research and study. Corpus linguistics and its use in teaching and learning attracted a lot of EFL researchers. The importance of corpus linguistics has also been widely accepted. A lot of linguistic studies on how linguistics can facilitate teaching and learning activities of various Language levels, has highlighted the scope of corpus linguistics.

The application of corpus linguistics in teaching can be divided into two aspects: the direct one; taking the relevant knowledge of corpus linguistics; means of developing a linguistic corpus; and applications of linguistic corpus as the teaching materials; and the indirect one: based on both corpus and computer technology, including compiling corpus-based dictionary; editing grammatical reference; textbook; developing multimedia courseware; language learning software, or evaluating or testing tools. Vocabulary is one of the three basic elements of language and is taken as the backbone of the whole language system as pointed out by Sinclair (1992) "If the language regarded as the bone of the tongue, the vocabulary offers vital organs, flesh and body." He further argued that without grammar, many can happen to express in languages, but nothing can be expressed without vocabulary.

Vocabulary is widely applied and its education is playing an increasingly important role in language learning. Linguists turn to the corpus linguistics for the ways of improving the language learning, which does make sense, especially for EFL vocabulary learning. As the learners lack a real context for the application of the language, sometimes, it is difficult for the learners to handle their learning and to acquire the usages and meaning of certain words accurately. The application of corpus linguistics into EFL vocabulary learning can almost perfectly deal with that, for the language data in the corpus are all from natural contexts, which can help the learners to use the word accurately and properly. Corpus was used to conduct research in Europe in the early eighteenth century; the procedure was followed, with a great deal of time and effort. In the nineties, the main application of the corpus was the study of the dictionary and the use of grammar. It started getting eminence in the 1950s but began to recede in the mid-80s.

Corpus refers to the composition of the natural words, containing everything from a few sentences in huge bulk, written or oral for language learning. Recently, the word "Corpus" has been reserved for a collection of letters (or parts of the manuscript) that are stored and stored electronically. Because computer can pack and process many files information, electronic banking is generally larger than the academic literature before used in the study of words. The process is dead but the odds can play a part in the text collection, and it is designed for some purpose. Specific purpose of the decision model choice of texts, and the purpose is rather to store the books themselves because they are internal force. This separates the corpse from a library or electronic archive. This storage is kept there in a way that it can be learned in a linear way and effectively.

Corpus linguistics has not same sense of potential as semantics, syntax, and pragmatics. Its real significance is in its empiricism as it takes into account the real patterns of language. Its analysis is objective and verifiability. It allows the collection of different genres and registers which makes it possible to show the wide repertoire of language. The corpus analysis is democratic in its approach as it provides equal chances to non-native speakers as to native speakers. Lastly, focused results can be obtained for pedagogical purposes.

During EFL learning, learners are more or less exposed to the native language, which is called transposition of the first language. The transfer can be both positive and negative. The negative change affects the traditional comprehension of the accepted language. However, the expression of the same situation is different in two languages. There may be differences in both the situations. Negative changes affect the learners' understanding of the new knowledge of the foreign language and communication. The consequences of negative changes are very high and are often found in EFL vocabulary learning.

Corpus of Pakistani English and the Frequency List

According to McCarthy (1998), corpus data is generally gathered with two approaches. The first approach is the genre approach that does not only rely on a pre-decided idea of a text but also endeavors to create a healthy balance between the context and the language use with repeated patterns. It focuses the population of language users, the context in which language is used and the environment. While the second approach is called demographic approach where the users of a target language are focused and there is a consideration of the span of the time as well during which that language has been used.

The genre approach is generally employed to compile corpus of a foreign language. The samples of corpus are different from one component than the other. Keen observation and intuition are used to decide the appropriateness of the corpus sample. The sample size is basically determined on the basis of two factors. The considerable factor is the availability of the text and the second important factor is the readership of the selected text.

In order to check the effectiveness of linguistic variation, the factor of the readership is considered very significant. In this regard, newspapers have maximum readership. As a result, the major bulk of the compiled corpus has been selected from the newspapers. English occupies an important place in Pakistan and is used in many domains for example administrative, educational, legal, judiciary, etc. There has been made intentional effort to include the data from all these sources to present a true reflection of Pakistani English. Some samples have also been selected from religious text as well. To undertake this research, a small Pakistani English Corpus (PEC) of the following newspapers for one month period (February 2018) has been compiled:

Table 1.

S. No	Newspapers	Token	Types
1	Dawn	2349698	57568
2	Daily Times	1628055	42831
3	The News	1995506	41752
4	Business Recorder	1283258	33251
	Total	7256517	85194

In order to operate the corpus, a software, Wordsmith Tools, has been used. “Wordsmith Tools” was launched in 1996. Through wordsmith tools, most frequent words have been retrieved. The proposed frequency list will help the ESL learners to learn only the most required vocabulary items.

Data Analysis

This section is related to data analysis and the tables show the trend of values to the respondents towards the statement in the questions. The results have been compared by using SPSS and T paired test is applied to compare the difference of performance between both groups (controlled and experimental) in the form of pre-test and post-test.

According to Coniam (1997), planning and the motivation behind every task play a valuable part which helps the doer to achieve the desired results in a better and accurate way. To achieve that goal for creating better motivation, planning has a vital role (Granger, 1998). Particularly for academic process where more success and achievement rely on the proper and suitable planning, certain means have been chosen. For the present study, the importance of the disciplined planning has been kept in mind. Material based on lexical items of different levels has been developed so that positive change in the students’ marks and learning with the help of incentive tools and positive support and bolstering could be measured. It has been designed to support the progress of the students and the ‘Like’ function has been used as students express their performance and capability well when they are given some prizes and awards. The lessons have been developed with a lot of colorful manner and other interesting activities. These activities are applied on the experimental group and objective behind is to show healthy response and remarkable progress in their lessons. Data has been collected with the help of tests (pre-post). The significance difference of means of controlled and experimental group is tested at 0.05 values. SPSS has been used and data is analyzed by using paired t-test. Certain measures have been calculated. The Tables below represents the results:

Table 2. Paired Statistics of Pre-test Post-test of Control Group

	Pre-test of Control Group	Post-test of Control Group
Valid	30	30
Missing	0	0
Mean	11.73	12.67
Median	12.0	12.0
Std. Deviation	1.778	1.151
Variance	3.66	1.54
Range	6	4
Minimum	8	12
Maximum	17	15

As Table 2 shows about the control group, the mean of pretest is 11.73 and mean of post-test scores (12.67) is somehow greater than the mean of pretest scores but not very significant.

Table 3. Paired Samples Statistics

Pair 1	Mean	N	St. Deviation	Std. Error Mean
Post-test of Control group	13.08	25	1.151	.257
Pre-test of Control group	12.84	25	1.778	.394

Table 3 shows the correlation between the mean of pretest and post test scores of the control group which exposes some improvement in the scores of but it is not of a greater value. The standard deviation of pretest is 1.778 and of posttest it is measured 1.151.

Table 4. Paired Samples Test

	Paired Differences					T	Df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	91% Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 Post-test control group-pre-test control group	.240	1.480	.295	-.371	.810	.811	24	.431

In Table 4, through Paired Statistics of t-test, the difference of pretest posttest score mean is measured and the t value (0.810) is checked which shows only a little difference and significance 2 tailed is 0.431.

Table 5. Paired Statistics of Pre-test, Post-test of Experimental Group

	Pre-test of Experimental Group	Post-test of Experimental Group
Valid	30	30
Missing	0	0
Mean	13.96	17.6
Median	14.0	18.0
Std. Deviation	1.908	1.431
Variance	4.04	1.96
Range	7	5
Minimum	10	15
Maximum	16	19

Table 5 relates the mean of pretest-post test scores of the “Experimental group” which exposes the significant improvement in the posttest scores of Experimental group. The median improved by 14 to 18 (median score of both Experimental and Control group was same in the pretest).

Table 6. Paired Samples Statistics

Pair 1	Mean	N	St. Deviation	Std. Error Mean
Post-test of Experimental Group	17.1	25	1.232	.252
Pre-test of Experimental Group	13.21	25	1.708	.394

Table 6 depicts about the Experimental group; the mean of pretest is 13.21 and mean of posttest scores (17.1) is greater than the mean of pretest scores, from these measuring figures can conclude that teaching vocabulary through Facebook has caused the improvement in relevant group.

Conclusion

Vocabulary fitness as of late has been distinguished to be the most critical indicator to general dialect capacity; however, it is additionally recognized by most learners to be one of the greatest difficulties of dialect learning. Luckily, with the coming of corpus innovation, another perspective of language learning has risen endeavors to coordinate PCs as apparatuses in language classrooms.

From a global point of view, the number of explicit vocabulary learning activities is quite high because it achieves 1/3 of all actions in the textbook. This it should be added that the percentage of explicit and accidental vocabulary activities are uniformly distributed along the units, as shown by the distribution of activities per unit. We must therefore conclude that in terms of the amount of work devoted when it comes to learning vocabulary, corpus strategies are well on the learners' way to achieving their goals of learning vocabulary.

While content is taken as one of the most important parts of a successful communication, the importance of familiarity in foreign language strategies has resulted in greater interference by speakers. Corpus linguistics, based on the application of technology, shows its advantage in the EFL vocabulary because of the ease of exploration and the natural content that it is given to the target word. As it has been identified above, the quality of the corpus language in the EFL guidelines is grounded. In order to eliminate its disadvantage and make good use of it, a corpus should provide a context in which a word is present while not yet representative of its angles of all meanings and relationships of the text. Moreover anything provided by the corpus can only be used as evidence, rather than information.

This exploration establishes that language corpora can upgrade the nature of vocabulary learning in second or outside language classrooms. By showing advantages of language corpora to learners, it is trusted that this examination can be useful to learners who are attempting to look for a productive method for learning vocabulary. The aftereffects of corpus based review will ideally guarantee the understudy's needs to create vocabulary at intermediate level. This is the primary investigation of its sort and fills a critical crevice in research on vocabulary techniques at intermediate level.

Recommendations

The undertaken research is far from being thorough but, it can be taken as a point of departure. The undertaken work may be expanded in different directions. It has still many unexplored dimensions. The above list has been retrieved on the basis of frequency only. It can be further analyzed on phonetic, semantic or syntactic level. The work motivates the future researchers to explore further dimensions of this vocabulary list that will help the learners and the teachers in an EFL classroom in future.

References

- Academic Reading Performance: an assessment perspective. *Language Learning* 52(3), 55-83. NY: Teachers College.
- Ajayi, L. J. (2005). Vocabulary Activities, and English Language Learners in a Second Grade
- Coniam, D. (1997). A preliminary inquiry into using corpus word frequency data in the automatic generation of English language cloze tests. *Calico Journal*. 15-33.
- Firth, John Rupert. (1957). *Papers in linguistics 1934-51*. London: Oxford University Press.
- Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. (pp. 3-18). In S. Granger (Ed.) *Learner English on Computer*. London: Longman.
- Graves, M. (2007). Long-term vocabulary instruction. In B. Taylor & J. Ysseldyke (Eds.),
- Gu, P. Y. (2003a). Vocabulary learning in second language: person, task, context and strategies" *TESL-EG*, 7(2), 1-25.
- Hoey, M. (2000). A world beyond collocation: New perspectives on vocabulary teaching. *Teaching collocations*, 224-245.
- Kumaravadivelu, B. (2003). *Post method pedagogy*. Oxford: OUP.
- Lewis (ed.) *Teaching Collocations*, Hove: Language Teaching Publications pp. 224-245.
- Mixed Classroom. *Journal of Instructional Psychology*, 32(3), 180-195.
- Moore, M., & Calvert, S. (2000). Brief report: Vocabulary acquisition for children with autism: Teacher or computer instruction. *Journal of Autism and Developmental Disorders*, 30(4).
- Mukoroli, J. (2011). Effective Vocabulary Teaching Strategies for the English for Academic
- Neuman, S. B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in pre-K. *The reading teacher*, 62(5), 384-392.
- Qian, D.D. (2002). Investigating the Relationship between Vocabulary Knowledge and
- Read, J. (2000). Assessing vocabulary. Cambridge University Press. *Reading Teacher*, 62(5), 384–392.
- Schmitt, N. (1997). Vocabulary Learning Strategies. In D. N. Schmitt, & M. McCarthy (Eds.),
- Sinclair, J. M. (1992). The automatic analysis of corpora John M Sinclair. Mouton De Gruyter.
- Sinclair, J. M. (1992). The automatic analysis of corpora. *Directions in Corpus Linguistics*, Mouton de Gruyter, Berlin and New York, NY, 379-397.