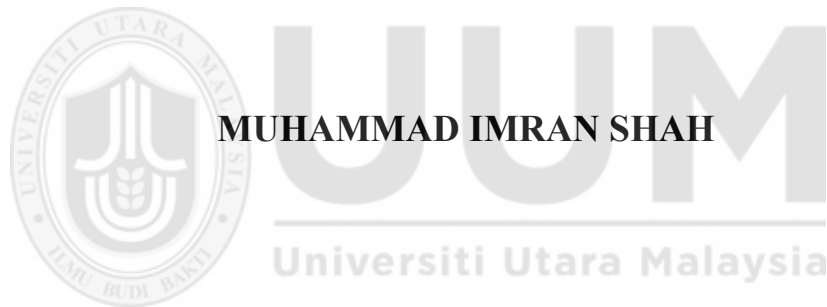


The copyright © of this thesis belongs to its rightful author and/or other copyright owner. Copies can be accessed and downloaded for non-commercial or learning purposes without any charge and permission. The thesis cannot be reproduced or quoted as a whole without the permission from its rightful owner. No alteration or changes in format is allowed without permission from its rightful owner.



**EXPLORING THE EFFECTIVENESS OF CORPUS-DRIVEN
LEARNING IN TEACHING LEXICAL COLLOCATION TO
PAKISTANI UNDERGRADUATES**



MUHAMMAD IMRAN SHAH

**DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2024**

Certification of Thesis



Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI (Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

MUHAMMAD IMRAN SHAH

calon untuk Ijazah
(candidate for the degree of)

PhD

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"EXPLORING THE EFFECTIVENESS OF CORPUS-DRIVEN LEARNING IN TEACHING OF LEXICAL COLLOCATION TO PAKISTANI UNDERGRADUATES"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada: **13 Mac 2023**.

That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: **13 March 2023**.

Pengerusi Viva:
(Chairman for VIVA)

Assoc. Prof. Dr. Hariharan a/l N Krishnasamy

Tandatangan
(Signature)

Pemeriksa Luar:
(External Examiner)

Dato' Paduka Prof. Dr. Supyan Hussin

Tandatangan
(Signature)

Pemeriksa Dalam:
(Internal Examiner)

Assoc. Prof. Dr. Hisham Dzakiria

Tandatangan
(Signature)

Nama Penyalia/Penyelia-penyelia:
(Name of Supervisor/Supervisors)

Assoc. Prof. Dr. Manvender Kaur a/p Sarjit Singh

Tandatangan
(Signature)

Tanakh:

(Date) **13 March 2023**

Permission to Use

In presenting this thesis in fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for the scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:



Dean of Awang Had Salleh Graduate School of Arts and Sciences

UUM College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

Abstrak

Pemerolehan bahasa asing adalah lebih berkemungkinan untuk bertambah jika penumpuan pelajar secara sedar dialihkan kepada ciri-ciri linguistik berdasarkan pendekatan leksikal (Lewies,1990). Pendedahan secara langsung ciri-ciri tertentu daripada korpora seperti kolokasi membantu pelajar bukan asli EFL/ESL dalam pemerolehan bahasa. Kajian penyelidikan ini menerangkan kesan pembelajaran bahasa berasaskan korpus (CDLL) ke atas pelajar EFL Pakistan dan meninjau pendapat pengguna bahasa mengenainya. Reka bentuk penyelidikan eksperimen berdasarkan pendekatan kaedah campuran (Cresswell, 2014) telah dijalankan terhadap 100 orang pelajar, terdiri daripada 50 orang pelajar untuk tujuan eksperimen dan 50 orang lagi untuk kumpulan kawalan dari Jabatan Linguistik Gunaan GC University Faisalabad Pakistan. Dalam ujian pra, pengetahuan tentang kolokasi leksikal pelajar telah dinilai dan data di kumpul secara kuantitatif untuk mengukur keputusan kedua-dua kumpulan untuk menguji nilai-p untuk skor-t. Dalam kalangan sampel kajian, 50 orang pelajar kumpulan eksperimen telah menjalani pembelajaran eksplisit dengan kompilasi *adhoc corpora editorial* akhbar Pakistan dan British yang terkemuka untuk mencari corak leksikal tertentu di bawah seliaan guru dan membandingkan kolokasi leksikal yang diekstrak dalam kedua-dua korpora. Dalam kes ini, pelajar menjadi penyelidik dan bahasa menjadi data (Boulton, 2018); pelajar belajar tentang perkara yang mereka hadiri dengan sengaja. Manakala, kumpulan kawalan seramai 50 orang pelajar telah menjalani pendekatan tradisional. Selepas menjalani eksperimen selama satu semester (16 minggu), keputusan ujian pasca mendedahkan bahawa kaedah CDLL memberi kesan positif dan meningkatkan pengetahuan kolokasi leksikal dalam kalangan kumpulan eksperimen berbanding kumpulan kawalan kerana nilai-p data menurun. Temu bual separa berstruktur telah dijalankan dalam kalangan peserta kumpulan eksperimen dan ahli fakulti jabatan untuk mendapatkan pendapat mereka tentang keberkesanan pembelajaran bahasa berasaskan korpus (CDLL). Kajian ini mempunyai implikasi pedagogi untuk ahli akademik dan penyelidik mengenai kepentingan CDLL dalam reka bentuk kursus Kaedah dan Pendekatan ELT.

Kata kunci: Korpora, Pembelajaran berasaskan data, Pembelajaran induktif, Kolokasi leksikal, Pengetahuan prosedural.

Abstract

Foreign language acquisition is more likely to increase if the attention of the learners is consciously drawn to linguistic features. The direct exposure of certain features from corpora like collocation helps EFL/ESL non-native learners in language acquisition. The present research study explains the effect of corpus-driven language learning (CDLL) on Pakistani EFL undergraduates and explores the opinion of language users about it. The experimental research design following mixed-method approach has been applied to 100 undergraduates, comprising 50 students for experimental and 50 for the control group from the department of Applied Linguistics GC University Faisalabad Pakistan. In a pre-test the knowledge of lexical collocations of learners were evaluated; data was collected quantitatively to measure the results of both groups to test the p-value from t-score. Then, among the sample of the study, the experimental group were undergone explicit learning with the compilation of adhoc corpora of editorials of Pakistani and British newspapers to look for certain lexical patterns under the supervision of the teacher and compared the lexical collocations they extracted in both corpora. In this case, learners became researchers and language became data (Boulton, 2018). Whereas, the control group was treated with traditional approaches in teaching English collocations. After experimental treatment of one semester (16 weeks), the results of the post-test revealed that CDLL method affected positively and enhanced the knowledge of lexical collocations more among the experimental group than the control group as the p-value of the data decreased. Then a semi-structured interview was conducted among the participants of the experimental group and the faculty members of the department to seek their opinion about the effectiveness of corpus-driven language learning (CDLL). The study has strong pedagogical implications for academicians and researchers regarding the importance of CDLL in the course design of ELT Methods and Approaches.

Keywords: Corpora, Data-driven learning, Inductive learning, Lexical-collocations, Procedural-knowledge.

Acknowledgement

"Allah sufficeth me: there is no god but He: On Him is my trust,- He the Lord of the Throne (of Glory) Supreme." I am adequately thankful to Almighty Allah, the All-knower for enlightening my thoughts and vision and helping me complete this PhD thesis.

I would like to express my heartiest gratitude to my PhD supervisor, Assoc. Professor Dr. Manvender Kaur Sarjit Singh for dedicating her precious time, valuable feedback and comment to illumine what is dark in me.

I am highly indebted to the Universiti Utata Malaysia for providing me all compatible and feasible environment of study so I have completed my research in serenity and comfort.

My love and thanks to all of my family members, students and colleagues, particularly, my mother whose frequent prayers strengthened and supported me inclusively.

My students are my assets, I hereby acknowledge their joyful contribution in my life as they have proved a source of solace and entertainment herein abroad. I herby, acknowledge the loveful contribution of Rohail, my best student, in the compilation of the research work.

I am also thankful to my children, Fahad & Hina who are my tension relievers and especially my elder daughter, Doctor Eisha sehar who is a source of inspiration and motivation for me to be doctor of philosophy. I also appreciate the cooperation of my caring and loving wife in completion of my thesis. My brother Muhammad Zeeshan (engineer) played a very responsible role in my absence. He looked after all the domestic issues and reduced my tension so I calmly carried out my research work.

I will never forget the all-time support of Assoc. Professor Dr Manvender Kaur in Malaysia. I am thankful for her assistance and guidance. She proved a blessing in disguise in the completion of my thesis

Table of Contents

Permission to Use	i
Abstrak	ii
Abstract	iii
Acknowledgement	iv
Table of Contents	v
List of Tables	x
List of Figures	xii
List of Appendices	xiii
CHAPTER ONE INTRODUCTION	1
1.1 Background of the Study	1
1.1.1 Status of English in Pakistan	2
1.1.2 Role of Corpus Driven Instructions in Language Teaching	6
1.1.3 Role of Corpus in Investigating Collocations	7
1.1.4 Preliminary Investigation	10
1.1.5 Implications of Preliminary Investigation	11
1.2 Problem Statement	12
1.3 Research Objectives	16
1.4 Research Questions	17
1.5 Significance of the Study	17
1.6 Scope of the Study	18
1.7 Definitions of Terms	19
1.8 Organization of the Thesis	24
1.9 Summary	26
CHAPTER TWO LITERATURE REVIEW	28
2.1 Introduction	28
2.2 Theoretical Framework	28
2.2.1 Noticing Hypothesis	30
2.2.2 Constructivist Learning	32
2.2.3 Sociocultural Theory	34
2.2.4 Corpus Driven Language Learning	36

2.3 Explanation of English Collocation	36
2.3.1 Major Approaches to Collocation in Linguistics.....	39
2.3.1.1 Firthian Approach	39
2.3.1.2 Semantic Approach of Collocation.....	40
2.3.2 Level of Co-occurrence of Restriction	44
2.3.2.1 Sinclair's Model.....	45
2.3.2.2 Allerton's Model	46
2.3.3 Approaches to Collocational Knowledge of EFL/ESL Learners	48
2.3.3.1 Major Approaches to the Analysis of Learner English.....	48
2.3.3.2 Corpus-based Approach & Data-Driven Learning (DDL)	50
2.3.4 Research Studies on Collocations among EFL/ESL Learners	53
2.3.5 Corpus Studies of Collocations in Learners' English Writing	54
2.3.6 Statistical Formulas in Corpus Linguistics.....	57
2.4 Literature Reviewed in Practice Perspectives of the Research Work	58
2.4.1 Corpus-based Studies on Collocations in Pakistani Learners' English.....	63
2.4.2 Previous Studies on Data-Driven Learning.....	65
2.5 Conceptual Framework	68
2.6 Summary	70
CHAPTER THREE RESEARCH METHODOLOGY.....	72
3.1 Introduction	72
3.2 Philosophy of the Research	72
3.3 Research Method.....	74
3.4 Population.....	76
3.4.1 Sampling.....	76
3.4.2 Participants of the Research Study	77
3.4.4 Comparison between both Groups of Research Sample	78
3.4.4.1 Language Proficiency and Exposure to the English language of the Sample.....	78
3.4.5 Awareness of Corpus Driven Language Learning	86
3.5 Description of Data	88
3.5.1 Material for Study.....	88

3.6 Data Collection.....	90
3.6.1 Research Instruments for Quantitative Data.....	91
3.6.1.1 Experimental Treatment	92
3.6.1.2 Statistical Information of Data.....	92
3.6.1.3 Comparability of COPLANE & COBLENE.....	94
3.6.2 Research Instruments for Qualitative Data.....	95
3.7 Tools for Data Processing	96
3.7.1 PoS (parts of speech) Tagger: CLAWS4.....	97
3.7.2 Accuracy Rate of Tagged Data.....	97
3.7.2.1 Why Tagging is Necessary	98
3.7.3 AntConc Tools.....	98
3.7.3.1 n-gram/cluster tool	99
3.7.3.2 Concordance	99
3.7.3.3 Word List	100
3.7.4 Experimentation.....	100
3.7.5 Interview	101
3.7.5.1 Nvivo-12 Software.....	102
3.7.5.2 World Cloud.....	103
3.8 Research Design.....	104
3.8.1 Rationale.....	105
3.8.2 Pre-test	107
3.8.4 Intervention.....	108
3.9 Data Analysis Strategy	122
3.9.1 Identification of Lexical Patterns	122
3.9.2 Thematic Analysis of Interviews.....	124
3.10 Implication from Pilot Study.....	126
3.11 Summary	128
CHAPTER FOUR RESULTS AND DISCUSSIONS	130
4.1 Introduction	130
4.2 Quantitative Data Analysis.....	130
4.2.1 Data Analysis Method of Research Question 1	131

4.2.1.1 Analysis Strategy of the Part-1 of Pre-test.....	133
4.2.1.2 Analysis Strategy of the Part-2 of Pre-test.....	134
4.2.1.3 Analysis Strategy of the Part-3 of the Pre-test.....	138
4.2.2 Analysis Strategy of the Post-test.....	142
4.2.3 Descriptive Statistics	144
4.3 Qualitative Data Analysis.....	146
4.3.1 Analysis Strategy of Research Question 2	147
4.3.2 Analysis Strategy of Research Question 3	155
4.3.4. Thematic Analysis of Interviews	164
4.4 Discussion on the Results.....	164
4.4.1 Declarative Knowledge of the Participants	165
4.4.2 Procedural Knowledge of the Participants	166
4.4.2.1 Pattern-hunting.....	166
4.4.2.2 Pattern-defining.....	166
4.4.3 Learners' Attitude towards CDLL.....	168
4.4.4 Teachers' Attitude towards CDLL	170
4.5 Summary	172
CHAPTER FIVE CONCLUSION AND RECOMMENDATIONS.....	173
5.1 Introduction	173
5.2 Findings for Research Question 1	173
5.3 Findings for Research Question 2	175
5.4 Findings for Research Question 3	176
5.5 Implications of the Research Study.....	177
5.5.1 Authenticity of Material	177
5.5.2 Discovery Learning	178
5.5.3 Noticing Hypothesis	180
5.6 Practical Implications of the Research Study	181
5.6.1 Development of Collocational Knowledge	182
5.6.2 Using Corpora.....	183
5.6.3 Training Learners	183
5.6.4 Preparing CDLL Material.....	184

5.7 Limitations of the Study	186
5.8 Summary	187
REFERENCES.....	188
APPENDICES.....	197



List of Tables

Table 1.1 English Collocation.....	5
Table 1.2 Statistics of Speakers of Other Languages in Pakistan.....	13
Table 1.3 Types of Lexical Collocations.....	19
Table 3.1 Course Outline of 1 st Semester BS English (Literature & Linguistics).....	79
Table 3.2 Course Outline of 2nd Semester BS English (Literature & Linguistics).....	79
Table 3.3 Course Outline of 3 rd Semester BS English (Literature & Linguistics).....	80
Table 3.4 Tabulation of Interest level of Experimental Group in the English Language.....	81
Table 3.5 The English Interest level of Control Group.....	82
Table 3.6 The Classroom Activity about English Learning of Experimental Group	84
Table 3.7 Tabulation of the Results of Control group English Learning Activities.....	85
Table 3.8 Results of Pre-test.....	91
Table 3.9 Statics of the Corpora Compiled by the Selected Sample.....	93
Table 3.10 Lexical Collocation.....	94
Table 3.11 Results of Pre-test.....	108
Table 3.12 The Assignment of Newspapers with the Respective Student Groups.....	109
Table 3.13 Lexical Patten-hunting with Respective Selected Group.....	111
Table 3.14 Pattern-hunting of Lexical Collocation.....	111
Table 3.15 Frequency Difference of Lexical Collocations.....	111
Table 3.16 Mis-collocations in Verb + Noun pattern of CPLENE.....	113
Table 3.17 Mis-collocations of Verb + Adverb Combination.....	114
Table 3.18 Mis-collocations Found in Noun + Verb combination.....	114
Table 3.19 Noun + Noun Mis-collocation.....	115
Table 3.20 Adjective + Noun Mis-collocation.....	117
Table 3.21 Adverb + Adjective Mis-collocation.....	120
Table 3.22 Adverb + Verb Mis-collocation.....	120
Table 3.23 Results of Post-tst.....	121
Table 3.24 Summary of the Methodology.....	129
Table 4.1 Types of lexical Collocations.....	132
Table 4.2 Statistical Measure of Pre-test Data.....	132

Table 4.3 Part-1 of Pre-test.....	133
Table 4.4 Part-2 of Pre-test.....	135
Table 4.5 Results of Part-3 of Pre-test.....	136
Table 4.6 Item-3 of Pre-test (Multiple Choice Questions).....	138
Table 4.7 Tabulation of the Results of Item3.....	140
Table 4.8 Statistics of the Results of Post-test.....	142
Table 4.9 Comparison between the Scores of Control Group and Experimental Group in Pre and Post-tests.....	143
Table 4.10 Comparison between the Declarative and Procedural Parts of the Tests of Both Groups.....	144
Table 4.11 ANOVA (Single Factor) Results.....	144
Table 4.12 Coding Summary of the Interview of the Students by NVivo 12.....	149
Table 4.13 Coding Summary of the Text of Responses of Faculty Interviews.....	157

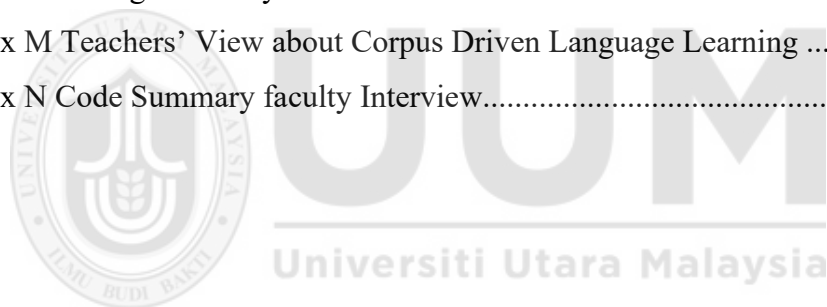


List of Figures

Figure 2.1: Underpinning theories of DDL.....	29
Figure 2.2: Theoretical framework of collocations.....	44
Figure 2.3: Corpus-based research studies on English collocation.....	63
Figure 2.4: Conceptual framework of the research study.....	69
Figure 3.1: Research design.....	75
Figure 3.2: English interest level of experimental group.....	82
Figure 3.3: English interest level of control group.....	83
Figure 3.4: Result of classroom activity of control group.....	85
Figure 3.5: Summary of data Collection for the research design.....	86
Figure 3.6: AntConc software.....	96
Figure 3.7: Nvivo-12 store and organize.....	99
Figure 3.8: Word cloud of students' response to the question "What is regression analysis used for.....	103
Figure 3.9: Frame-work of study.....	104
Figure 3.10: Outline of research design.....	106
Figure 4.1: Results of part two of the pre-test questionnaire.....	137
Figure 4.2: Comparison of the results of post-test.....	140
Figure 4.3: Comparison of the results of post-test.....	143
Figure 4.4: Hierarchy of nodes of the interview text.....	148
Figure 4.5: Analysis of the entire text of the responses of faculty members regarding the effectiveness of CDLL	154
Figure 4.6: Hierarchy of the coded text of the responses of the faculty interview.....	155
Figure 4.7: Text search query of the data obtained from the responses of faculty members.....	160
Figure 4.8: Item clustered with word similarity of the text of interviews of the faculty member.....	162
Figure 4.9: Item clustered with word similarity of the text of interviews of the faculty member.....	163

List of Appendices

Appendix A LNG-403 ELT Methods & Approaches.....	197
Appendix B Questionnaire 1.....	200
Appendix C Pre-test.....	203
Appendix D Post-test.....	206
Appendix E ADV+ NN.....	211
Appendix F ADV+VV.....	246
Appendix G NN+NN.....	250
Appendix H NN+VV.....	267
Appendix I VV+NN.....	272
Appendix J VV+ADV.....	276
Appendix K Students Interview Questions.....	282
Appendix L Coding Summary.....	283
Appendix M Teachers' View about Corpus Driven Language Learning	310
Appendix N Code Summary faculty Interview.....	311



CHAPTER ONE

INTRODUCTION

The first chapter discusses the status of English language in Pakistan along with the lexical behavior of the foreign language learners in the country which leads to the main issues relevant to the research investigation; these issues pertaining the problem with ELT methods & approaches in Pakistan, the unawareness of the modern teaching/learning approaches like Corpus Driven Learning and its significant importance in language teaching/learning at micro level, particularly, the use of lexical combinations like collocations. Moreover, it encompasses the statement of problems which is the bedrock of this study. Subsequently, it discusses the research objectives followed by research questions which are based on insight encaptured from the preliminary investigations. The last portion of the chapter comprises of the significance and scope of the study and ends with the segmat of organization of the thesis.

1.1 Background of the Study

English has emerged as language of transnational communication and people around the globe are adopting and adapting it for their needs (Wessendorf, 2015). Obviously, it has become incumbent upon all world communities, particularly, developing countries to enhance their competence in English language for progress and development. English has dominated the official as well as political discourse in Pakistan as in other developing countries for a long time now (Shamim, 2013). Although much effort has been exerted on the part of academia and policymakers to promote the level of English as a second language

(ESL) since the emergence of the country on the globe, yet the existing scenario is not so encouraging to be claimed as the fruitfulness of these efforts.

1.1.1 Status of English in Pakistan

With the promulgation of English medium schooling policy, the people of Pakistan have witnessed the intensity in the demand of English since the beginning of the 21st century. The National Education Policy has not only recommended English as a compulsory subject up to graduate level but also emphasizes to use English language as medium of instruction for content subjects with effect from 2014 (Channa, 2017). However, The National Education Policy (2018) of the country has admitted the fact that:

“The quality of learning across the country is very low, this is especially true for the public sector and low-cost private schools. National level students’ assessment surveys have, continuously over the years, indicated low mastery in Science, Mathematics and languages. Low level learning outcomes are directly associated with the instructional quality by teachers. Teachers across the country have the low content knowledge and weak pedagogical skills. There is also a deficiency of qualified subject specialists particularly in Mathematics, Science and English across all provinces.”

(Government of Pakistan, 2018, p.8)

Keeping in view of the language disparities, the instructional language has again been revised in the National Educational Policy 2019 which claims:

“ Urdu will be the revised language at the primary level. It will help teachers as well as students in using a language which is more familiar than the current so called medium of instruction (English) and more likely to be used inside and outside the class-room. The change of language will include a revision of text, teacher training material and many assessments to Urdu”.

English will be taught as a subject throughout the journey of primary level student to a standard that will enable students to successfully shift to English medium instruction at the post primary level.”

(Government of Pakistan, 2019, p.12)

As far as the teaching level is concerned, it has also been claimed:

“ Teachers around the country have low content knowledge and weak pedagogical skills. There is also a deficiency of qualified subject specialist teachers particularly, in Mathematics, science and English across all provinces”.

(Government of Pakistan, 2018, p.4)

Currently, In the country, English is being used as a medium of communication in parliament and other important military and civil functions of the government (Channa, 2017). The medium of instruction at a higher level has been declared English, which testifies the status of English as a Second Language in Pakistan (Celce-Murcia & McIntosh, 1991). No doubt, “English has been nativized in Pakistan” (Mehmood, 2009, p.2); the students are learning the English language as a compulsory subject from childhood, from the first class and reach the university level but they cannot speak English

fluently (Adil, 2018) and can not show their competence in writing. Several research studies have been undertaken to explore the nature of the problems faced by Pakistani Learners in learning English as ESL (English as second language) such as the research work of Ahmad et al, 2011; Akram & Qureshi, 2012; Bilal et al, 2013; Haider, 2014; Jahan, & Nasrullah, 2016. Many researchers have worked on the variation exists in Pakistani English (Baumgardner, 1993; Mehmoob & Ahmar, 2004; Rehman, 2011; Uzair, 2012; Demir, 2016). Apart from that, few have tried to prove Pakistani English as a distinct variety of English and formulating its mark within the sphere of “new Englishes” (Mehmood, 2009; Bilal et al 2011; Humaira, 2012; Parveen, 2013; Jameel, 2014; Jilani & Behzad, 2018). Many former researchers have mainly probed into the administrative and other macro-level issues like environment or deficiency of well-trained instructors as key factors hinder the adoption of English as second language in Pakistan; some have tried to prove Pakistani English as a special variety in South Asian Englishes. A very little attention has been paid to the development of vocabulary among EFL/ESL learners of the country, particularly, the way the lexical patterns are being used by these learners. Approvingly, the teachers as well as the learners are not well aware of the modern approaches like lexical approach of language learning which is greatly important for lexical competence. .

In Pakistan, there is still research work to be done to investigate the effect of modern ELT methods and approaches on lexical behavior (the natural trend of the usage of English lexicons) of language learners and their use of lexical patterns, for example; “two or more words that tend to occur together (collocations)” (Lewis, 2000, P.73), as shown in table 1.1, which reveals the language behavior of language users.

Table 1.1

English Collocations

Definition	Example
(arbitrary	to reach a verdict, to commit (a) murder,
recurrent word combinations)	to launch a missile, to withdraw an offer,
(Benson, 1997)	to make an impression, homework, etc.

As claimed by David P. Wilkin in 1992, that little could be conveyed without grammar but nothing could be conveyed without vocabulary. The crucial task faced by foreign language learners is acquiring a sufficient amount of vocabulary. It has been observed by Michael Lewis (1945) that usually our vocabulary consists of prefabricated chunks of various kinds; ‘the single most important chunk is known as collocation’ (Lewis, 2000). Traditional teaching methods at Pakistani schools have led to inadequate vocabulary knowledge.

“The language learners may have ‘competence’, the mental representation of grammatical rules but poor ‘performance’, the use of language for comprehension and production”

(Ellis, 2015, p.18).

Keeping in view of above mentioned fact, the present study endeavors to move knowledge ahead in the field of modern teaching/learning approaches especially the lexical approach with corpus-driven learning method in teaching the English language to EFL learners, particularly, their usage of prefabricated chunks (collocations) in academic writings.

1.1.2 Role of Corpus Driven Instructions in Language Teaching

Corpus linguistics is a branch of linguistics in which language is expressed in corpora including samples of talk and text occurred in real-world. It proposes that reliable language analysis is more feasible with corpora collected in natural contexts with minimal external interference (Almutairi, 2016). Corpus linguistics (CL) has contributed much to growing research addressing the effective description of language used in various genres of writing (Mouringhan, et al., 2014). Since the emergence of 21st century CL has promoted autonomous learning ability among students as an important indicator of successful reform in teaching/learning. Earlier, in the 1990s, Tim Johns has put forth the concept of data driven learning (DDL) initiative across the globe (Boontam & Phoocharoensil, 2018). DDL is the prototype of corpus-driven learning (CDL). The application of corpus-driven learning (CDL) in EFL teaching and learning has marked a fundamental shift in methodology and ideology concerning linguistics studies and language learning (Binkai, 2012).

CDL has been originated from data-driven learning (DDL); the term was coined by Tim Johns (1991) that refers to learners' striving through a large amount of authentic language data to arrive at patterns or rules of language. CDL approach treats language, collected in natural context (corpora), as data and students as researchers undertaking guided discovery tasks. Its major purpose is to identify patterns of language at all levels. It can encourage learners' generalizing skills and raise learners' consciousness of a particular linguistic aspect which leads to greater autonomy and better language learning skill in the long terms (Boulton, 2010). One of the benefits of CDL is it supports effective collocational learning

methods which allow learners to access language corpora, a large amount of authentic data, in the form of concordance lines as examples of particular features of collocations patterns (Chambers, 2010; Gilquin & Granger, 2010). These language features derived from the corpus are more accurate because they are based on real data and more practical as they are expressed in learners' terms (Yaemtui, 2018). Hence, CDL is exploration-based learning which encourages language learners as researchers whose learning is driven by access to corpus data.

The scope of the corpus in teaching academic writing, creative writing, and the contextual use of the English language is worthwhile (Adolph, 2008). In the present scenario, the easy access to large computer-based corpora – language chunks like collocations, naturally occurring in written or spoken text, can be statistically analyzed and we have a better description of the English language available to us ever before (Lewis, 2000). Moreover, it has forced the researchers and linguists to change their views about language learning and teaching. It has even re-evaluated many accepted ideas of language learning/teaching and changed the content what the language teachers teach radically and “these ideas on collocations are only the first ramblings” (Lewis, 2000, p. 59).

1.1.3 Role of Corpus in Investigating Collocations

Corpora represent a rich source of information about the regularity, frequency, and distribution of lexical patterns (collocations) in a language. Collocations constitute a problematic aspect for EFL learners. From a linguistic perspectives, it seems reasonable to

assume that the arbitrary nature of collocations is responsible for the attested difficulties of non-native EFL learners (Boulton, 2017). Over the past few decades, studies of EFL/ESL vocabulary acquisition have identified the significance of collocation in language studies (Gablasova, 2017). Rebecca (2000) has investigated collocational deficiency among Taiwanese learners by comparing the corpus of non-native Taiwanese learners with BNC (British National Corpus) only on the bases of lexical combinations over certain similar words: *'big, large and great'*. Gonzalez in 2013 has worked on a comparative study of collocation in a native corpus and learner corpus of Spanish; he measures the collocational richness of the native and finds them lacking in density, variety, and sophistication. Similarly, in 2017 Demir has worked on lexical collocation and made a significant comparison between non-native and native scholars of English in Turkey.

The inappropriate use of collocations among foreign language learners is a pervasive phenomenon. They often fail to choose the correct combination of two or more words due to their lack of collocational knowledge in vocabulary (Hsue, 2000). The fewer collocations the learners can choose, save them from using longer expressions with much more grammaticalization to communicate an idea or something which a native speaker can express with precise lexical choice. For example, *'adequate supplies meet the demand'* instead of longer expression *'we do not have things enough so that every person who will have one can have one'* (Lewis, 2000, p.59). They prefer to use the lexical simplification strategy which is influenced by the L-1 expressions.

It has been proved that “Speaking natively is speaking idiomatically using frequent and familiar collocations, and the job of language learners is to learn these familiar word sequences. The native speakers are doing this is not only demonstrated by the frequency of the collocations in language but also the fact that conversational speech is broken into “fluent units of complete grammatical clauses of four to ten words uttered at or faster than the normal rates of articulation” (Ellis, 2005, p.128). Hence, the largest load of language learning is mastering the lexicons; and collocational knowledge is an important domain in which all foreign language learners need to be introduced from class one at the very beginning level of learning.

In the aforementioned continuum, to have a better understanding of the language learning of the Pakistani English language learners a corpus-based study of lexical behavior in their writing was investigated and examined in the preliminary investigation of the main study with the perspective of the usage of collocations which is the habitual co-occurrence of a particular word concerning other word or words with a frequency more than usual. The lexical Patterns identified in corpora are used to form a hypothesis about language learning and processing (Gablasova et al., 2017). Hanks (1996) has pointed out this fact that there is a vast amount of collocational knowledge in our everyday utterance. Investigating the use of collocation in English language learning is important as such a study may inform us of the use of restricted collocations in English language teaching and learning (Halim, 2018). Research-based corpora can be useful to language teachers in course design as corpus linguistics research offers exploration and informs the end-users regarding frequently used lexical items (Zahra, 2018). In the context of the role of collocation in

language acquisition, a preliminary investigation was undertaken to rationalize the research problem.

1.1.4 Preliminary Investigation

The preliminary investigation was undertaken to contribute an effort by exploring the lexical behavior (the way of using lexicon in expressions) of Pakistani English language learners, particularly, their usage of lexical collocations in academic writing (Shah et al., 2020). The data selected for investigation was comprising of two corpora: one comprising of 100 argumentative essays composed by Asian graduate university students from ICNALE (International Corpus of Non-native Asian Learners' Essays) and other 100 essays composed by native speakers extracted from LOCNESS (Louvain Corpus of Native English Essays). The selected corpora was annotated with PoS tagger to investigate the comparison between the frequencies (Benson et al, 2005) through the software AntConc, n-gram tool. The levels of co-occurrence restrictions were analyzed while using the models of co-occurrence of restrictions for collocation propounded by Sinclair (1991) and Allerton (1984). Online Oxford Collocation Dictionary was used as reference source for the identification of appropriate collocation. It was revealed that non-native university students often failed to use the appropriate combination of two words due to their unawareness of collocational knowledge in vocabulary. Their choice of collocation lacked the variety and richness in comparison with the native learners. The results have strong pedagogical implication addressing the issues relevant to ELT methods & approaches being practiced in the countries where English is being taught as foreign or second language.

1.1.5 Implications of Preliminary Investigation

The purpose of the PI (preliminary Investigation) was mainly to examine the lexical behavior of non-native foreign language learners based on lexical approach (Shah et al., 2020) and to promote the applications of the corpus-driven learning method. It has actually, rationalized the problem statement of the main research work.

The initial analysis proved that the native speakers used a wider range of the variety of collocation whereas Pakistani foreign English language learners have low variety and density of lexical collocation as compare to the native speakers. The frequent use of mis-collocation is mainly the result of unfamiliarity with collocational knowledge and the influence of the L-1 expressions (Shah et al., 2020).

As far as the pedagogical implications of the study are concerned, the stakeholders of academia, particularly, the policymakers and syllabus designers were made teacher and learners realized the worth of DDL for correct usage of collocation by revising the syllabi contents and materials as the words naturally demand their combination so the learners could sound natural and acquire native-like competence. The study argues that expanding learners' phrasal lexicon and knowledge of word grammar – the pattern, in which it is regularly used, are two most important elements in the study of foreign language learners (Lewis M, 2000). It follows that language learners can widen their knowledge about combination of words newly or previously met and widen their understanding of what those words mean and, more importantly, how they are used. One immediate implication of this

study for the main research work is that it has also widened the range of teaching/learning language approaches to corpus-driven learning.

Based on preliminary investigations, the hypothesis of the main research work has been conceptualized which redresses the core problematic issues; the EFL/ESL learners are facing in Pakistan. Hence, the purpose of the present study is to explore the effectiveness of Corpus Driven learning in teaching collocation to the Pakistani undergraduates and its role in reducing the teaching/learning issues relevant to ELT in Pakistan.

1.2 Problem Statement

English has become a global language; to gain proficiency in this language is the need of the hour. All over the world, particularly, the developing countries are making their best efforts to enhance the competence and proficiency among English language learners. Like developing countries, Pakistani nation is also striving to make her niche among world nations. Although various strategies are being promoted in Pakistan to gain proficiency and competence among the English language learners yet the national Education Policy 2018 has claimed (as referred above) that the level of proficiency in English language among EFL learners is very low, particularly the lexical behavior among nonnative Pakistani English learners is substandard as well as there is lack of skilled subject specialist teachers in English (Government of Pakistan, 2018). The problem is multi-faced which is discussed in the following paragraphs.

Pakistan has multi-ethnic communities that show different cultures and use various languages. Urdu is major language along with the huge percentage of regional languages as shown in table 1.2:

Table 1.2

Statistics of speakers of Other Languages in Pakistan

Language	2008 estimate		1998 census		Areas of Predominance
1 <u>Punjabi</u>	76,367,360	44.17%	58,433,431	44.15%	Islamabad, Punjab
2 <u>Pashto</u>	26,692,890	15.44%	20,408,621	15.42%	KPK
3 <u>Sindhi</u>	24,410,910	14.12%	18,661,571	14.10%	Sindh
4 <u>Saraiki</u>	18,019,610	10.42%	13,936,594	10.53%	Southern Punjab
5 <u>Urdu</u>	13,120,540	7.59%	10,019,576	7.57%	Urban Sindh and Pakistan
6 <u>Balochi</u>	6,204,540	3.59%	4,724,871	3.57%	Balochistan

Pakistani foreign language learners belonging to various ethnic and cultural communities as shown in table 1.2 have the greatest problem of the influence of their L-1 while exposing themselves in the foreign language as has been investigated in the preliminary investigation. Their socio-cultural scenario is the major factor causing hindrance in the way of foreign language proficiency (Channa, 2017). The familiarity once they develop with the linguistic expressions always dominates in their natural utterances and they fail to attain a standard lexical behaviour. for this, English is being taught as a compulsory subject from the first level to the university level but the learners do not show satisfactory competence

in this area. Several research works have been done in the context of the problematic issues facing Pakistani non-native English learners in the field of EFL.

As Akram, (2017) has proved that Urdu orthography affects the procurement of English Phonetics/phonology. The system of sound and spelling of the mother tongue deforms the structure of the words and their associations of the foreign language. In schools, colleges and even universities classroom environments are not suitable for learning the English language as the whole scenario is predominated with regional and cultural perspectives with convention teaching approaches which impede the EFL learning process.

The major problem faced by nonnative Pakistani learners is the use of traditional approaches in ELT classrooms; even with the end of the second decade of 21st century, there is, still the same Grammar Translation Method, valid in most of the public and private sector of education. Most teachers use the mother tongue for the translation of target language concepts and the language learners generally cover their syllabi to take the exams and score marks in the paper rather than the development of the language (Channa, 2017). The teachers also focus on the contents rather than the objectives of the syllabi. There is a lack of qualified subject specialist in English across all the provinces of the country. English language teaching is regarded as being far from a satisfactory level (Fareed, Jawed, & Awan, 2018). Realizing these concerns, the government of Pakistan has taken positive initiatives to launch ELT training programs like ELTR (English Language Teaching Reforms) and Train the Trainer through ICD (Institute of Career Development); yet there

is very little attention has been paid on the incorporation of modern methods such as Corpus Driven Language Learning in the ELT training programs.

There is a serious problem with the English language learners in Pakistan that they try to learn the grammatical structures conventionally rather than acquiring language skills. The traditional method of language teaching/learning led to inadequate knowledge of EFL Pakistani university students (Al-mahbashi, 2015). The learners focus on the individual meaning of the vocabulary items paying less attention to formulaic expressions like collocation (Yaemtui, 2018). Moreover, there are external and internal factors badly effecting the lexical behaviour of the EFL/ESL learners including the learner's family background, his environment, and his way of learning (Nawaz et al, 2015). He/she has no direct exposure to the English language and is not introduced inductive approach with explicit and implicit knowledge of field-depending at all which helps to promote foreign language acquisition (Luo & Zhou, 2017). The learners depend upon the teachers and the materials in case of language learning that's why they fail to use the foreign language in their common communications. They are usually trained under the umbrella of declarative knowledge rather than procedural knowledge which impedes their foreign language proficiency (Fuentes, 2001).

The non-native Pakistani learners are unaware of the knowledge of collocation which plays an important role in promoting lexical behavior of EFL learners. They do not know the natural tendency of words proximity. Native English speakers can easily acquire

collocations because they hear and use English every day. They use collocations without conscious choice to produce proper expression (Kim, 2012). Whereas, the nonnative do not have the opportunity to observe the language patterns as collocations in practice and performance that causes many problems in the acquisition of foreign language.

The role of formulaic sequence and multi-words expression that appear to be stored and retrieved in the form chunks from the memory of native speakers are overlooked in the composition of ELT syllabus contents for the nonnative foreign language learners (Yaemtui, 2018). One of the formulaic sequences that are almost ignored in Pakistan by the ELT syllabus designers in many vocabulary acquisition studies is that of collocations. These prefabricated language chunks need to be stored in the mind of foreign language learners for smooth retrieval at the instant of language usage; it helps to maintain the native-like language proficiency (Men, 2018).

Keeping in view of the aforementioned facts the study has been focused on specific objectives to be achieved for the solution of the problems relative to the EFL/ESL teaching and learning in Pakistan.

1.3 Research Objectives

The study aims to delve into the role of Corpus Driven Language Learning in English language teaching/learning and to make language teachers and learners aware of and better able to recognize the advantage of corpus-driven language learning and to investigate the effectiveness of such teaching approach on the lexical behavior of learners. This study aimed to:

1. observe the effectiveness of CDLL in teaching collocation to Pakistani university undergraduates.
2. seek the Pakistani university undergraduates' opinion towards learning the English language through CDLL.
3. Seek the Pakistani university teachers' opinion towards CDLL in English language teaching/learning.

1.4 Research Questions

Based on the exploratory research design of mixed-method (Cresswell, 2014), the present study endeavors to meet the objectives while answering the following close-ended and open-ended questions:

1. Does CDLL technique promote the knowledge of lexical collocation among Pakistani university undergraduates?
2. What is the Pakistani university undergraduates' opinion about learning English collocation through CDLL?
3. What is the Pakistani university teachers' opinion towards the effectiveness of CDLL in English language teaching/learning?

1.5 Significance of the Study

Modern teaching/learning techniques in the foreign language learning are the need of the hour. These can improve the language proficiency more effectively than the conventional ones. According to Ellis (2005), speaking natively is speaking idiomatically using frequent and familiar collocations. The present study renders a contribution to the importance of DDLL (data-driven language learning) approach (Boulton, 2008) in teaching lexical

collocations. It will make the learners realized that knowing vocabulary is not merely learning new words; it needs to have familiarity with familiar words in proper combination. The study can facilitate language acquisition by making learners notice the features of language inductively they come across while reading or observing in their surroundings autonomously. In this way, learners can develop their mental lexicons to have native-like fluency and competency. The significant implication for teachers is that they need to re-examine the course contents of syllabi of ELT/EFL courses by incorporating certain exercises which focus on co-text and which draw student attention to significant lexical combinations. The study will look at the way how the English language can be improved as a result of the analysis of a large amount of naturally occurring written/spoken text available in the form of corpora on computers through CDLL approach. It will explain the breakdown of the dichotomy between grammar and vocabulary and emphasizes the multiword phrases we use as language chunks in our usage. The study will also highlight the potential of corpora in the context of writing instructions.

1.6 Scope of the Study

The participants selected for this study comprised of 100 undergraduates of Government College University Faisalabad Pakistan who had enrolled themselves for the degree program of BS English (Literature & Linguistics) for the session 2018-22 and 15 faculty member of the same department. These participants belong to the various region of the country and can help to generalize the results easily. To make them aware of the importance of CDLL the undergraduates were treated for 16 weeks (complete one semester) to have pre and post-test scores while evaluating the use of collocation in argumentative essays.

As the research study based on a mixed-method approach, therefore, for the quantitative approach an experimental method is adopted in which the independent variable of the study was the focusing attention of students through corpus driven language learning technique in ELT class room and dependent variable was the development of the knowledge of lexical collocation among the learners. As for as a qualitative approach is concerned, the descriptive variable is the attitude of the teachers and learners toward CDLL.

The target collocation involved in this study were seven types of lexical collocation as shown in table 1.3. The lexical collocations were represented by word pair which were lexically restricted (Yaemtui, 2018). These word pairs have been mentioned in the table.

Table 1.3

Types of Lexical Collocations

Sr No	lexical pattern	examples
1	<i>Verb + Noun</i>	<i>(make a mistake)</i>
2	<i>Verb + Adverb/Adjective</i>	<i>(to show clearly,)</i>
3	<i>Noun + Verb</i>	<i>(results proved)</i>
4	<i>Noun + Noun</i>	<i>(a ceasefire agreement)</i>
5	<i>Adjective + Noun</i>	<i>(heavy traffic)</i>
6	<i>Adverb + Adjective</i>	<i>(extremely generous)</i>
7	<i>Adverb + verb</i>	<i>(simply show)</i>

1.7 Definitions of Terms

The following terms have been recurrently used in the document of the research study.

These need to be elaborated for the convenience of the reader.

1.7.1 Assisted Performance

Assisted performance defines that a learner can do better with help, with the support of the environment or other individuals and of the self. It has its genesis in Vygotskian sociocultural theory which asserts that human learning is a social process; a learner learns or acquires much with the help of social members and other environmental factors. In the modern world of science and technology, we see a new environment (online access to network computing) enables learners to interact for learning as long and whatever they want inductively (Jalil & Mcfarlane, 2008). For example, the sample in the study carried out the assigned tasks with the assistance of their colearners and the teachers.

1.7.2 Deductive Reasoning

Deductive reasoning works from general to specific, a ‘top-down approach’ in which a researcher follows a theory and confirm it with the hypothesis and subsequent observation (Soiferman, 2010). The quantitative methods are followed by deductive reasoning. Herein this research investigation the researcher is following Schmidt’s theory (1990) of Noticing Hypothesis based on Tim John (1936) work on data driven learning and quantitatively measure the effect of CDL on EFL learning in learning the lexical items.

1.7.3 Explicit Language Learning

Explicit language learning refers to direct language instruction in the classroom (Tu & Talley, 2016). It is a type of communicative approach in language teaching/learning in foreign language instruction that arguably approximates the social, professional, and

stimulating travel setting to a certain extent. It is a conscious process that is also likely to be intentional. It can be investigated by giving learners an explicit rule and asking them to apply it to language data (Tu & Talley, 2016). The learners focus certainly on language aspects such as grammatical forms and another linguistic pattern like collocation. In the present study, the term explicit language learning refers to the direct exposure of language for the learners.

1.7.4 Foreign Language Acquisition

Foreign Language Acquisition refers to the learning that typically takes place in the classroom through instruction where there is no or limited opportunity to use the second language in daily life (Ellis, 2015). In Pakistan, English is acquired as a foreign language as the learners have limited opportunities to use in daily life.

1.7.5 Formulaic Sequence

A formulaic-sequence is a ready-made chunk of language that is accessed as a whole rather than generated by joining its elements (Ellis, 2015). For example, ‘rotten eggs’ and ‘rancid butter’ are acquired as a whole by the language learners. In this research study, “collocations” have been used as formulaic sequence. The knowledge of this notion tried to be enhanced through modern technique like CDLL.

1.7.6 Implicit Language Learning

Implicit learning takes place when it is imperative to determine how speech habits are

developed outside the classroom paradigm (Tu & Talley, 2016). Implicit stands for something is being implied but not stated directly. The language learners themselves develop a cognition for certain language patterns while having recurrent exposures of these features from data or language occurred in a natural context. Herein this study, it was conceptualized that the non-natives habitually use inappropriate collocation due to implicit language learning.

1.7.7 Inductive Reasoning

Inductive reasoning works in such a way as the present researcher moves from specific observations to broader generalization and theories (Soiferman, 2010). It is a ‘bottom-up’ approach (Cresswell, 2014) in which conclusions are based on experiments. The qualitative approach in research studies is followed by inductive reasoning.

1.7.8 Lexical Behavior

Most of the naturally occurring language consists of recurrent patterns due to implicit language learning, many of which are phraseological, lexical bundles, and somehow collocations which determine the nativity or other particularities relevant to the speaker (Nesselhauf, 2005). The use of words in speaking or writing exposure of the user shows the lexical behavior of them.

1.7.9 Miscollocation

A collocation is a term used to address the possibility of the occurrence more than usual of the two or more words in syntactic relation (Demir, 2017). The incorrect use of collocation or collocation inappropriacy that is the unnatural use of word combination which asserts non-nativeness is considered miscollocation. This property of lexical behavior has gained currency for the linguistics to determine the linguistic features of foreign language users. The same trait has helped to conceptualize the hypothesis of the current research study.

1.7.10 Pattern-defining

Pattern-defining is a process which refers to finding language patterns when the language learner has this particular pattern in his mind to be used at specific point of text. Generally, it is a matter of knowing some of the component words and seeking a model for the exact structure required (Kennedy & Miceli, 2010). For example, in the present study, the sample of the study does the pattern-defining for English collocations, particularly, lexical combinations.

1.7.11 Pattern-hunting

Pattern-hunting refers to looking for the corpus on particular words or pattern associated with the topic concerned and then scanning the concordance lines for potentially useful patterns (Kennedy & Miceli, 2010). For example, when the students like to write about their personal space, the teacher suggests what sort of things they might like to talk about

their living space. Herein the current research work the teacher assign the learners to do pattern-hunting of ‘lexical collocations’ from the compiled corpora.

1.7.12 Second Language Acquisition

Learning of another language in a context in which it is used as a medium of wider communication is called second language acquisition, for example, learning of English in the United States (Ellis, 2015). Here, the term is used for the efforts of the third world countries like Pakistan to adopt English as second language.

1.8 Organization of the Thesis

This research study shows how modern language teaching techniques like corpus-driven language learning can help foreign language learners to enhance their language competence effectively more than the conventional ways of teaching. Therefore, a dissertation that consists of five chapters has been organized to present a realistic model that is helpful not only for ELT teachers and learners but also for the syllabus designers. The organization of this dissertation is as follows:

The first part of this chapter consists of the background of the study involving the worth of the English language around the globe and more in third world countries like Pakistan along with the foreign language teaching techniques. The second portion of this segment shows how the linguists and the other researchers have contributed their efforts to redress the foreign language teaching/learning problems. It is then followed by the purpose

statement which exactly highlights the purpose of this research work. The second part of the first chapter presents a detailed problem statement that has brought forth the objectives of the research. Then, the research constructs have been coined in the form of research questions on which the investigation has been undertaken. Apart from that, this chapter also highlights the significance and scope of the study. The researcher has also tried to explain the important terms being used in the thesis for the convenience of the reader.

The second chapter provides a detailed description of all the relevant theories which remain helpful in the conceptualization of the hypothesis of this research investigation including the theories of language learning and models propounded by the predecessors. It, then, discusses the concepts of collocation and the models of their restriction in usage. The last portion of this chapter describes the practical work which has paved the way for this research investigation, done by other researchers in the same area.

The third chapter comprises of the overall research methodology involving the philosophical stance of the research, population and sampling, the research data description and organization, and the design of the research study. It also presents the methodology involves in the data analysis using both qualitative and quantitative approaches.

Chapter four extends performance studies from point-to-point analysis of the data to a highly congested communication pattern, the many-to-one collected way. It focuses on the constructs of how a reliable result can be reached. The pre-post test results have been

discussed to validate the hypothesis which has further been endorsed by the opinion sought from the selected sample.

Chapter five not only gives the summary of this research study and how the research questions have been answered but it discusses the directions for future work. Last but not the least, all the relevant material including the question-set items and the results of the corpus tools are contained within the appendices.

1.9 Summary

This introductory chapter has presented the background of the study concerning the status of English in Pakistan and issues of English language teaching/learning approaches in the country. Having reviewed the studies focusing on the importance of inductive learning and learners' autonomy, corpus-driven learning has been proposed as a possible solution with the aim of its learning effects on the development of lexical behavior, particularly, the usage of collocation by EFL learners while comparing the pre and posttest of selected samples of the study. The impact of CDL instruction as an effective learning tool requires more empirical evidence; particularly, in Pakistan where no such study has yet been undertaken in this field. A preliminary investigation on the role of collocation in foreign language learning based on noticing hypothesis has helped to conceptualize the hypothesis of the main research study; making teachers and learners aware of the importance of modern approach like corpus-driven learning in the development of foreign language

proficiency particularly through noticing the collocations and their role in inter-language grammatical patterns.



CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter is mainly bifurcated into two parts the first part envisages the linguistics phenomenon which forms the central focus of the thesis: the concept of ‘Corpus Driven Language Learning (CDLL)’, to outline some of its underpinning theories as these have been portrayed within the different descriptive framework and to indicate why CDLL is important for EFL/ESL teaching and learning. Section 2.3 discusses the role and nature of collocation and co-occurrence of restrictions in English collocation and available literature in this respect. The next section 2.4 encompasses various approaches to study collocations propounded by famous theorists and linguists. The second part of this chapter discusses a detailed coverage of previous researches on Data-Driven Learning from the perspective of second language acquisition.

2.2 Theoretical Framework

The merging thought of data driven learning by Tim John (1991) has opened new skylines in the field of the foreign language learning. It is one of the inductive methodologies by involving corpus data in which language becomes data and the student becomes analyst (Boontam and Phoocharoensil, 2018). It is a find based language learning approach among EFL students who extrapolate rules in view of their scrutiny of the concordance yield. Whereas, conventional language learning stresses rule-based learning in which language

educators lay a lot of accentuation on the significance of syntactic standards and students keep them latently relying upon their educators in all regards.

In spite of the fact that data driven learning is considered to epitomize an alternate way to deal with language learning yet it demonstrates an inductive way to deal with the foreign language learning; it would be useful to perceive how it fits with principal language learning speculations that support it, specifically the noticing hypothesis, constructivist learning and vygotskyan sociocultural hypothesis (Boulton, 2012) which are being examined exhaustively in the accompanying passages.

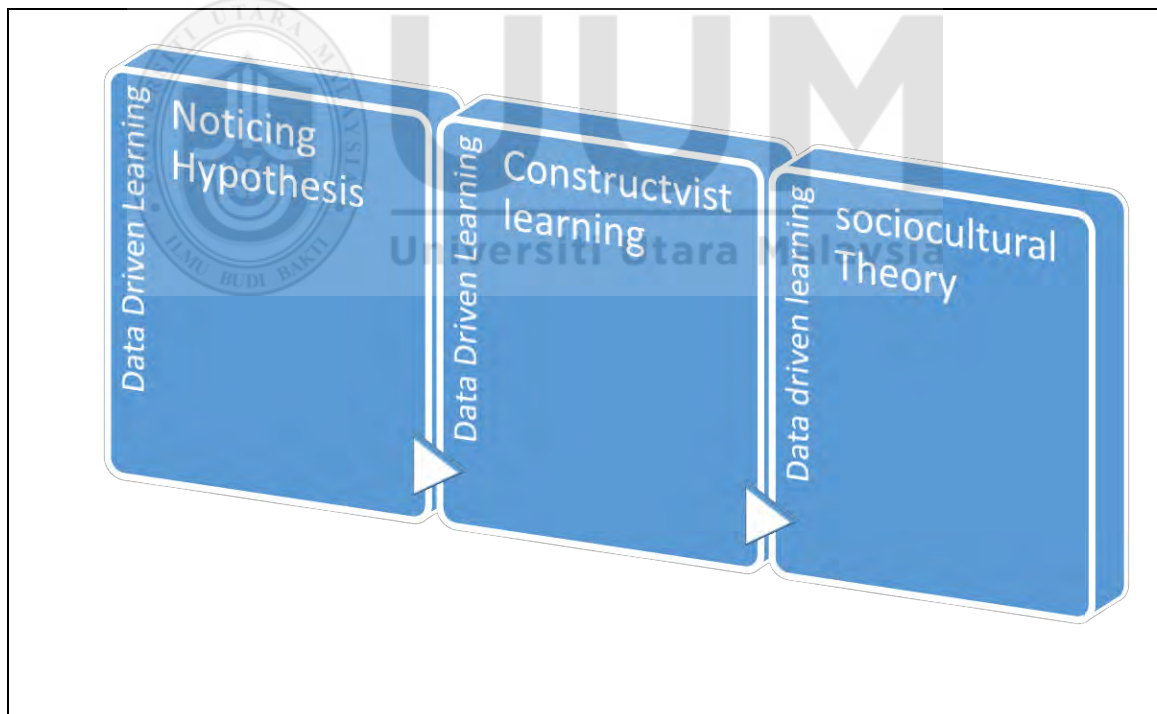


Figure 2.1. Underpinning theories of DDL

2.2.1 Noticing Hypothesis

The noticing hypothesis has been unmistakably examined with regards to second language learning studies. It holds that 'students' obtaining of semantic information is bound to increment assuming if their attention is intentionally attracted to linguistic elements (Schmidt, 2010). A not many slivers of proof make express references in writing; in any case, Papp (2007) has shrewdly propounded while checking on the articles of different linguists, the mental cycles supporting students' noticing the distinctions of the language they created and the language they experienced:

Firstly, students need to focus on structure (Schmidt, 2010); then, at that point, they need to see highlights of their interlanguage grammar (ILG) and target language (Truscott, 1998); then they should be capable inductive learning component to have the option to make a speculation, relationships and recognize designs in the objective language (Luo and Zhou, 2017); at last, they need to contrast their ILG and target language with chalk out confounds and discrepancies (Boulton, 2012). This is what kavaliauskiene (2013) called "subjective jump to cognizant insight" (Jiao, 2016).

Early corpus exercises connected with SLA studies include the mental course of noticing hypothesis, which commonly of these procedures will in general have a place with the inductive methodology. This approach includes unconstrained noticing by the actual students or coordinated by the instructors. In specific cases, where the students are provoked by the educators to search for explicit language examples or elements could be

seen as "educational mediation"(Gilquin and Granger, 2010a). Flowerdew (2009) has proposed four cycles of instructor coordinated noticing exercises: Representation (taking a gander at data), Cooperation (conversation and imparting insights), and Intercession (educator's intervention, and Enlistment (making one's own standard for a specific component.

Kennedy and Micelis (2010) have proposed two sorts of noticing exercises: Pattern hunting and pattern defining. Pattern hunting procedure incorporates perusing the entire text in view of title or test type and searching for recurrence list for familiar word blends while, pattern defining functions for the particular reason when the understudy has a specific example to him to see in the data (Kennedy and Miceli, 2010). The two kinds include noticing as example characterizing with respect to the educator and example hunting with respect to understudies. In addition, asking learners to look at potential keywords or certain blends a long time before they counsel a corpus should have been visible as noticing-hole hypothesis as the learners attempt to contrast their conjectured highlights and the corpus data. Along these lines, students can resolve designs inductively even with next to no direction. In this manner, Karashen's (1984) guarantee that understandable information is an adequate and vital condition for second language learning needs modification in light of DDL. As Lixum (2001) calls attention to the particular element of DDL way to deal with inductive instructing is that the language data is fundamentally, and the educator doesn't realize ahead of time what rules or example the student will find (Lixun, 2001).

There are two types of noticing: strong and weak form (Boulton, 2012). The solid structure guarantees that consideration should be paid to a few part of the stimulus environment and that viewpoint should be seen before a psychological portrayal of that improvement previously framed. While, the more weak structure claimes that individuals find out about the things they take care of and find out a lot of about the things they don't take care of. In the event of noticing hypothesis purposeful consideration that is explicit learning is vital for contribution to become consumption. Subsequently DDL recommends that language learning ought to comprise generally of conscious-raising (CR) exercises instead of the educating of rules (Hadley, 2002). CR alludes to make language students mindful of the specific examples of the objective language and he/she intentionally utilizes concordance to see the elements of this specific example.

2.2.2 Constructivist Learning

Constructivist learning is an instructive way of thinking that sees the procurement of information as a unique process, with students in the driving seat (Flowerdew, 2015). It is an inductive way to deal with learning in which students include in discovery learning with higher-order coherent abilities like making speculations and deductions. Collentine (2000) has obvious constructivism with regards to the inductively-arranged worldview of language learning:

“Cognitively speaking, constructivism is a learning theory that departs from the traditional objectivist perspective of knowledge acquisition....Objectivism in its purest form assumes that the essential elements of instruction are communication and deduction. In this view, when educators adequately comprehend a phenomenon, they

can describe it and transfer that knowledge to students. To verify that knowledge has been transferred, a teacher asks learners to utilize an explanation in tasks where they apply a construct or rule (e.g. a translation exercise, a close passage). Constructivism, on the other hand, presumes that the learners build knowledge actively, largely through the inductive process”.

(Collentine, 2000, pp 47-48)

Successful language learning through a constructivist learning methodology permits investigation and gives various points of view on the designated peculiarity. Through investigation, students investigate, contemplate, and consider while refining a hypothesis that makes sense of how a peculiarity functions. General mental critical thinking systems and background knowledge pertinent to the peculiarity assist student with getting a handle on new data (Collentine, 2000). Seeing peculiarity according to different viewpoints (source of legitimate data) improves the probability that the peculiarity will become remarkable to the students. For instance, two alternate points of view of the basic development of language tests could give the perspectives of meaning and structure.

The plan of web index interfaces for syntax guidance which permits students to either work insightfully to get to grammar rules or inductively to see language pattern is additionally considered to advance constructivist learning (Flowerdew, 2015). Concordance is one more strategy for language learning approach in which students follow constructivist learning to concentrate on structure or lexical examples like collocations (the routine juxtaposition of single word with different words) tracked down in the computerized

database (Yaemtui, 2018). Corpus-based devices manage the cost of students a constructivist climate, advancing their higher-order mental abilities related with inductive learning to deduce linguistic pattern to have a more profound grasping (Costa, Plonsky, and Starfield, 2018).

2.2.3 Sociocultural Theory

Vygotsky's (1978) sociocultural theory expresses that human learning is a social process having its starting point in the public eye and culture wherein the individual is residing. Social connection assumes a principal part in the improvement of discernment among students. Vygotsky accepts that everything is learned at two levels: First, through connection with others then incorporated into the person's psychological construction.

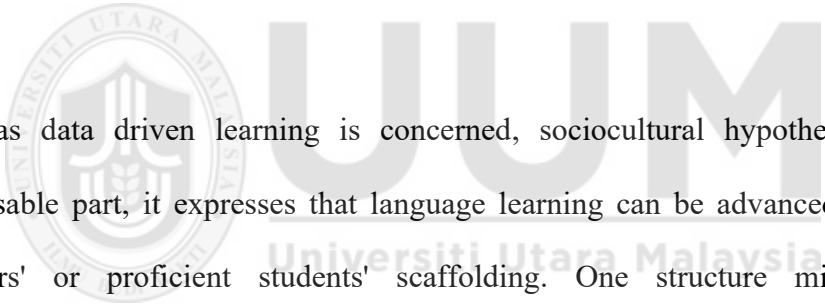
“Every function in a child’s cultural development appears twice: first on the social level and then on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological). This applies equally to voluntary attention, to logical memory, and to the formation of concepts. All the higher functions originate as actual relationships between individuals”.

(Vygotsky, 1999, pp. 86)

The second piece of sociocultural hypothesis portrays the possibility that the potential for mental advancement is restricted to the "zone of proximal development" (ZPD). This is the area of investigation for which the student is intellectually ready yet needs assistance through friendly association to completely create (Hadley, 2002). An instructor or more

productive friend can give the student 'scaffolding' to help him developing comprehension of information spaces and improvement of more significant level mental abilities.

Vygotsky's (1934-86) sociocultural hypothesis has been highly talked about and explained in quite a bit of SLA literature. That's what it contends "the development and working of all more elevated level mental cycles (perception) are undeniably interceded, and that language is one of the most significant intervening apparatuses of the mind". Perception is created through students' social connection by means of discourse, either dialogic or private, to get the sense of meaning (Flowerdew, 2015).



As far as data driven learning is concerned, sociocultural hypothesis assumes an indispensable part, it expresses that language learning can be advanced through either instructors' or proficient students' scaffolding. One structure might be shared communication for checking and criticism in little gatherings; that has been extended to DDL. Flowerdew (2012) covers a composing module which efficiently integrated DDL into following three phases:

Stage 1: Instructor distinguished mistakes on a diagram.

Stage 2: Student distinguished blunders presentation.

Stage 3: Students produced inquiries for reviewing a last report

Rambling stage 1 and 2 of the corpus conference meeting, in the soul of the sociocultural methodology, students were purposefully assembled while blending weaker students with proficient ones to cultivate cooperative discourse through "assisted performance" for forming searches and conversation of corpus yield. Along these lines, information is built and assimilated.

2.2.4 Corpus Driven Language Learning

The term Corpus Driven Language Learning (CDLL) has been taken from data-driven learning (DDL) in which language becomes data and students become analysts (Johns, 1991). Similarly, when students get ready adhoc corpora to determine specific linguistic patterns or qualities inductively they include in CDLL. In the current study, the selected sample is supposed to examine the collocational information which they got from the corpus arranged separately with the help of the educator while following the previously mentioned speculations of language learning. The following segments of this chapter depict the idea of collocation extensively.

2.3 Explanation of English Collocation

Speculations generally act as a good source of thoughts that typically function to generate a hypothesis (Costa et al., 2018). The hypothetical and reasonable piece of my research study starts with the following of hypothetical impacts on the dependant variable of the study, ' the utilization of collocations by nonnative EFL students'. The term, collocation, has its beginning in the lexical-field theory propounded by a German etymologist Joster

Trier in 1931; he contends that words procured their significance through the relationship of different words inside a similar field. He set out the underpinning of the lexical methodology in EFL/ESL learning. This approach advocates that a significant piece of language learning comprises of having the option to comprehend and deliver lexical expressions as chunks (Lewis, 1993). These lexical expressions and chunks are known as collocations. The current study has hypothetical establishments in this theory which has later been embraced and extended by a few language specialists by embellishment and reshaping the ideas propounded in lexical field theory.

As per the internet based Oxford dictionary reference, the word collocation stands for: The routine juxtaposition of a specific word with other word or words with a recurrence of occurrence more prominent than once.

The idea of collocation was instituted by J. R. Firth in 1951, the father of collocation in his renowned citation "you will know a word by the company it keeps" (Firth 1957, p.11). Numerous different definitions were propounded by language specialists to allude to collocation including, 'composite element' (Mitchell, 1971), 'idioms' (Fuentes, 2001), 'gambit' (Farghal, 2007), 'lexical phrase' (Nattinger, 1992), 'fixed expression' (Moon, 1994), 'multiword lexeme' (Jackson, 2002), 'formulaic expression' (Kuiper, 2004), 'fixed order of connection between words or jargon' (Ismail, 2006). Other than different thoughts propounded in regard of collocations, linguists share a typical view that collocations are words blend focusing on 'the event of at least two words inside the short space of one

another in a text' (Sinclair, 1991). There are different conclusions on the idea of collocations; one thing whereupon everyone concurs is that, not normal for syntactic developments, there are no immovable guidelines for the construction of collocations. Just the utilization of extra time figures out which words go together. We have no clue about why individuals throw and hold parties as opposed to make them however we realize it is an overall way since that individuals say (Lackman, 2011). Ken Lackman further depicts that some say that phrasal verbs are collocation, for example, 'take up, set in, and so on; semi-fixed articulations which can be utilized over and over again by transforming single word or a piece of it is viewed as collocations by certain individuals. Fixed expressions like phrases are additionally included in the classification of collocations. Notwithstanding, in fact, applicable to educating, it is proposed that prepositional expressions, phrasal verbs, semi-fixed or fixed articulation stay separate classification and collocations be limited to principally two words blends comprised of fundamental grammatical forms (noun, verbs, adverbs, and adjectives...).

The previously mentioned definitions expressly allude to two primary ways to deal with characterizing collocation: recurrence based and phraseological definitions. Various terms have been utilized in writing to allude to this differentiation: empirical versus hypothetical quantitative versus qualitative methodology (Soiferman, 2010) separately. The first is the manner by which regularly words are seen together in a corpus and utilizations factual proportions of solidarity of relationship as rule for collocation extraction. On the other hand, The subsequent methodology bargains collocations as a grouping of words that meets certain phraseological standards somewhat and involves local speakers' instinct as an

approval of their status. The following segment of this section will examine the expansion of the frequency based way to deal with collocation.

2.3.1 Major Approaches to Collocation in Linguistics

In this phase, three major approaches to collocation are discussed:

- I. Firthian approach and Neo-Firthian approach
- II. Semantic approach
- III. Structural approach

2.3.1.1 Firthian Approach

In the context of Firthian approach, the meaning is considered as 'a complex of logical connection' (Williams and Millon, 2012). Firth weights on the crucial role played by collocation in deciding the significance of a word, as should be visible from a renowned comment that "you will know a word by the company it keeps." (Firth, 1957, p.179). He makes sense of the place of words in collocation is routinely positioned and the meaning of these words is commonly expected and perceived. Though, for colligation, he affirms that 'meaning at linguistic level comprises of the interrelation of the classifications which are conceptual in nature'. Firth's differentiation among collocation and colligation is a valuable one since the two arrangements with importance at a syntagmatic aspect. Nonetheless, the significance of colligation works at the linguistic level, which is not the same as the importance of collocation at the collocational level.

M.A.K Halliday and J.M.Sinclair further fostered the Frithian approach. This Neo-Firthian way to deal with collocation advances a progression of ideas, for example, "lexical items", "lexical set", "node", "spane" and so on to lay out a lexical theory free of language structure. As indicated by Halliday not all semantics peculiarities can be represented inside the structure of language. Co-occurrence in any succession is the essential primary standard; regardless of mediating material: features of language, for example, favored groupings or ongoing mediation are auxiliary in structure (Sinclair, 1991). He, however, recognized 'casual' and significant collocation in light of a couple of events and manifestation of above possibility of recurrence respectively.

2.3.1.2 Semantic Approach of Collocation

Like Firthian and Neo-Firthian approach separated from grammar, the semantic methodology additionally examines syntagmatic relations yet inside a semantic structure. As Lyons brings up in 1966, sentences might be linguistically very much framed, however they will be 'simply rubbish' on the off chance that they don't have the ramifications of expression according to a semantic perspective. There are three significant places of the semantic methodology:

1. The degree to which lexical items are consolidated in the clearest normal for collocations. Lexemes differ gigantically concerning the opportunity with which they can consolidate in syntagms with different lexemes. For instance, the adjective "good", "bad" and so on might be utilized as a modifier of any noun however "rancid" consistently predicates the noun "butter".

2. There exist limitations between the lexical items in collocation with the condition the lexical co-occurrence semantically. These limitations are semantically characterized as far as units like arguments and predicates (Wanner et al., 2013). For example, "gold claims a man" is a semantically bizarre sentence as "man possesses gold".

3. "Encapsulation" is one trait of the connection between the lexical things in a collocation. On the off chance that a lexical noun is habitually co-occured with a limited arrangement of other lexical noun, it might come to embody their senses (Lyons, 1977). An illustration of such embodiment is presented by Lyons is the regular collocation of "drive a/the vehicle". The sense "a/the vehicle" is frequently epitomized with the action word "drive".

2.3.1.3 Structural Approach to Collocation

This approach depends on syntactic designs, lexical noun, and the nouns with which they arrange. It has three significant focuses as examined beneath:

1. Collocation can be sorted into two kinds: lexical and grammatical collocation. The distinction between them is:

Grammatical collocations comprise of the prevailing word - noun, modifier/participle, and verb - and proposition or syntactic developments.

Lexical collocations, then again, don't have predominant words; they have structures as action word + noun, modifier + noun, noun + verb, noun + noun, intensifier + adjective, qualifier + verb (Benson, et al., 1997).

2. The primary way to deal with collocation looks at 'how sentence structure is incorporated into lexical as well as the other way around through collocation and designing (Nelson, 2000). The example of a word is characterized as "every one of the words and designs which are routinely connected with the word and which add to its signifying" (Hunston and Francis, 1999, p. 37).

The connection between linguistic designing and lexical importance found by Hunston and that's what her co-specialists is "every one of the words can be displayed to have examples and words which have similar examples will generally share a part of significance (Hunston, at el., 1997). For instance, a few action words have designs "V by - ing" where the action word follows by the relational word 'by' and 'ing' condition. The majority of the verbs fall in this classification shares the importance:

- I. those implications "begin" or "finish". Counting: start, close, end, finish, begin, and so forth. Those signifying "answer or make up for something", including offer penance, redress, answer, respond, counter, respond, fight back, and so on. Huston and Francis (1999) some way or another, share similar examples as depicted in BBI word reference e-g "modifier + noun", "noun + to-infinitive", "verb +noun" and so forth. The examples utilized by the Hunston conspire don't include the customary utilitarian classes like an item, supplement, assistants, and so on in view of their restricted examination.
- II. It is valuable for instructive situated collocation studies. The blend of grammar and lexis into language depiction addresses a gathering point between the worries of teaching method what is that students need to be aware - and those

of hypothesis - how language can generally sufficiently be described (Hunston and Francis, 1999).

A few specialists have done their research work in light of a primary methodology in the field of EFL. For example, Gitsaki (1999) inspects the advancement of collocational information among Greek understudies in view of 37 lexical and linguistic collocation. Yuanwen Lu in 2017 has examined 19 lexical and syntactic collocation examples of Chinese students and demonstrated the significance of collocation in EFL educating and learning.

The underlying methodology focuses on the combination of grammar and lexes into depicting syntagmatic relations of lexical items. It is, hence, more legitimate to refer to it as "integrated approach" and it shows that an integrated approach is the most ideal way to portray the syntagmatic connection of lexical items. The current study is to a great extent founded on a coordinated way to deal with collocation. The previously mentioned conversation in the viewpoint of different ways to deal with collocation should be visible as a perceptible perspective on collocation. We will presently switch towards a tiny perspective on collocation in the English portrayal.

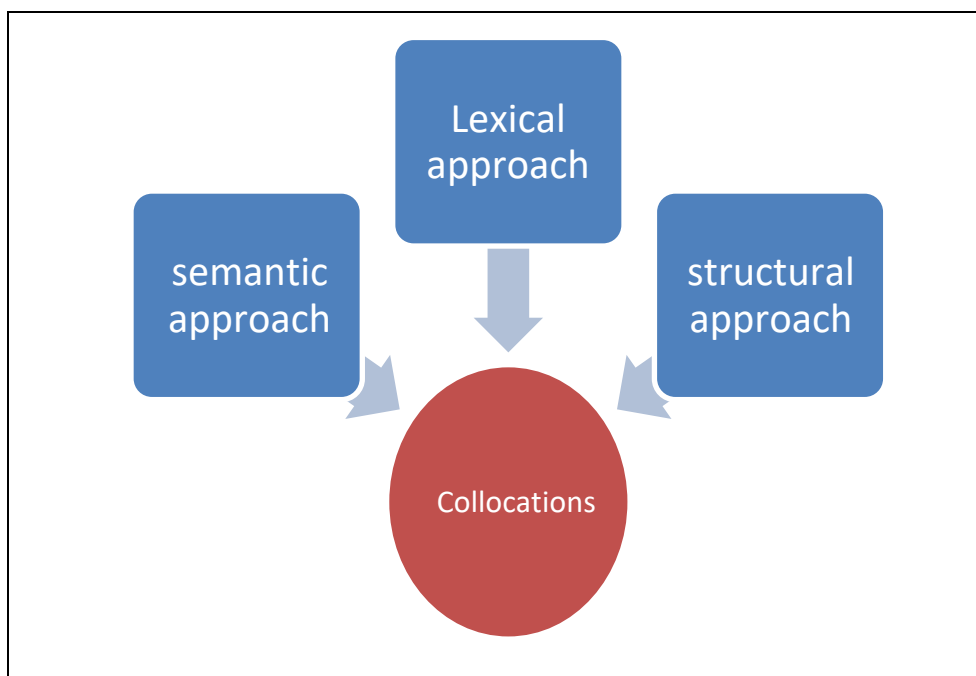


Figure 2.2. Approaches to collocation

2.3.2 Levels of Co-occurrence Restrictions in English Collocations

English collocations can be depicted from various levels in view of co-occurrence restrictions. For the frame work of the current review, two arrangements of four degrees of collocation co-occurrence limitations as proposed by Sinclair in 1996 [collocation, colligation, semantic inclinations, and semantic prosody] and Allerton prior in 1984 [syntactic restriction, semantic restrictions, location restrictions, and pragmatic restriction], are followed. They give a precise method for portraying the syntagmatic relations between lexical items, from the substantial degree of genuine collocation and syntactic standards to the theoretical degree of pragmatics. The two models are consolidated to begin the assessment of genuine combination of a word its linguistic and semantic restrictions to its semantic prosody.

2.3.2.1 Sinclair's Model

Sinclair proposes four levels for the examination of co-occurrence of restriction: collocation, colligation, semantic inclination, and semantic prosody.

1. According to Sinclair, "Collocation is a regular co-occurrence of words". Collocations are real words and they can be straightforwardly seen in literary data. Their likelihood of co-occurrence can be genuinely estimated by t-score or MI-score (Stubbs, 1995)

2. John Firth prior has expressed that Colligation alludes to the interrelation of syntactic classes. While, Sinclair alludes to colligation as "co-event of linguistic decisions". Hoey characterizes it in an unexpected way, as per him colligation is "a syntactic organization word keeps and the position it likes". For instance, there is areas of strength for a for "accountant" with a classifier as in "a wages accountant" (Hoey, 2000).

3. "Semantic inclination" is a lexical arrangement of habitually happening gathers sharing chosen semantic elements (Sinclair, 1996; Stubbs, 1999). The lexical set is conceptual and not straightforwardly seen in literary data (Stubbs, 1999). For instance, an unmistakable semantic inclination for "genuine inclination" is "articulation", which is induced from the action words with which "genuine inclination" oftentimes gathers "express", "convey", "show", "uncover", "share" and so on. (Sinclair, 1996)

4. "Semantic prosody" shows the speakers' mentality towards what is being discussed and is at the even minded degree of language (Louw, 1993; Sinclair, 1996; Stbbs, 1999). This level is additionally theoretical and not recognizable straightforwardly. For instance, Sinclair notices the striking element of the phrasal action word "set in" like subject is in

every case, for example, "rot", "sick well" or "bias" and so on for the most part alludes to a terrible situation. Stubbs in 1996 has additionally separated 'semantic prosody' into three classes: normal, positive, and negative. As "set in" is an illustration of negative semantic prosody.

2.3.2.2 Allerton's Model

Allerton (1984) has additionally proposed four degrees of co-occurrence of restrictions: syntactic, semantic, locational, and pragmatic.

1. "Syntactic limitations" alludes to syntactic principles between lexical items in a collocation. In the event that a speaker or essayist neglects to keep the syntactic guideline, for example, an understanding among subject and action word, a non-sense sentence will be delivered.
2. "Semantic limitation" is the 'shared semantic similarity' of each word in a sentence, which is its general significance (Allerton, 1984). It is many times the situation for language students that they produce a syntactically wonderful sentence however it sounds unnatural to local speakers. This peculiarity shows that the significance of word-blends isn't just the amount of the importance of each word: there should be some more unobtrusive sort of semantic word choice.

Semantic restrictions are additionally recognized under three relations: "fitting", "reshaping" and "portrayal" (Allerton, 1984). "Semantic fitting" alludes to the way that

lexical assemblies change the importance of their accomplices. For instance, 'an English Piano' would ordinarily be expected that piano is made in England; though, "an English musician" need not have been brought into the world there. The basis passing judgment on Britishness is different due to various gathers of "English" (Allerton, 1984). Semantic "reshaping" is the reevaluation of a word because of its collocations. For instance, "dentures" are not genuine teeth but rather a substitute for them (Allerton, 1984). Semantic "portrayal" alludes to the peculiarity by which a quality requires the presence of specific attributes in an element it describes. For example, "the old maid's better half" is odd because of inconsistent properties of "old maid" and "spouse" (Allerton, 1984). "Semantic portrayal" is something very similar, as the "relevant overt repetitiveness rules" proposed by Parasite (1974).

3. "Locution restrictions" are known as "colloquial" in Allerton's model. It alludes to the etymology peculiarity that the syntagmatic relations of lexical things on this level are erratic (Allerton, 1984, p. 28-36). For instance, the relational word in numerous verbal expressions is erratic e-g "lose faith in regards to", "trust for", "snicker at" and so on.

4. "Pragmatic restrictions" allude to 'a custom for favoring one out of a few (close) interchangeable expressions or sentences (Allerton, 1984). For example, individuals like to pick "lower group articulation" as 'twelve months' is liked to 'a year' (Allerton, 1984).

From the previous conversation, it very well may be expected that both of the above-depicted models the initial three levels manage the linguistic features at both lexical and syntactic level, while, forward one includes the practical thought. Notwithstanding, the two

arrangements of four degrees of restrictions proposed by Sinclair and Allerton give an efficient method for portraying the syntagmatic relations between the lexical things and the capability of language toward the end. Their examination shows the undeniably dynamic connection between co-occurring lexical items (Lu, 2017).

2.3.3 Approaches to Collocational Knowledge of EFL/ESL Learners

As talked about above, corpus-based examinations are frequently associated with language learning/instructing. It is subsequently normal to find a lot of corpus-based investigations of student English. In this segment, I will start with a show of the significant ways to deal with the examination of student English, then, at that point, continue on to present the corpus-based investigations of the collocations in EFL/ESL students English, trailed by a conversation of corpus-based investigations of crude data and label blends.

2.3.3.1 Major Approaches to the Analysis of Learner English

Two ways to deal with the investigation of student English have affected language educating and learning: blunder examination and corpus-based approach.

(1) Error Analysis (EA)

Error Analysis is the examination of errors committed by language students, including correlation and depiction of student's language creation and target language to find crisscrosses (James, 1998, p. 5). EA has its beginning in the contrastive examination (CA). It is, subsequently, important for us to investigate CA. The term acquired money during

the 1950s 1960s. Fries one of the main defenders of CA presents his view on second language educating and learning:

“The most efficient materials are both that is based upon a scientific description of the language to be learned, carefully compared with the parallel description of the native language of the learner”.

(Fries, 1945, p. 9)

(2) Contrastive Analysis (CA)

Contrastive Investigation includes two stages. The initial step is to depict similar elements of L1 and L2, and afterward contrast the two languages along with the confuses that would typically cause obstruction and mistake (James, 1998,). In light of the view that troubles in second language learning could be anticipated by CA and legitimate advances could then be taken to diminish them, a colossal number of showing materials were delivered. Be that as it may, a few shortcomings are likewise guided out concerning the materialness of CA toward language learning and educating. For instance, the distinctions somewhere in the range of L1 and L2 may not cause hardships; CA can anticipate a portion of the learning issues, not every one of them; and so on. (Jhonson, 1975). Because of these deformities of CA, language specialists started to move their thoughtfulness regarding EA.

By the last part of the 1960s, EA had replaced CA (James, 1998). EA is the examination of blunders committed by language students. It includes portraying the student's language creation and target language and contrasting these two with find jumbles (James, 1998). EA takes the position that 'mistakes can be completely cleared up' regarding L2 without

reference for L1 (James, 1998). This conviction has likewise met with analysis from EFL specialists. Howl has raised three issues with respect to EA:

- 1 It can be deceiving to depict errors exclusively from the objective language;
- 2 It disregards the student's ability, which is what the student can do at some random transformative phase;
- 3 ‘Counting the raw number of errors might make people overlook those structures learners do not attempt to use as much as other structures, that is, ‘avoidance of understanding of learner English’.

(Howl, 1995, pp. 4-5)

Nonetheless, both CA and EA have areas of strength for them flimsy spots. The most effective way of managing student English is to coordinate the two of them to contribute a lot to the comprehension of student English.

2.3.3.2 Corpus-based Approach & Data-Driven Learning (DDL)

Computerized corpus studies showed up in the mid 1990s as another promising examination field. With the appearance of corpus innovation, analysts enjoy the benefit to assemble and explore enormous measure of data. As Leech (1998) notes, interestingly it likewise permits an orderly and nitty gritty investigation of the student's semantic way of behaving from 'abuse' and 'underuse' (Parasite, 1998). In any case, Leech (1998) likewise takes note of that the expression "underuse" and the differentiating "abuse" ought not be utilized in a critical soul; since it is quarrelsome to set the composition of local speakers as

the standard for students to keep, they ought to be deciphered as a helpful shorthand for 'fundamentally more/less regular use by non-local speakers than local speakers'. Hence, the current review won't utilize "abuse" and "underuse" in that frame of mind of Chinese student English.

The fundamental way to deal with the corpus-based investigation of student English is contrastive, however it is not quite the same as customary CA 'in the new feeling of looking at/differentiating what non-local and local speakers of a language do in a practically identical circumstance' (Prey-Woodley, 1990). This corpus-based contrastive methodology lies at the core of EFL student corpus-based investigations. There are two correlations in this contrastive analysis(Granger, 1998):

- 1 Between student creations of various native language foundations. It means to acquire a superior comprehension and understanding into the idea of student language. By contrasting student English and the foundation of various first languages, it can assess the impact of various factors on student creation, for instance, students' age, capability level, L1 foundation, task type, learning setting, medium, and so on moreover, cross-semantic highlights can be uncovered in view of the correlation of these different student languages.
- 2 Between student creation and local speakers' creation. Uncovering the elements of unfamiliarity in student production is expected. Corpus-based investigations of student English can complete EA as well as quantitative examination. For instance, the abuse, underuse and abuse of specific words, expressions or designs, and so on can be uncovered plainly in this sort of study. These elements of student English can be all the more

effectively found. Granger (1998a) brings up that this examination requires a control corpus of local speaker English since numerous language highlights are "style-touchy" (1998a: 13). One of the most popular undertakings of corpus-based investigations of student English is the Global Student English Corpus (ILEC), which includes articles by cutting edge EFL students from different nations. After this short conversation of the two ways to deal with student English, let us currently go to the new way to deal with language instructing/learning.

Johns (1936-2009) has spearheaded data-driven learning and instituted the term DDL

which has demonstrated a cutting edge unknown dialect approach.. In this methodology of language learning, language is treated as data and understudies as specialists undertaking directed disclosure errands while in the regular approach to educating/learning a large portion of the language learning is directed by the educators and reading material. In an article, it showed up, Thomas' errand based exercises found English examples with Sketch Motor exemplified DDL as proposed by Johns. Other late books on DDL which acknowledge Johns as the originator of the methodology incorporate those by Anderson and Corbett (2009), Reppen (2010), Bennett (2010), Flowerdew (2012) and Friginal (2018). As the title of the current review proposes it particularly affected by DDL. The review is grounded in friendly constructivism hypothesis, which draws vigorously on crafted by Vygotsky (1978). The hypothesis depends on the reason that information isn't sent to the students rather it ought to be developed by them (Ayas, 2006). Social useful has serious areas of strength for a for the overwhelming majority instructive practices to advance powerful learning one of which is DDL.

2.3.4 Research studies on Collocations among EFL/ESL Learners

Grammar is constantly given a lot of significance in a language educating as contrast with the Lexis. In any case, lexis is as of now not the survivor of segregation in SLA since the meaning of lexis has been acknowledged in ongoing many years (Laufer, 1997). More consideration has been paid to collocation among the investigations of lexis. Granger (1998b) has made sense of the utilization of collocations has demonstrated a significant focal point of interest in EFL since ninties. As she notes:

“First, the emergence of lexico-grammar, inspired by Halliday and Sinclair, has promoted the syntagmatic investigation of lexis. . . . Secondly, corpus linguistics plays an important role to uncover and analyze the lexical patterns by giving the linguists computational means”.

(Granger, 1998b, p. 145)

Despite the fact that collocation in student English has been getting expanding consideration among EFL analysts, not much examination has been finished around here. The purposes behind it is possible that it isn't not difficult to gather student data and 'student corpus arrangement is a careful interaction's (Granger, 1998a, p.10). The accompanying segments will survey a couple of concentrates on collocation in students' English composition, trailed by a conversation of studies including a label mix of data from student English.

2.3.5 Corpus Studies of Collocations in Learners' English writing

In this part, I have reviewed the work done by De Cock, et al. (1998), Granger (1998b), Kaszubski (2000) and Milton (1998) on collocations in learners' English, as their work has affected the corpus investigations of collocation in raw data.

De Cock, et al. (1998) revealed a mechanized way to deal with the collocations utilized by EFL students. They investigated the collocational ability of students in view of the examination between two corpora: French speakers and local speakers. Their PC program can naturally remove current word-blends. By looking at the general frequencies of two-, three-, four-and five-word current mix and a few formulae in the two corpora, they presumed that exceptional and French English students utilized collocations, at times considerably more than local speakers rehearsed. Nonetheless, the students involved these word-mixes uniquely in contrast to local speakers with regards to the decision of explicit collocations and their various frequencies, syntactic purposes, and down to earth capabilities.

Along these lines, Granger (1998b) has likewise played out a collocational investigation on cutting edge French English students' language creation. She stands out their composition from that of local speakers, zeroing in on two kinds of word-mix: collocations and formulae. She has concentrated on enhancers finishing off with "- ly" and working as modifiers. Her review shows that French students utilize less such collocations than their local speaker peers do and the majority of the collocations utilized by French students are

affected by their primary language. The discoveries in Granger's review (1998b) additionally uncover that French students' collocational abilities are genuinely restricted: they utilize not many local like collocations and numerous unnatural ones (1998b, p. 158). The general image of this collocational concentrate by Granger (1998b) is that French students tend to 'utilize intensifiers more as building blocks than as parts of pre-assembled segments', which makes sense of why collocational information on students isn't 'just powerless yet additionally somewhat misinformed' (1998b, pp. 151-152).

Comparative discoveries have likewise been gotten from Kaszubski's work (2000). He led a corpus-based investigation of 'sic' English center action words - "be", "do", "have", "make", "take" and "give" utilized in descriptive and contentious papers by Clean high level EFL understudies, EFL understudies from different nations and local English clients. The viewpoints of his review were from the colloquial, collocational, and free-combinational utilization of these 'sic' action words in English. His examination discoveries show that understudies utilize less informal articulations and collocations than local speakers and will quite often turn to elaborately dubious articulations albeit not mistaken 100% of the time. He additionally sees that EFL students, particularly progressed Shifts, have their collocational inclinations, which is inferable from L1 obstruction.

Rather than the L1 impedance, the accentuation of certain collocations in English guidance additionally brings about understudies' collocational inclination (Milton, 1998). Milton looks at the word arrangement in his review to distinguish the distinctions between Hong

Kong student English and that of local speakers. One of his discoveries shows that Hong Kong understudies utilize a lot more modest supply of word groupings than local speakers do. The greater part of the multi-word articulations regularly utilized by Hong Kong understudies are fixed lexical expressions, and a large number of these capabilities as sensible connectors, for example, "as a matter of some importance", "then again", as I would see it", and so on. He likewise sees that there are clear contrasts in the discourser jobs of the articulations utilized by local speakers and non-local speakers. Local speakers will generally try not to clarify and coordinate proclamations, while non-local speakers normally overstate their assertions by the utilization of escalating and downright articulations, for example, "obviously", and so on. (Milton, 1998).

It ought to be noticed that Chinese learners from Hong Kong and PRC have extensively unique learning settings. In Hong Kong, English as a mechanism of guidance in most optional schools during the 1960s and 1970s. Since the 1980s, blended mode educating (Cantonese and English) has taken the predominant situation in most auxiliary schools (Jhonson, 1998). In Central area China, nonetheless, English is shown exclusively as an unknown dialect. For by far most individuals, English is just utilized in the study hall setting. The high level English courses at colleges are frequently intended to give a quick development of jargon and presently more consideration is being paid to empowering understudies to communicate in English smoothly founded on the open way to deal with language educating. We will go to corpus-based investigations of PRC Chinese student English presently.

The four corpus-based investigations of collocations just examined have harmonious discoveries: EFL students in all actuality do involve collocations in their composition, yet less than local speakers and their utilization of collocations are enormously impacted by their native language. EFL students will generally utilize unnatural collocations, which are not unsatisfactory however provide perusers with an impression of non-local similarity.

Consequently, three significant classes of data are proposed for taking care of the issues EFL understudies face: a nitty gritty depiction of English pre-assembled language, the pre-assembled language in the students' first languages, and students' utilization of prefabs (Granger, 1998b). These three gatherings of data can give helpful data to the scholarly world and different partners which surely benefit EFL instructing and learning, and the additionally unequivocally show that customary CA and EA are adequately not to research student English.

2.3.6 Statistical formulas in Corpus Linguistics

For linguistics, raw corpora are an un-equal source of quantitative data. In this manner, language specialists frequently sum up their quantitative discoveries from corpora through measurements. The greater part of the examinations in corpus semantics work upon essential measurements which don't look for additional test is designated "Enlightening Measurements". These just depict the data somehow or another, for instance, the most essential measurement measure is recurrence count and straightforward estimations of cases of something that happens in a corpus. A particular sort token proportion is another

essential corpus insights. The token is the all out number of cases a word shows up in the text while 'type' is remarkable from - can perceive us how huge a scope of jargon is utilized in the text. Ordinarily, we decide the sort/token proportion by separating the quantity of types by the quantity of tokens and afterward get the rate by increasing with 100 (Templin, 1957). "This permits us to gauge jargon variety between corpora" (McEnery, 2012). For an alternate sort of corpora, a Normalized type-token proportion (STTR) is utilized as a normalizing rendition of the method. For the various sizes of data log TTR is applied for exactness rather than crude TTR. It tends to be determined by taking the log worth of the two kinds and tokens present in the message.

2.4 Literature Reviewed in Practicle Perspectives of the Research Work

As grammar is preferred to lexemes in EFL educating and learning. However, vocabulary is 'no longer a victim of discrimination' since the significance of lexis in SLA has been realized in recent decades (Laufer, 1997, p. 140). More consideration has been paid to collocation, especially, with regards to lexical investigations. Granger (1998) claims that since the last part of the 20th century the investigation of collocation procures a lot of significance EFL instructing and learning. As she notes:

In the first place, the development of the idea of lexico-sentence structure, enlivened by Halliday and Sinclair, has advanced the syntagmatic examination of lexis... ..besides, corpus semantics plays had a significant impact, giving language specialists computational means to reveal and explore lexical example... .. At last, pragmatics has turned into a

significant field of study, by its own doing, in phonetics and presently in EFL. Pragmatics ability has come to be seen as a fundamental piece of students' capability (Granger, 1998b).

Albeit, the collocational examination has turned into a great concentration in the EFL field yet not much exploration has been completed around here. In any case, the general picture that arises through past L2 collocation research is that collocation creation is a specific issue space in SLA, in any event, for students at a high level, contrasted and their better responsive collocation information (Marton, 1977; Gyllstad, 2005). Besides, EFL collocation studies show a collocation slack, where collocation information lingers a long ways behind the improvement of linguistic structure and lexis (Haiyan, 2015). In this part of the section, work is finished by Granger (1998), Rebecca (2000), Haiyan (2015), EserOrdem& Erdogan Bada (2016), Yuanwan Lu (2016) and Gablasova et al (2017), has been talked about in the accompanying passages.

Granger (1998) has explored the collocation information on French English students. She stands out their composition from that of the local speakers, zeroing in on two kinds of word mixes: collocations and formulae. Speakers finishing with "- ly" and working as modifiers are chosen for her collocational study. Her review shows that French students utilize less such collocations than their local speaker peers and the majority of the collocation they use is exceptionally affected by their native language. The finding in her review uncovers that French students' capability in collocational expertise is missing and restricted. The general image of Granger's (1998) concentrate on shows that the French

students tend to "use enhancers more as building blocks than as a piece of the pre-assembled segment", which makes sense of why students' capability of collocational information isn't just powerless yet in addition somewhat misinformed.

Rebecca (2000) has examined the lexical way of behaving of Taiwanese English students. She has concentrated on the collocations utilized by non-local English students and cases that EFL students frequently neglect to pick the right mix of at least two words because of the absence of information on collocational properties in jargon. She has handled two pre-labeled corpora: Taiwanese student corpus of English and BNC, to look at the students' utilization of collocation over a bunch of comparative words: 'huge, enormous, and fantastic'. The discoveries show the abuse and underuse of these words as collocations. The collocations of 'enormous' are essentially abused by the students when it is utilized to allude to extract ideas. This peculiarity on the other hand is researched and she presumes that most successive collocations in students' use will generally communicate obscure thoughts when more unambiguous significance ought to be imparted. In addition, students have the choice to involve those collocations in specific conditions where more succinct implications are liked. Further, she finds that mis-collocations, about 'huge, enormous and incredible' are utilized because of students' obstruction of L-1 and equivalent techniques, which they take on for lexical disentanglement (Shih, 2000).

Eser Ordem and Ordogan Bada (2016) have attempted to depict and dissect lexically started 'action word + thing' lexical collocation across composed scholarly class in Turkey, and

find whether there are prototypical 'action word + thing' collocation or developments across inside three sciences: physical, wellbeing and sociology and analyze these collocations regarding development language. Their review investigations the lexical collocation through recurrence based and chi-square examinations. It tracks down the peculiarities between the action words with their collocations. The corpus-driven results show rather more likenesses and the connection among wellbeing and actual sciences while; sociologies show a tremendous distinction from the other two. Then, at that point, the review demonstrates that comparative action words with their collocation across composed scholarly classes might be trailed by cutting edge EFL/ESL clients so they can make their scholastic composition and distributions advantageous (Ordem, and Bada, 2016).

Likewise, Yuanwen Lu (2016) presents an extensive examination report, 'A Corpus Investigation of Collocation in Chinese Student English'. The current review is a lot of impacted by the philosophy, he has utilized in his examination. He says "collocation is a significant apparatus in portraying lexical conduct in language and has gotten expanding consideration lately" (Lu, 2016). He has explored the lexical way of behaving of Chinese English language students by contrasting their scholastic works and the composed material made by local students out of English and America. The review analyzes two PoS labeled corpora to look into huge lexical and syntactic examples in the two corpora to see their disparities quantitatively and subjectively. It further makes sense of the elements of Chinese student English and examines critical ramifications emerging from the finding. Both lexical and linguistic collocations have been concentrated on in the chose corpora; their frequencies have been determined through the concordance device of scholar

programming, then, at that point, for the subjective examination Kjellmer's (1984) definition to the collocation is applied on Chinese students' composition. He finds Chinese students use collocation with extensively less assortment and their arrangements uncover their own particular syntactic and lexical collocation emphatically affected by Chinese culture and L-1. The review suggests that educators ought to plan exceptional exercises to show commonplace English collocation and bring issues to light of them so the students could have local live ability in language capability in EFL (Lu, 2016).

Dana Gablasova and her co-analysts in 2017 have broadened the investigation of collocation in the field of language learning research (LLR) by giving specific consideration to the examination of collocability across various corpora addressing various sorts, registers, and modalities. They raise different issues, include in the translation of collocational designs in the development of first language and second language. They have chosen three sorts of collocation: action word + supplement, modifier + thing, and intensifier + descriptor from the subcorpora of BNC. The strength of collocation is estimated through the three sizes of affiliation measures (AMs). The variety in the collocational strength among BNC and its subcorpora recommends that huge accumulated data, for example, BNC might conceal various appropriations of standard across registers and sorts as well as across composed and spoken.

The outcomes show that the connection among predictable and language delivered in various settings needs cautious consideration. Efficient variety in collocational designs

across classes, registers, and temperaments have suggestions for research that looks to reveal insight into the connection between etymology experience (openness) of language-clients and their collocational information (Gablasova, Brezina, and McEnery, 2017). The accompanying figure 2.3 shows the related impact of the past explores upon the current review.

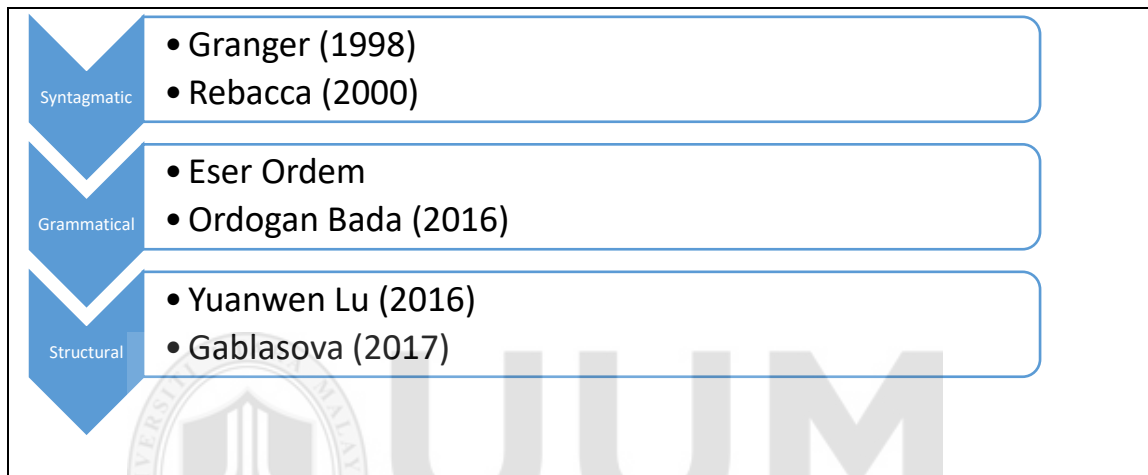


Figure 2.3. Corpus-based research studies on English collocation

2.4.1 Corpus-based Studies on Collocations in Pakistani Learners' English

Attributable to its English medium tutoring strategy people of Pakistan have seen a blast in the interest of English since the rise of the 21st century. Presumably, English has been nativized in Pakistan (Mehmood, 2009); the understudies are learning the English language from youth from the five star and arrive at the college level yet they can't communicate in English easily (Adil, 2018), however, they show their capability, to some extent, in written form. Many research studies have been completed to investigate the idea of the issues looked by Pakistani Students in learning English as EFL or ESL, for example, the research

work of Ahmad et al, 2011; Akram and Qureshi, 2012; Bilal et al, 2013; Haider, 2014; Jahan, & Nasrullah, 2016, and so on. Numerous scientists have dealt with the variety that exists in Pakistani English (Baumgardner, 1993; Mehmoob and Ahmar, 2004; Rehman, 2011; Uzair, 2012; Demir, 2016, and so forth). Aside from that, Couple of have attempted to demonstrate Pakistani English as a particular assortment of English and figuring out its imprint inside the circle of "new Englishes" (Mehmood, 2009; Bilal et al., 2011; Humaira, 2012; Parveen, 2013; Jameel, 2014; Jilani and Behzad, 2018, and so forth). Not very many exploration work has been finished to research collocation in the lexico-grammatic way of behaving of Pakistani English students. Notwithstanding, Hafeez (1988) has endeavored to dissect the conversational collocation blunders that Pakistani-Canadians make in ESL, a non-corpus-based examination. Afterward, Akhter et al (2017) have utilized a corpus-based study to deal with give English/Urdu collocational portrayal in light of contrastive examination to research same and non-identical collocation in the field of SLA in Pakistan.

Hafez (1988) has played out a microanalysis of inter-language (L-1) utilized by Pakistani learning English as a second language (ESL) on ten subjects' discourse tests to acquire bits of knowledge into their second language securing. He has demonstrated that second language students neglect to perceive collocation as fixed articulation and view each word inside them as autonomous and in this manner, replaceable. He has recorded the meetings of the members and afterward deciphers them to work out the recurrence and different qualities of collocations. The outcomes show that collocations are space explicit; the continuum found in this setting runs between more objective like collocation in work talk area/article movement and less in biography area/story action; and officially taught subject

show more objective like utilization of collocation as contrast with the casual taught ones (Mian, 1988).

Akhter, Behzad, and Qureshi (2017) have likewise dealt with collocations to introduce a substantial portrayal of English/Urdu collocations and accentuation their degree in educating and learning of English as a second language in Pakistan. They have utilized a corpus-based way to deal with portray English/Urdu collocation in light of contrastive examination to explore same and non-comparable collocations in corpora of million words from Pakistani papers. AntConc has been utilized to separate collocation from the data and afterward Oxford Collocation Word reference has been utilized to actually take a look at the collocations of Urdu proportionality. Lexical collocation, open-class word mixes have been dissected and found that ESL students have issues in the use of word blends yet it very well may be limited by the deliberate arrangement of native same or non-comparable use of collocations (Akhter, et al., 2017).

2.4.2 Previous studies on Data-Driven Learning

With the rise of the 21st century, the methodology of data-driven learning has acquired popularity. Many linguists and scientists have researched and demonstrated the value of the inductive methodology through DDL in the field of language learning which comprises of devices and procedures of corpus linguistics for educational purposes (Gilquin and Granger, 2010a). It gets the validness in the classromm as well as students can give them

countless true examples of a specific linguistic items out of class. This openness can add to lexical extension and other grammatical elements of the target language.

The greater part of DDL exercises include concordance or the same. Concordance has been introduced in different ways by different scientists. As Gilquin and Granger (2010) characterized concordance lines as far as shortened (supposed KWIC, keywords in context) or took the state of the total sentence, the entire concordance may be given to the students or simply a segment of it, the concordance lines may be altered or introduced in their unique structure or they may be displayed on screen or in printed structure. This KWIC (keyword in context) view ended up being centering for the novices with the events of search word-adjusted under one and different makes the example more apparent than the sentence view.

Boulton (2009) has detailed a trial with lower middle students where KWIC gave improved results than full sentence settings. He controlled the determination of the subset of concordance aiming to lessen the data into sensible amounts. This determination was performed in light of a few standards including, meaningfulness (the most troublesome concordance lines are disposed of, cf. Kuo at el, 2001), Recurrence (just the concordance lines delineating the most incessant use are kept, cf. Demand, 1990) and handiness (just those concordance lines that are judged helpful are kept, cf. Tribble 1997: 4). Control comprises of altering the concordance, especially, improving on them, Boulton (2009) isn't agreeable to such practices as this as per him sabotages the credibility of the material and

doesn't plan students for the real factors of the genuine language they are probably ready for. However, Nesselhauf's (2004: 143-44) conversation of a concordance of the action word recommend in the LOCNESS corpus (a corpus of factious expositions by English and American understudies) and the German subcorpus of the Global Corpus of Student English is a genuine illustration of why control is some of the time essential for DDL (particularly with novices):

“As the lines are now, they could be confusing for learners in many respects: there is at least one typographical error (*suggested than* instead of *that* in LOCNESS); one of the occurrences of *to* after *suggest* does not constitute wrong complementation (*could suggest to her two colleagues*); and, as *suggest + ing* only occurs in the learner but not in the native speaker corpus, learners might even conclude that this construction is not possible in English”.

(p. 14)

Researchers demonstrated the critical part of DDL in the improvement of bigger units of grammar like a paragraph. Sepehri (2015) explored the effect of some DDL-based units to a gathering of Iranian EFL students in the passage improvement course. He explored that the understudies taking DDL classes were in a superior situation in fostering their decisive information than their partner peers who were treated in non-DDL class. It showed that the foreign language students utilizing DDL-based materials could foster their insight into the unknown dialect components and rules at lexical and syntactic levels like words, articulations, and examples much of the time utilized in associating thoughts more than the non-DDL students utilizing the traditional course reading and the educator's class introductions. In addition, the subsequent seeing as worried about the procedural

information on students on the ground of creating composing abilities by and by, the outcomes showed that the members of DDL bunch further developed their 'language use' includes more than that of the non-DDL bunch. 'Language use' highlights by definition are phonetic developments, word request, and semantic exactness. This finding was upheld by the outcomes acquired from breaking down the 'Exactness' measures in the CAF (intricacy, precision, and familiarity) study. He accepts that one of the objectives of the course which is by all accounts the improvement of the students' miniature level abilities recorded as a hard copy has been met. He has depicted the mentality of understudies towards DDL overall and the DDL material arranged for this concentrate specifically. There existed an overall uplifting outlook of students towards DDL and the materials arranged for DDL activities. The understudies appeared to be happy with taking part in DDL-based class exercises, happy with the quantity of learned materials in a class meeting and think that the technique is new and creative. Nobody among the interviewees showed his absence of interest in being a piece of the DDL study hall nor they knew anyone among their friends who grumbled against this methodology. The respondents invited the books with enormous proportions of KWIC configuration to check specific linguistic features.

2.5 Conceptual Framework

Examining the effect of a modern technique while teaching English collocations, two groups of students, in this regard, was the main objective of the study. Here, I focused on the issues regarding development in the foreign language use, particularly, lexical knowledge in the form of correct use of collocation of the target language. Actually,

language development is meant the characteristics of learner's output that reveal some point or stage along the developmental continuum.

The PI has helped to conceptualize the hypothesis of the current research work which consists of the adaptation of modern technique like CDLL in the foreign language teaching and learning. A significant gap was identified in terms of lack of studies in the research area of the foreign language teaching and learning that paved the way for current research investigation and implications.

The conceptual framework shows the placement of this particular study within the perspective of ELT methods and techniques aligned with the DDL definitions of Tim John (1991) further endorsed by Granger(1998), Baba (2000), Boulton (2012), Smidh Flowerdew (2015), Sepehri (2015) and Lu (2016) based on the zone of proximal development (ZPD); the gap between what is known and what can be known in accordance with socio cultural theory propounded by Vygostky. The work of these theorists helped to chalkout the conceptual framework of the research work undertaken by me.

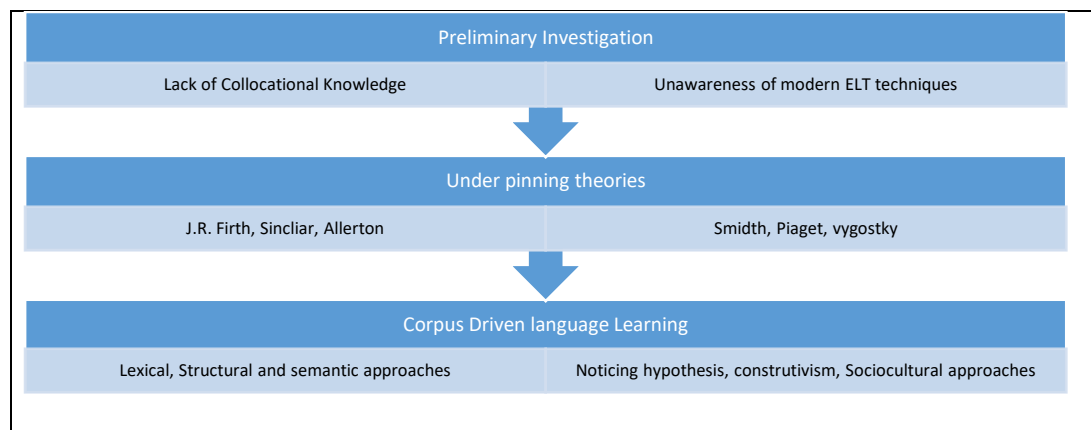


Figure 2.4. Conceptual framework of the research study

It was conceptualized that if the EFL learners would be provided with scaffolding to elicit their knowledge in a procedural way they could learn better. Obviously, they were undergone the process of pattern hunting and then pattern defining to reachout the objectives of language learning in an independent atmosphere. They would themselves develop corpora and then looked for the particular cunks of the language like collocation in concordandane lines then they found the appropriateness of these lexical items with a standard source. Ultimately, they could have a durable knowledge about that particular language item.

2.6 Summary

Since the emergence of the 21st century, linguists and academicians have laid much emphasis on the importance of multi-word expressions (collocation) while determining the lexical behavior of foreign language learners. Linguists all over the world have proved that the knowledge of how to combine words into chunks inefficient language use is imperative. The most common approach underpinning the investigation of these chunks is a structural approach in which the learner master the patterns involved in a sentence. Structures are the different arrangement of words in one accepted style or the other that leads to the lexical association rather than the spelling and grammatical structure of vocabulary. A paucity of collocational proficiency is associated with lexical proficiency i-e collocations tie with the lexical development. Hence, this chapter has viewed the detailed description of the theories and models which help to conceptualize the hypothesis of the research study. Moreover, corpus-based investigation of the lexical patterns has revolutionized the world of lexicography and language-learning behavior. Several corpus-based types of research have

been viewed that have paved the way for the formulation of the methodology of the research study and highlight the gap in this particular area of foreign language learning research. The different methods of analysis have also been reviewed that are incorporated to carry on the mixed-method approach of the present study. This chapter deals with the elaboration of how other researchers and theorists have tried to work in the relevant areas of the topic under investigation of the current study.



CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter explains the overall methodology undertaken in this research study. It starts with the paradigm and philosophical stance which informed the research design in section 3.2. The characteristics of the subject have been described in section 3.3. It elaborates the research sample and the variables considered for the data collection of the study given the corpus-based selection of data. A detailed account of the treatment of the experimental group regarding awareness of corpus driven language learning (CDLL) is discussed in section 3.4. Then the next section of the chapter, 3.5 is allocated for the data collection methods relevant to the research hypothesis. Later, it describes the detailed framework of the study how data was retrieved for quantitative as well as qualitative purposes with the Explanatory Sequential design of a mixed-method (Cresswell, 2014) for the authenticity of validity and reliability of the study. Subsequently, the implications from piloting was also referred to the next segment. The last portion encompasses the summary of the whole research methodology involved in the present study.

3.2 Philosophy of the Research

The philosophical stance of the present study is very much influenced by the post-positivistic school of thought in which causes determine the effects and outcomes (Cresswell, 2014).

“The knowledge sought through the lens of post-positivism is based on careful observation and measurement of objective reality that exists “out there” in the world”.

(Cresswell, 2014:p.7)

Post-positivists believe that they can develop a regular relationship between the social phenomena by using theory to generate hypothesis which can then be tested by direct observation.

The current study holds that language is an artefact that is jointly created by its speakers and that reality is similarly created through language learning. It doesn't go against the research's postpositivist epistemological stance or foundationalist ontological perspective. As stated in the previous explanation, reality (language) exists, but only after conscious meaning-making beings (learners) have attempted to make sense of it. Therefore, collocational knowledge is a type of reality that actually exists and occurs in the real world. Instead of attempting to understand how reality first came to be, this research aims to analyse and explain how it arises under certain circumstances. As a result, the mixed methodology has been used in this study. Post-positivist work must demonstrate knowledge empirically by objective modification and control of factors (Boulton, 2017). The investigations are concerned with the causal inferences which will be discussed in the following portion of this chapter.

3.3 Research Method

The design of the present study is mainly chalked out according to the explanatory sequential mixed method of research design (Cresswell, 2014). The present study explored the effect of CDLL in the teaching of lexical collocations to Pakistani non-native ungraduated university students. The subject of the study was tested earlier to find their level of English; then, the sample of study was bifurcated into experimental and controlled group. The experimental group was treated for a complete semester under the hypothesis of the study. Two types of corpora were compiled to investigate seven types of lexical collocation which were automatically extracted from corpora by the software AntConc. The lexical combination (collocations) was described by the sub-groups of the experimental group to have a deep understanding while comparing these unusual patterns with BNC. A pre-test comprising of 25 items, for evaluation of the English collocational knowledge of the selected sample, was formulated and distributed among the participants. The sepc-driven test items (selection from the tests we have already seen or from pilot work) were composed while following the series of guidelines by (Kehoe, 1995) and originated from the idea of “Critical Reverse Engineering (RE)” coined by Davidson and Lynch (2002). Data was collected quantitatively and operated upon inferential statistics to measure the P-value of results from experimental group and controlled group.

Later, the experimental group was treated with CDLL exploiting the participants’ dominant intelligences, namely visual-special and linguistic-verbal intelligences (Gardner, 1983), for a complete semester of 16 weeks and more attention was focused on the study of

collocation. Whereas, the controlled group was treated with the traditional way of teaching. After a complete semester post test was conducted to check the results of the treatment.

Aftermath, for qualitative data, a focused interview questions (Appendix-M) was mailed to the participants of the selected sample to seek their opinion regarding the effectiveness of the CDL; the obtained data were analyzed through Nvivo-12 software to evaluate the level of perception of the CDL approach in teaching/learning.

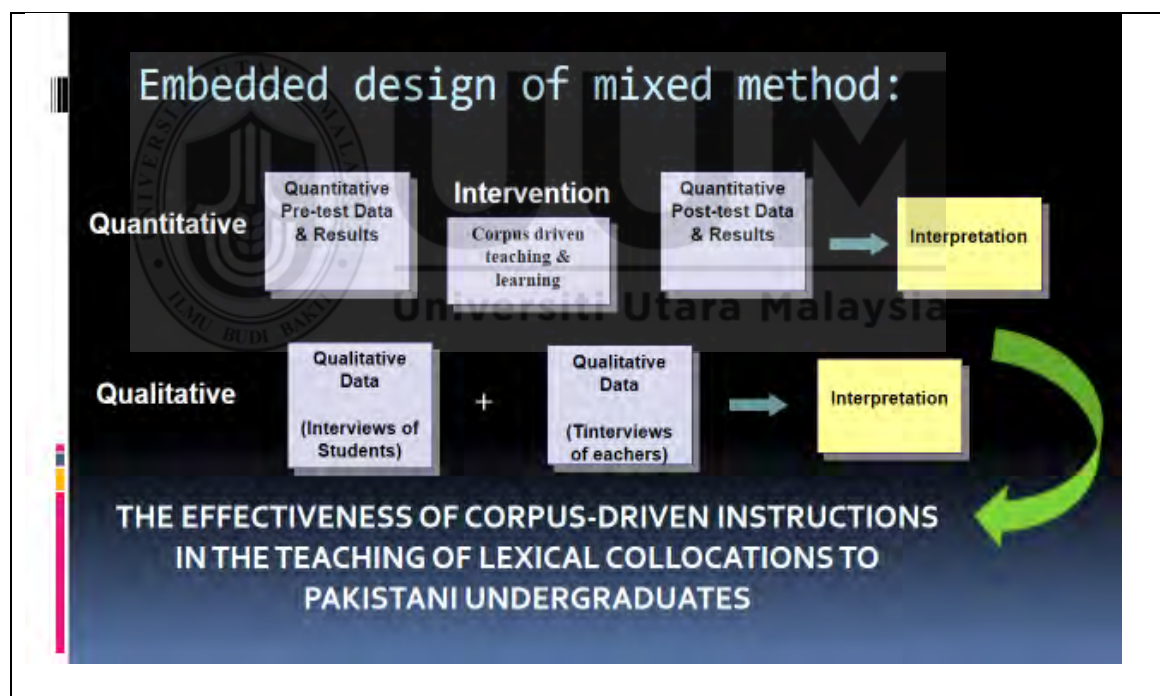


Figure 3.1. Research design

3.4 Population

The investigation of the study was generalized for the Pakistani non-native EFL university undergraduates and the teachers teaching English language courses at the university level. As per Higher Education Commission Pakistan statistics (2019) there are 174 universities in the country including 1194185 undergraduates who are studying English language courses compulsorily among the other courses of their degree programs. 11246 faculty members are serving in the English departments of the universities.

3.4.1 Sampling

The present study used a non-probability purposive method for sampling (Finfgeld-Connett, Finfgeld-Connett, & Johnson, 2018); this sampling technique is most effective when the researcher needs to study a certain cultural domain with knowledgeable experts within (Tongco, 2007), it was convenient for the researcher to rely on his own judgement. Being a researcher of Applied Linguistics, I selected Government College University Faisalabad Pakistan undergraduates of BS English (Literature & Linguistics) degree program as a frame of sampling. During the research investigation, there were 31518 students enrolled in undergraduate degree programs being guided and taught by 1068 full-time and part-time faculty members.

100 undergraduates of 3rd semester BS English (Literature & Linguistics) morning & evening session 2018-22, and 11 faculty members from the Department of Applied Linguistics G C University Faisalabad were selected as the sample of the study. As the

selected students and faculty members of the department of Applied Linguistics, belong to various regional and ethnic groups of the country that helped to generalize the results so it would be convenient for me to approach them easily keeping in view of time and cost factor. The sample is purely a part of English literature and linguistics where English language teaching and learning practices are common. Moreover, the pandemic situation prevailed due to COVID-19 forced me to select the sample as I had an easy access to the selected sample.

3.4.2 Participants of the Research Study

Being the faculty member of the Department of Applied Linguistics GC University Faisalabad Pakistan, the researcher himself is the part of the research study. The sample for experimentation will be selected from the students of the same department. The non-native Pakistani 100 undergraduates of 3rd semester BS English (Literature & Linguistics) for the session 2018-22 were selected as the sample of the study. The morning group of 50 undergraduates was treated as experimental group and evening group of 50 undergraduates of the same level of study were treated as a controlled group for the present study.

As far as the ethical requirements of the research were concerned, all the participants were asked to give their consent to the researcher to use their uploaded material as the required data for corpus compilation to be used for the research study. A consent form was prepared in which the participants were familiarized with the research aims and data collection procedures.

3.4.4 Comparison between both Groups of Research Sample

The participants of both morning and evening groups of the same session (2018-22) were not randomly assigned to the experimental and controlled groups, they seem to be comparable in many aspects.

- Both groups belonged to the same department of the university.
- Both had the same educational level that is the 3rd semester of the same session (2018-22) degree program.
- Both had taken the same courses introduced by the university
- Both had the same age level
- The information about their English learning background, learning strategies preferences, familiarity with different language learning techniques and skills were sought through a questionnaire (see appendix B) that could help to determine to what extent the participant groups are comparable.

3.4.4.1 Language Proficiency and Exposure to the English language of the Sample

The selected sample was possessing almost the same proficiency level in English. They were the undergraduates of public university in the region. All of the participants had never been in a foreign country at all and they never used English as a second language for their general communication. All had gone through various courses relevant to the language studies in previous semesters of degree program BS English (Literature & Linguistics) as shown in the following table 3.1 and 3.2 before they were selected as the sample of the study.

Table 3.1

Courses Outline of 1st Semester BS English (Literature & Linguistics)

Sr. no	Code	Title	Credit hr
1	LNG-301	Introduction to Linguistics	3(3-0)
2	LNG-302	Study Skills	3(3-0)
3	LNG-303	Sentence Analysis/Grammar	3(3-0)
4	LNG-304	Introduction to Literary Studies	3(3-0)
5	LNG-305	English For Academic Purposes	3(3-0)
6	CSI-321	Introduction to Computing Applications	3(3-0)
	Total		18

The following table 3.2 shows the course outline for second semester of the selected sample:

Table 3.2

Courses Outline 2nd Semester BS English (Literature & Linguistics)

Sr. no	Code	Title	Credit hr
1	LNG-306	Life Skills	3(3-0)
2	LNG-307	Phonetics & Phonology	3(3-0)
3	LNG-308	Poetry-1	3(3-0)
4	LNG-309	History of English Literature	3(3-0)
5	SOC-307	Introduction to Sociology	3(3-0)
6	PST-321	Pakistan Studies	2(2-0)
	Total		17

The selected sample went through the above-cited scheme of study during the previous semester and they were studying in the third semester since September 2019 under my treatment. The following courses were introduced in the third semester as shown in table 3.3:

Table 3.3

Course Outline of 3rd Semester BS English (Literature & Linguistics)

Sr. no	Code	Title	Credit hr
1	LNG-401	Sociolinguistics	3(3-0)
2	LNG-402	Advance English Grammar	3(3-0)
3	LNG-403	ELT method and Approach	3(3-0)
4	LNG-404	Drama-1	3(3-0)
5	STA-321	Introduction to Statistical theory	3(3-0)
6	ISL-321	Islamic Studies/Ethics	2(2-0)
	Total		17

However, a summated rating method (Binkai, 2012) aquestionnaire along with a Likert scale was administrated to measure the proficiency and level of English among the participants (see appendix-B). The results were processed through statistical operations to check the comparability of the both groups of the study. The following results show the comparability of both groups selected for the research study:

- I. The first two items of the questionnaire were about general information regarding the biological and academic information which seemed to be relevant in the context of comparability

- II. While answering the third item most of the participants answered the same option that was about why they learn English; the participants of both groups answered that “Because it helps me to get a good and well-paid job”. A few participants tick the other options.
- III. Item number 4 and 5 were about the foreign stay of the participants who mostly answered “never” only two participants each from either group claimed once they visited abroad for a few months
- IV. Item number 6 was all about their interest and level in English which was measured through Likert scale that was later processed through excel sheet to find the percentage of both results as shown in the following table 3.4:

Table 3.4

Tabulation of Interest Level of Experimental Group in the English Language

	Count	Very Often	Often	Some time	Rarely	Never	
Respondants	N	%	%	%	%	%	total
Q 1	30	0%	7%	33%	37%	23%	100%
Q 2	30	0%	0%	0%	47%	53%	100%
Q 3	30	0%	10%	27%	53%	10%	100%
Q 4	30	0%	7%	47%	27%	20%	100%
Q 5	30	10%	63%	17%	0%	10%	100%

The graphical representation of the responses of the participants about first segment of question set is shown in the following figure 3.2:

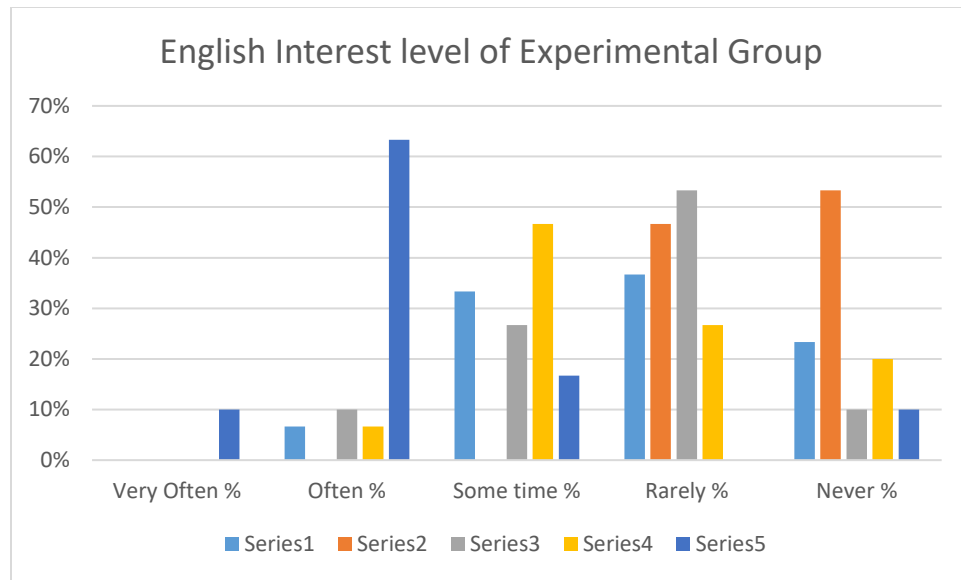


Figure 3.2. English interest level of experimental group

Moreover, the second segment of the question set regarding the interest level of the control group in English is presented with tabulation as shown in the table 3.5.

Table 3.5

The English Interest Level of Control Group

Respondant	Count	Very Often	Often	Some time	Rarely	Never	
s	N	%	%	%	%	%	total
Q 1	30	0%	7%	27%	33%	33%	100%
Q 2	30	0%	0%	0%	50%	50%	100%
Q 3	30	0%	13%	23%	57%	7%	100%
Q 4	30	0%	7%	43%	33%	17%	100%
Q 5	30	10%	60%	17%	7%	7%	100%

The graphic presentation of the interest level of the participants in English is shown in the following figure 3.3:

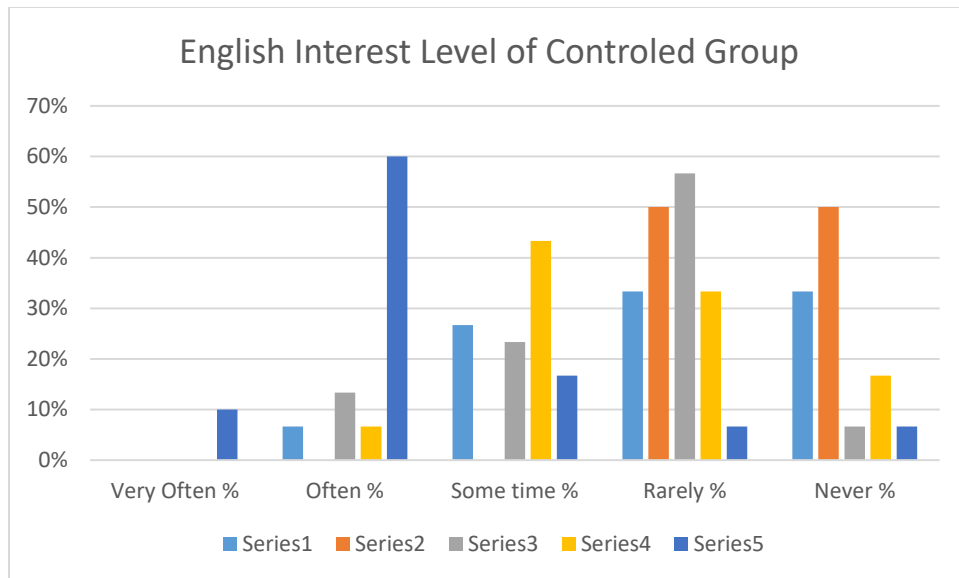


Figure 3.3. English interest level of control group

- V. Item numbers 7 and 8 of the questionnaire deal with the English courses the participants took outside the classroom the answer of this question is almost same that the both groups had not taken any English language courses outside the classroom other than a semester. Only one participant claimed that he did ILETS other than the university courses.
- VI. Item number 9 dealt with the English learning process in the classroom. The results of this Likert scale were also analyzed from the below-mentioned tabulation on the excel sheet. It also depicted the relevance of comparability of the both groups.

The following table 3.6 shows the percentage of the responses of the experimental group regarding the class-room activities about English learning (How often the The teacher e

engages the learner in language learning activities as mention in the Appendix-B). The responses show the meagre involvement of the teacher regarding the foreign language usage in class-room.

Table 3.6

Class room Activity about English Learning of Experimental Group

		Very	Often	Some time	Rarely	Never	
Respondants	Total	Often %	%	%	%	%	total2
Q 1	30	0%	7%	33%	37%	23%	100%
Q 2	30	0%	7%	30%	53%	7%	97%
Q 3	30	0%	10%	24%	55%	10%	100%
Q 4	30	0%	7%	47%	27%	20%	100%
Q 5	30	10%	63%	17%	0%	10%	100%
Q 6	30	3%	7%	10%	40%	40%	100%
Q7	30	13%	40%	30%	3%	13%	100%

The following figure 3.4 presents the graphic exposure of the class-room activities about the foreign language teaching/learning in the class-room

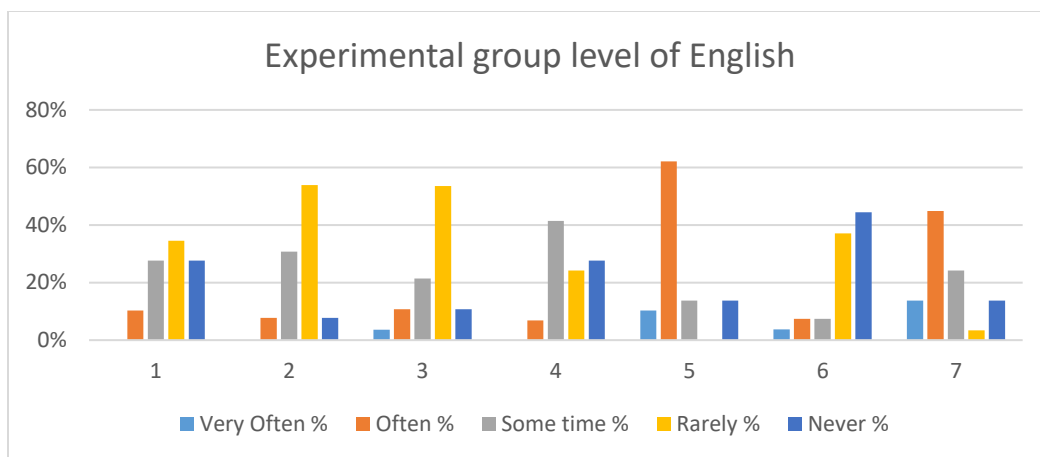


Figure 3.4. Result of the classroom activity of experimental group

Similarly, the following table 3.7 shows the results of the responses obtained from the control group about the class-room activities of English language learning:

Table 3.7

Tabulation of the Results of Control Group English Learning Activities

		Very Often	Often	Some time	Rarely	Never	
Respondants	Total	%	%	%	%	%	total2
Q 1	30	0%	8%	31%	35%	24%	100%
Q 2	30	0%	7%	30%	53%	7%	97%
Q 3	30	0%	10%	24%	56%	10%	100%
Q 4	30	0%	7%	47%	27%	20%	100%
Q 5	30	10%	63%	17%	0%	10%	100%
Q 6	30	3%	7%	11%	40%	40%	100%
Q7	30	13%	40%	30%	3%	14%	100%

The figurative representation of the responses about the English language learning activities of the control group of the study is shown in the following figure 3.5.

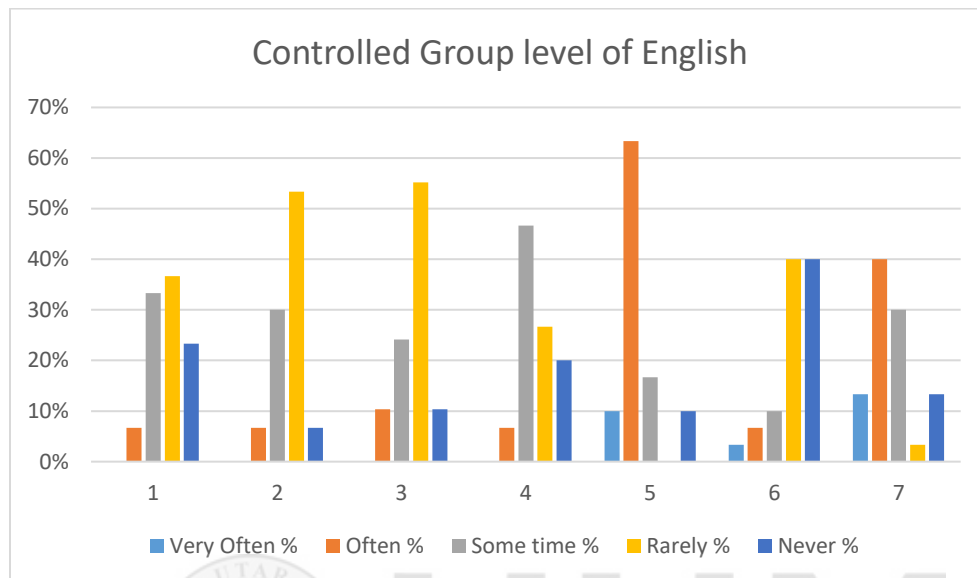


Figure 3.5. Result of classroom activity of control group

The results of question number 9 of the question-set (Appendix-B) administrated upon both groups showed a reliable comparison between both experimental and controlled groups for further investigation of the study. The responses of the participants of the both experimental and control groups show their equivilqance in English language learning and teaching,

3.4.5 Awareness of Corpus Driven language Learning

The selected sample of 100 student of degree program BS English (Literature & linguistics) in Government College University Faisalabad Pakistan during fall semester 2019 were

introduced CDLL in their course ELT Method & Approaches. The course out-line along with the objectives of the course is given in appendix-A. The experimental group, comprising of 50 students were introduced to the historical perspectives of English Language Teaching. Then, they were taught traditional approaches and methods of English Language Teaching as proposed by Jack C. Richards and Theodore S. Rodgers (1999) up to mid-semester exams which were conducted in November 2019. After mid-exam in the 9th week of the course, the selected sample was introduced corpus Linguistics and Data-Driven Learning Approach in foreign language teaching/learning. Whereas the controlled group of 50 students was introduced the traditional ways of the development of the foreign language at lexical, morphological, and syntactical levels as discussed by Rodgers (1999).

During the 10th week of the semester, the experimental group was taught theories as discussed by Flowerdew (2016) including Noticing hypothesis, Constructivist theory, and Sociocultural theory. The students were introduced various software like Quizlet, COCA, and concordance in the 11th week. During the following weeks of the semesters, the role of DDL in the development of vocabulary, syntax, and EAP was discussed respectively in detail. During the last week of the semesters, the students were asked to revise the same ideas in the form of a presentation. In the final assessment of the semester, the student was evaluated regarding their knowledge about DDL in written form of paper.

During the fourth semester, The morning group of 50 students who were treated as an experimental group. The group was taught a course of Computer Assisted Language

Learning (CALL) in their degree program by one faculty member; whereas. The controlled group was taught by the other teacher. The course outline for both groups has been mentioned in the appendix-A. The experimental group was facilitated with a computer furnished language lab. They were, particularly, motivated to prepare adhoc corpora for autonomous investigation of linguistic features through CDL (corpus driven learning).

3.5 Description of data

This is a typical type of language learning research (LLR) focusing on the lexical behavior of foreign language learners. The purpose of this study is to explore the experimental treatment upon the lexical behavior of Pakistani English language learners through corpus-driven learning (CDL) of collocations. To investigate the lexical behavior of the subject under investigation an explanatory sequential design of mixed-method research (Cresswell, 2014) is adopted. The detail of the selected material for the study is described in the following sections

3.5.1 Material for Study

To collect data, a prerequisite for the research study, the following types of materials were prepared to undertake and used appropriately while collecting data relevant to the study.

- Teaching material: the revision of course outlines for ELT Method & Approaches and CALL. (Appendex A)
- Set of a questions based on Kehoe's guidelines (1995) to check the comparability of the selected sample in the context of English proficiency level. (Appendix B)

- Compilation of corpora from editorials of leading newspapers of the Britain print media and Pakistani print media. The activities were based on participants' dominant intelligences, namely visual-special and linguistic-verbal intelligences proposed by Gardner
- preparation of spec-driven items relevant to the research hypothesis in the pretest (Davidson & Lynch, 2002). (Appendix C)
- Preparation of spec-driven items relevant to the research hypothesis in the post-test. (Appendix D)
- Development of open-ended interview questions probing into the effectiveness of CDLL on the selected sample. (Appendix K)
- Development of semi-structured interview questions to seek the opinion of teachers regarding the effectiveness of CDLL in English language teaching. (Appendix M)

Being the faculty member of the university, the course outline of the course LNG-405 (ELT Method & Approaches) was revised as shown in the Appendix A. Topics of data-driven learning were particularly included in the scheme of study. The students were intentionally acquainted with the basic concepts of DDL along with the underpinning theories which had shaped the idea of DDL in language learning. Apart from that, the course of CALL (Computer Assisted language Learning) was also revised (see Appendix A) keeping in view of the needs of the selected sample to be able to use various software for language investigations.

To measure the attitude of the selected sample towards the English language, the summated rating method was followed on Lickert scale (Jonas, 1992) to develop a prerequisite

questionnaire was developed and distributed among the selected undergraduates of the Government College University Faisalabad Pakistan. The results were then statistically analyzed to measure the comparability of both groups.

After making the students aware of the DDLL, they were directed to prepare ad-hoc corpora individually. Every student was provided with a computer to work on it. Later, he/she maintained his /her corpus of the selected genre. Then they are allowed to work on the collocation autonomously. At the end of the 4th semester, a post-test was administrated upon the selected sample of the study to check the hypothesis regarding the effectiveness of the CDL on the development of English collocational knowledge.

To revalidate the results, open-ended interview questions were developed and circulated among the selected sample of the experimental group to probe into their attitude towards corpus-driven learning.

3.6 Data Collection

The data collected for the present study was mainly comprised of the responses obtained from the questions administrated upon the selected sample of the study. A typical self completion set of questions with most items either asking about information or seeking opinion with the serious response from respondents is a highly structured data collection instrument (Dörnyei, 2003). Two types of spek-driven questions sets were composed for the present study: closed-ended formate for quantitative data analysis and open-ended

formate for qualitative data collection and then analysis. Keeping because of the research questions, the data for the study was collected in four phases.

3.6.1 Research Instruments for Quantitative Data

Earlier, a pre-test comprising of 25 sepc-driven items (selected from the already chosen items) were composed while following the series of guidelines by Kehoe (1995) and originated from the idea of “Critical Reverse Engineering (RE)” coined by Davidson and Lynch (2002), which evaluated the knowledge of lexical collocations of the selected sample (see Appendix C), was administrated upon both groups and the results were operated upon descriptive statistics on excel sheet as shown in the following table 3.8. the average of the test scores of experimental group is 13.95 and controlled group is 14.03. It means that no significant difference existed between the results of both groups' pre-test.

Table 3.8

Results of Pre-test

Participants	Number of	Number of		Aerage	SD
	responses	Blanks	Total		
Experimental group	40.00	10.00	50.00	13.95	3.99
controlled Group	39.00	11.00	50.00	14.03	3.34

The experimental as well as control group have almost the same competence about the knowledge and usage of collocation before the experimental treatment.

3.6.1.1 Experimental Treatment

In the beginning of experimental treatment adhoc corpora which was compiled by the selected sample are described as follow:

- COPLNE: Corpus of Pakistan Leading English Newspaper Editorials composed by the undergraduates of BS 4th Morning (Applied Linguistics GCUF) before experimentation.
- COBLENE: Corpus of Britain Leading English Newspaper Editorials, compiled by the undergraduates of BS 4th Evening (Applied Linguistics GCUF) before experimentation.

3.6.1.2 Statistical Information of data

The statistical information of COPLNE and COBLENE was obtained by the wordlist tool of Antconc software. This tool provided us the total number of tokens and types from both corpora as shown in table 3.9. Although COPLNE & COBLENE might not be too large corpora to examine the overall lexical behavior of language learners yet it was large enough to test the hypothesis regarding the importance of CDL in language learning. As Granger (1998) claims ‘one cannot imagine expecting learner corpora to reach the gigantic size of native corpora’, as it is difficult to collect learners’ data. The figure of normalized type & token ratio was listed to show their complexity in both of corpora which later proceeded on the famous Log TTR formula to maintain the accuracy as shown in table 3.9.

Table 3.9

Statics of the corpora Compiled by the Selected Sample

<u>Corpus</u>	<u>Word type</u>	<u>Word token</u>	<u>Ratio @ 10000 words</u>
COBLENE	16628	288996	575.37
COPLNE	16283	212369	766.73

After the compilation of data, the experimental group of the selected sample was further divided into 7 subgroups and asked to do the pattern-hunting of lexical collocations patterns as mentioned in the following table 3.10. Each group was assigned to find their respective collocational feature from the corpora compiled by them under the supervision of the teacher.

Table 3.10

Lexical Collocations

Sr No	lexical pattern	examples
1	<i>Verb + Noun</i>	<i>(make a mistake)</i>
2	<i>Verb + Adverb/Adjective</i>	<i>(to show clearly,)</i>
3	<i>Noun + Verb</i>	<i>(results proved)</i>
4	<i>Noun + Noun</i>	<i>(a ceasefire agreement)</i>
5	<i>Adjective + Noun</i>	<i>(heavy traffic)</i>
6	<i>Adverb + Adjective</i>	<i>(extremely generous)</i>
7	<i>Adverb + verb</i>	<i>(simply show)</i>

Each group had to find the frequency and distribution of the specific collocation which was later presented by them among the other fellows. They derived the inappropriate collocations and focused on the features of these misusages. The experimental group was

taught the models propounded by Sinclair and Allerton for collocational restrictions (Lu, 2016), The experimental group did pattern-hunting according to the assigned task then they did pattern-defining and shared their work among other participants. Whereas, the control group used the traditional methods by the choice of teacher to learn about lexical collocations..

After the treatment of one semester (16 weeks), data was again collected in the form of a post-test from the selected sample the statistical information of the data is given in table 3.10. The analysis of the results derived from both tests has been discussed in the next chapter of the study. The p-value of the t-test showed a significant different between the results of experimental and controlled groups.

Table 3.10

Statistics of the Results of Post-test

Groups	Number of Responses	Number of Blanks	Total	Mean	SD	T-test
						P-value
Controlled	46	4	50	13.67	2.74	
Experimental	49	1	50	22.41	2.09	1.2923E-29

3.6.1.3 Comparability of COPLNE & COBLENE

Both of the corpora: COPLNE & COBLENE – were comparable in terms of the genre of the newspaper editorials and almost it was the voice of mature authors who criticized the socio-political and socio-economic affairs of the state. Although the topics in both corpora are different yet the quantitative analysis of lexical collocation such as “Noun+Noun”,

Noun+Verb” etc. was not affected significantly by the frequency of occurrence of specific words, for it is reasonable to assume no matter what topics are written about, these patterns are essentially used by the language users.

3.6.2 Research Instruments for Qualitative Data

To go into the in-depth study of the research a qualitative approach was followed (Costa et al., 2018). To answer the second question of the research and while listening to the voice of the participants regarding the effectiveness of CDL, the students from the experimental group participated in a focused interview questions were asked at the end of the semester. These were the participants who volunteered to be interviewed; they were contacted by e-mail and provided with a certain time duration to think about the questions in advance. The questions were organized keeping in view of the familiarity with the participants as shown in Appendix-F. The thematic analysis was then be conducted through Nvivo-12 software and discussed in the following chapter.

Similarly, to answer of the third question of the research study, the opinion of faculty of the department of Applied Linguistics GCUF regarding the effectiveness of CDL in English language learning was sought through focused interviews which were conducted among 11 teachers of the Department of Applied Linguistics GC University Faisalabad Pakistan. These faculty members belonged to the various region of the country and they also obtained their degrees from the various universities of the country. The interviews were recorded and then analyzed later which was described in the following chapter 4.

However, the summarized form of data collection has been presented in the following figure 3.6.

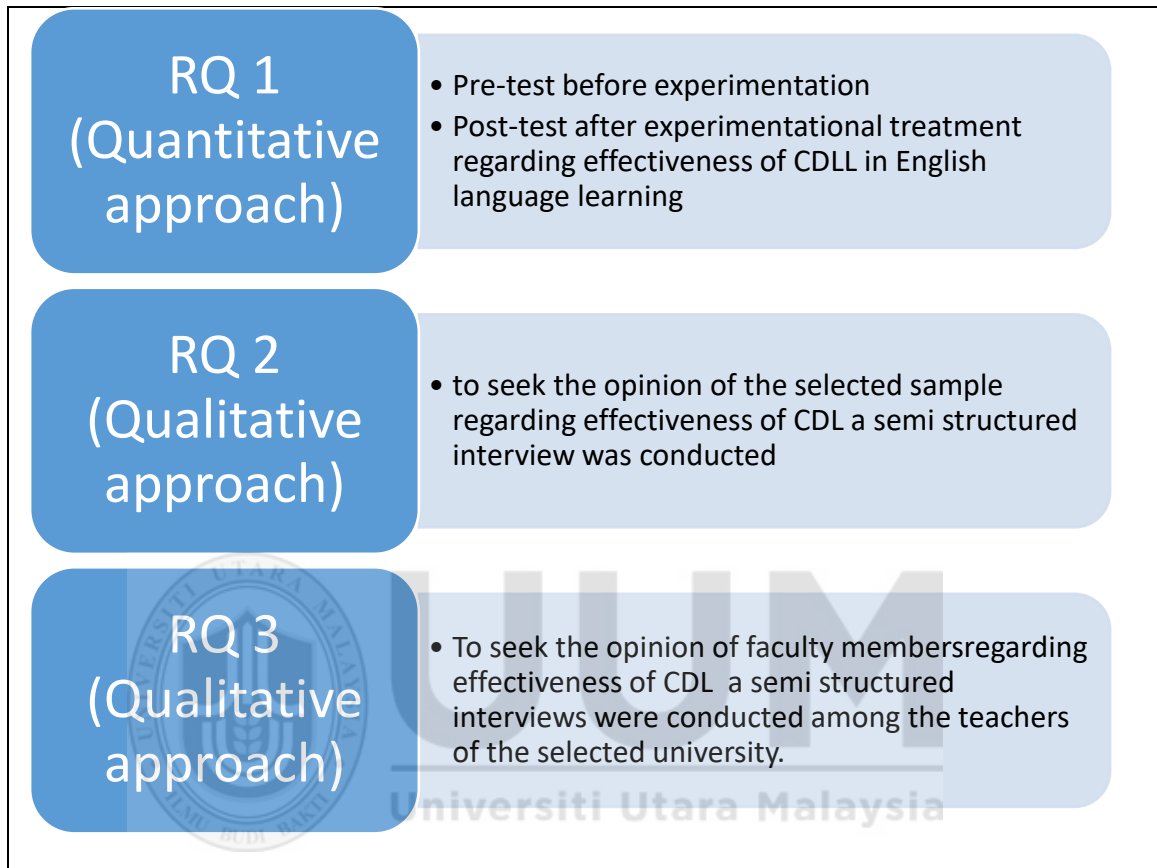


Figure 3.6. Summary of data collection for the research design

3.7 Tools for Data Processing

In this section, we turn to the tools used to collect and then manipulate the data for the present study: the PoS (parts of speech) tagger and tools of the Antconc software. First, the raw data was converted into HTML or text file then proceeded to tagging process.

3.7.1 PoS (parts of speech) tagger: CLAWS4

PoS tagger is a software device used to tag each character (lexicon) in the text with a PoS marker. This process is called PoS tagging which is developed at the University of Lancaster that was used in the present study to tag both corpora under investigation. It is a hybrid tagger based on both probabilistic and rule-based approaches. The probabilistic approach means that it selects ‘a preferred tag for a word by calculating the most likely tag in the context of the word and its immediate neighbors’ (Garside and Smith, 1997, p. 102). The rule-based enables the tagger to carry out exceptional tagging for multi-words as “family burden” a “Noun+Noun” combination rather than “Adjective+Noun”. The combination of a probabilistic approach with rule-based will lead to the maximum rate of accuracy in tagging the general text (Garside & Smith, 1997). A list of C5 tagset has been mentioned in Appendix B. For the current study both corpora: COPLNE & COBLENE were tagged by CLAWS4 with the PoS marker in tagger C5. In this way, each word in both of the data was tagged with the relevant ‘part of speech’ tagger for example:

The_AT0 whole_AJ0 book_NN1 is_VBZ a_AT0 criticism_NN1 of_PRF
l'optimisme_NN1

3.7.2 Accuracy rate of Tagged Data

To check the accuracy rate, two tagged passages from tagged COPLNE & COBLENE were randomly selected. In this way, the problem in tagging was detected and then solved manually if such a tagger appears in the whole text because it may be problematic for auto tagger to detect the discourse function of a word in the text.

3.7.2.1 Why Tagging is necessary

Tagging of corpora is necessary because it has the following advantages:

- Lexical patterns can easily and automatically be extracted. As far as the present study is concerned, seven lexical patterns such as “Noun+Noun”, “Noun+Verb” and “Adjective+Noun” etc were automatically obtained through the n-gram tool on Antconc software.
- Much more information is readily extracted within a few moments. As it has been referred above that each word in the data has its grammatical mark, hence, direct information regarding its lexical position can be readily obtained.
- Growing awareness o the importance of PoS tagger and tagged data in the emerging linguistics. Ooi (2001) argues that grammatical patterns can be easily extracted by PoS tagging, and ‘such annotation has been undervalued as a pedagogical tool until now’.

3.7.3 AntConc tools

AntConc 3.4.1 is a software tool that was used for the present study. It is a freeware, multi-platform, multi-purposes corpus analysis toolkit, designed specifically for the usage in the classroom (Anthony, 2004). It includes a comprehensive set of tools including powerful concordance, word and keyword generator, tools for the lexical bundle and cluster analysis as well as word distribution plot. The major tools used in this research study are being discussed in the following sections.

3.7.3.1 n-gram/cluster tool

The focus point of the study ‘collocations’ were investigated while using clusters/n-grama tool of the software. This tool displays the combination of words (collocations) and phrases in an arranged form according to frequency or alphabetically. The desired term can be specified in the inquiry box as a word, phrase, or substring with the specification of the number of words to left or right. The frequency threshold can also be set for the cluster generator as minimum frequency 3 has been fixed to have the collocations or present study.

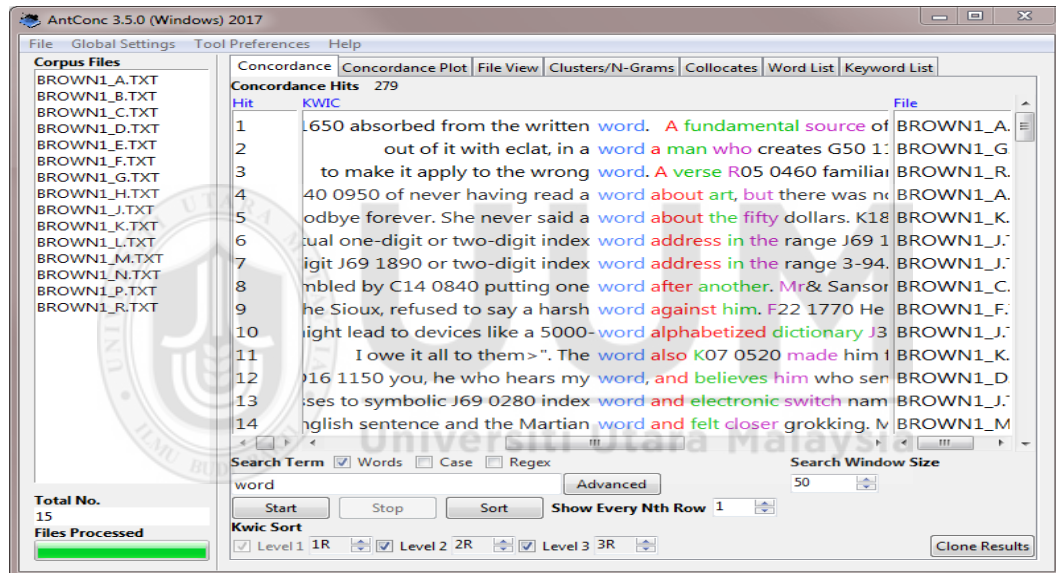


Figure 3.7. AntConc software

3.7.3.2 Concordance

The most used tool in corpus studies is a concordance. As Sun and Wang (2003) describe, concordance has proved an effective aid in the acquisition of foreign language, facilitating the learning of vocabulary, collocation, grammar, and syntax. In the present study, Concordance was applied to find the number of hits of the particular combination in the

tagged data for the comparison with the other data to check the authenticity of the combination. “If a user clicks on any search term in the KWIC results display, the program will automatically open the View Files tool (described later) and show the search term hit in the original data file” (Anthony, 2004).

3.7.3.3 Word List

The first thing that should be done before undertaking a new corpus analysis or process is to have a list of all the word present in a corpus. It is useful to point out the relevant area or items in the corpus. Bowker & Pearson (2002) note, it can also be used to calculate lemma words in a corpus of families of the related word forms. It also enables the researchers to find types and tokens running in the corpus under investigation. It can also be calculated through chi-squared or log-likelihood the statistical measures (Kilgariff, 2001), and offering the end-users the option of hiding or displaying the keyness (the unusual frequent or infrequent words) in the preferences window.

3.7.4 Experimentation

Experimentation is the best method to validate the results while manipulating the independent variable upon the experimental group. Experimentation involves the provision of treatment, the effects of which are measured in pre and post-tests and it is most relevant to language teaching (Plonsky, 2014). It enables the researcher to measure and compare the behavior of the experimental group of selected participants with the controlled group before and after treatment. In the present study, this tool addressed the first question: Does the CDL affect the learning of lexical collocation of non-native Pakistani undergraduate?

The sample of the study: the experimental group was treated to complete two semesters (Fall-2019 & Spring-2020). During the earlier semester, they have introduced the basics of DDLT then in the next semester each student was allowed to prepare his corpus of assigned text to do pattern-haunting. Later, these patterns (lexical collocations) were defined by them individually in the presentation demonstrated by them.

3.7.5 Interview

An interview is a popular tool in the case of qualitative research design. A Focused interview attempted to capture attention on the known experience of the interviewees (Buriro, Awan, & Lanjwani, 2017). “This interview seeks to obtain a description of interviewees’ lived world concerning the interpretation of the meaning of the described phenomenon”(Kvale, 2007, p.107). The interview questions are designed in a sequential way to inter-view the subjects’ opinions about a phenomenon. The sequence is given as follow:

Phases of interview questions by Steinar Kvale (2007) in ‘Doing Interview’

1. The introductory question
2. The follow-up question
3. Probing question
4. Specifying question
5. Direct question
6. Indirect question
7. Structuring questions

In a qualitative research design, the interview texts were thematized based on the descriptive and empirical knowledge provided by the participants on the ‘what’ and ‘why’ questions respectively. These interviews comprising of few unstructured and open-ended questions to elicit the opinion and views from the interviewees (Creswell, 2014). To answer the second question of the study, what is the Pakistani university undergraduates’ attitude towards language learning through DDLL? a focused interview was organized to seek the opinion of a selected sample of study to observe the attitude of the participants towards the emerging approach CDL in foreign language learning. The data collected through the interview was then be processed through Nvivo-12 software.

As far as the third question of the research study was concerned, similarly, a focused interview was conducted among the faculty members of the Department of Applied Linguistics G C University Faisalabad Pakistan to seek their views regarding the language learning through CDLL. The received data was processed on NVivo 12 to thematize the meaning-based units of the text of knowledge the participants provided in the answers to the questions asked in the interview. The data was analyzed on the content and thematic base analysis to generalize the results.

3.7.5.1 Nvivo-12 Software

Nvivo is a powerful software to gain richer insight from the qualitative and mixed-method data. It helps the researcher to organize, store, and retrieve his data so he can work efficiently, save time and rigorously backup findings with the exposure of evidence (Godau, 2014). In the case of the present study, data was imported from the e-mail and

processed with Nvivo-12 to have a thematic analysis of the query generated through questions of a focused interview.

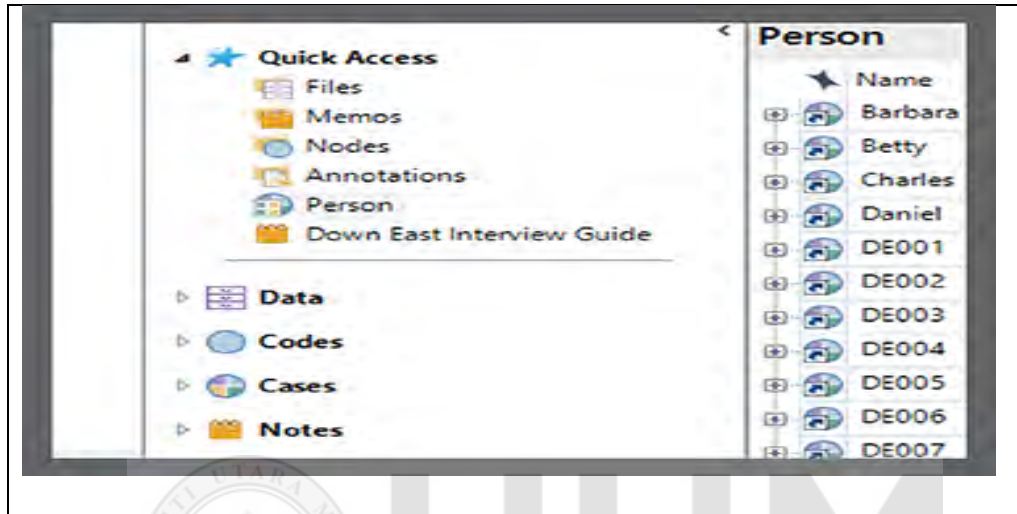


Figure 3.8. Nvivo-12 store and organize

3.7.5.2 World Cloud

Word cloud is a visual communication tool that displays text data in graphic form. It demonstrates the visual cloud of words present in the text according to the attributes and frequency of occurrence with the size, weight and different colours (DePaolo & Wilkinson, 2014). With the help of the word cloud, the researcher can easily assess the knowledge produced by the respondents in textual form by investigating the phrases or words of lack thereof in the textual responses. The researchers can easily generate themes keeping because of the most frequent word of phrase highlighted in the word cloud as shown in figure 3.9.

- Pre-test
- Intervention
- Post-tst

3.8.1 Rationale

The research type of the present study was mainly exploratory, based on the embedded design of the mixed-method under the umbrella of the postpositivist paradigm. The present study explored the effect of CDLL in the teaching of lexical collocations to Pakistani non-native ungraduated university students. The subject of the study was tested earlier to find their level of English; then, the sample of study was bifurcated into experimental and controlled group. The experimental group was treated for a complete semester under the hypothesis of the study. Two types of corpora were compiled to investigate seven types of lexical collocation which were automatically extracted from corpora by the software AntConc. The lexical combination (collocations) was described by the sub-groups of the experimental group to have a deep understanding while comparing these unusual patterns with BNC. After the treatment of a complete semester. A pre-test comprising of 25 items, to check the English collocational knowledge of the selected sample, was formulated and distributed among the participants. The sepc-driven test items were composed while following the series of guidelines by (Kehoe, 1995) and originated from the idea of “Critical Reverse Engineering (RE)” coined by Davidson and Lynch (2002). Data was collected quantitatively and operated upon inferential statistics to measure the P-value of results from experimental group and controlled group.

Later, the experimental group was treated with CDLL exploiting the participants' dominant intelligences, namely visual-special and linguistic-verbal intelligences (Gardner, 1983), for a complete semester of 16 weeks and more attention was focused on the study of collocation. Whereas, the controlled group was treated with the traditional way of teaching. After a complete semester post test was conducted to check the results of the treatment.

Aftermath, a focused interview questions (Appendix-M) was mailed to the participants of the selected sample to seek their opinion regarding the effectiveness of the CDL; the obtained data were analyzed through Nvivo-12 software to evaluate the level of perception of the CDL approach in teaching/learning. The over all frame-work of the research study has been chalked out in the following figure 3.10.

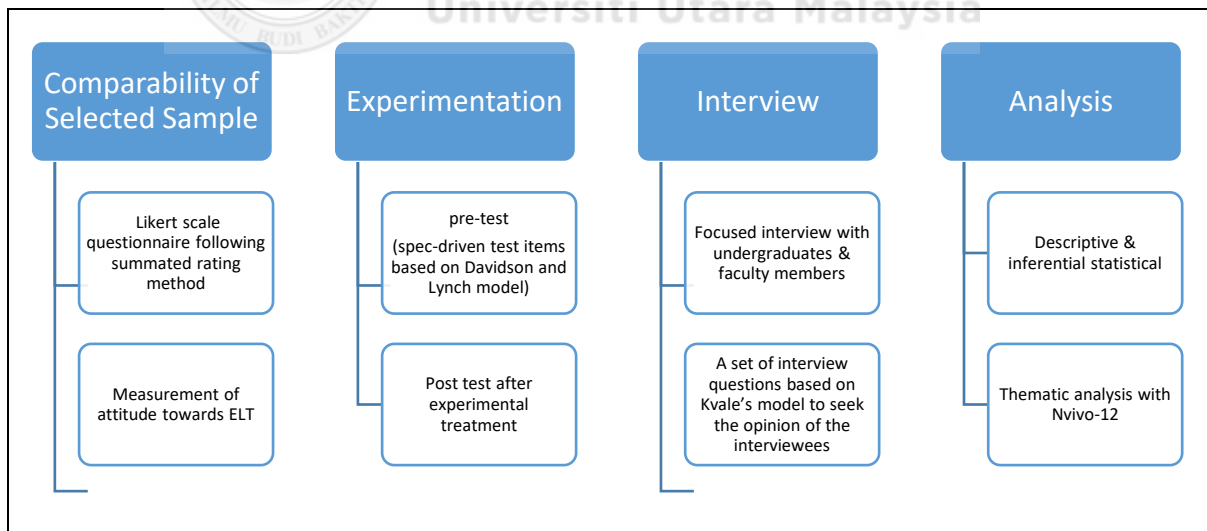


Figure 3.10. Frame-work of study

3.8.2 Pre-test

The selected sample of 100 students were taught revised course of ELT Methods & Approaches, in 3rd semester (Morning + Evening groups) of degree program BS English (Literature & linguistics) during the fall semester 2019. The course outline along with objectives of the given course is given in the appendix-A. A Likert scale questionnaire (see the Appendix-B) was developed while following method of summated rating (Kehoe, 1995) to measure the attitude of learners towards English learning separately. The results were processed through statistical operations as shown in tables 3.3. The students were introduced to the historical perspectives of English Language Teaching. Then, they were taught traditional approaches and methods of English Language Teaching as proposed by Jack C. Richards and Theodore S. Rodgers (1999) up to mid-semester exams which were conducted in November 2019. A corpus of newspaper editorials comprising of one million words was compiled by the students of both morning and evening groups separately of the selected population. The collocation richness from the corpora was measured through the n-gram tool of AntConc software. The data was tagged through PoS tagger then the frequency of Noun + Noun collocations and type/token ratio is measured in the pre-test.

A pre-test comprising of 25 sepc-driven items were composed while following the series of guidelines by Kehoe (1995) and originated the idea of “Critical Reverse Engineering (RE)” coined by Davidson and Lynch (2002), which evaluated the knowledge of lexical collocations of the selected sample (see Appendix C), was administrated upon both groups and the results were operated upon descriptive statistics on excel sheet as shown in the following table 3.9. the p-value of the t-score from the below-mentioned values was 0.212

which was greater 0.05. It means that no significant difference existed between the results of both groups' pre-test.

Table 3.11

Results of Pre-test

Participants	Number of	Number of			
	responses	Blanks	Total	Aerage	SD
Experimental group	40.00	10.00	50.00	13.95	3.99
controlled Group	39.00	11.00	50.00	14.03	3.34

The above mentioned result in table 3.11 indicates the equitable and reliable competence of the both groups, the experimental and the control before experimental treatment.

3.8.4 Intervention

The experimental group was earlier divided into seven groups and asked to prepare corpora of editorials from British and Pakistani leading newspapers. Each group was assigned two newspapers each from a separate country shown in table 3.12:

Table 3.12

The Assignment of Newspapers with the Respective Student Groups

Experimental Groups	Pak Newspaper	British News Paper
A	Dawn	Daily Mail
B	The Nation	Daily Mirror
C	The Express Tribune	Daily Express
D	Daily Times	The Guardian
E	The News International	The Daily Telegraph
F	Business Recorder	The Sun
G	The Frontier Post	The Independent

The students have compiled corpora of about one million words for further description. The corpora were tagged while processing upon CLAWS web tagger. Each group was then assigned to do pattern-hunting of each lexical combination mentioned in table 3.13 and then presented the difference between the frequencies of the selected collocations used by native as well as nonnative editors.

Table 3.13

Lexical Patten-hunting with Respective Selected Group

Sr No	Lexical Pattern	Group	Pattern
1	<i>Verb + Noun</i>	<i>A</i>	<i>(make a mistake)</i>
2	<i>Verb + Adverb</i>	<i>B</i>	<i>(to show clearly,)</i>
3	<i>Noun + Verb</i>	<i>C</i>	<i>(results proved)</i>
4	<i>Noun + Noun</i>	<i>D</i>	<i>(a ceasefire agreement)</i>

5	<i>Adjective + Noun</i>	<i>E</i>	<i>(heavy traffic)</i>
6	<i>Adverb + Adjective</i>	<i>F</i>	<i>(extremely generous)</i>
7	<i>Adverb + verb</i>	<i>G</i>	<i>(simply show)</i>

They also searched the mis-collocations used by the non-native authors of Pak print media while matching with the results of native British print media, and then they looked for the reasons behind the usage of that mis-collocation according to the models' correlative restrictions propounded by Sinclair and Allerton. In this way, they did the process of pattern-defining and used the correct collocations in their personal expressions while demonstrating their presentation.

Earlier, each group did pattern-hunting according to the assigned lexical pattern to find the frequency differences and the respective collocational richness (the type-token ratio of the relative lexical pattern under investigation) as shown in the table 4.10. While processing the corpora on AntConc, the participants were directed to look for all kind of lexical items for example, in case of 'verb + noun' lexical pattern, all types of verbs including progressive, functional, a verb with participles combine with all types of nouns: singular, plural, concrete and abstract were extracted from the data as shown in table 3.14.

Table 3.14

Pattern-hunting of Lexical Collocation

Lexical Pattern	Types of	Types of	TTR in % of	TTR in % of
	COBLENE	COPLANE	COBLENE	COPLANE
<i>Verb + Noun</i>	972	841	94.19%	93.14%
<i>Verb + Adverb</i>	1037	724	89.32%	80.62%
<i>Noun + Verb</i>	1944	1342	96.62%	96%
<i>Noun + Noun</i>	2475	2366	85.64%	94,75%
<i>Adjective + Noun</i>	5292	4469	85.34%	82,71%
<i>Adverb+ Adjective</i>	822	429	96.58%	90.89%
<i>Adverb + verb</i>	1107	839	96.93	93.64%

The students were directed to separate the collocations of said lexical patterns with frequency 2 and more than two. The following table 3.15 shows the lexical-patterns found in the corpora with frequency 2 and more than 2.

Table 3.15

Frequency difference of Lexical Collocations

Lexical Pattern	COPLANE	COBLENE
<i>Verb + Noun</i>	33	42
<i>Verb + Adverb</i>	45	76
<i>Noun + Verb</i>	39	51
<i>Noun + Noun</i>	224	231

<i>Adjective + Noun</i>	438	461
<i>Adverb+ Adjective</i>	28	32
<i>Adverb + verb</i>	31	28

Each sub-group of the experimental group was asked to look for the abnormal co-occurrence of lexical combination in their respective data while comparing the lexical-pattern used by the non-native Pakistani authors with the native ones and with the concordance of BNC (British National Corpus). Each group separated the inappropriate-collocations used by the non-native authors and then discussed them in their presentation in the classroom under teacher's supervision. Each of lexical-pattern has been discussed in the following segments.

Verb + Noun

Among 100000 tagged words of each corpus, sub-group A worked on *verb + noun* lexical pattern. There were 33 such combinations with a frequency of more than one in COPNELE and 42 in COBLENE respectively. The frequency and range of the pattern were greater in British editorials as compare to the Pakistani data. The group presented various patterns that they found extraordinary while personal observation and then found their concordance while processing the item on BNC, either the combination was acceptable or not. They presented the following mis-collocation concerning the pattern-hunting of *verb + noun* collocations.

Table 3.16

Mis-collocations in Verb + Noun pattern of CPLENE

Serial Number	Frequency	Mis-collocation
1	2	Find relief
2	2	Forgetting words
3	2	Hate speech
4	2	Kill words
5	2	Lynch mobs
6	2	Stop hate

All the above-cited mis-collocations, used by the non-native Pakistani authors revealed the dominance of L-1 and socio-cultural influence that implied the semantic and pragmatic correlation restriction of the regional or national languages of the language users.

Verb + Adverb

The sub-group of the experimental group was assigned to do the pattern-hunting of *verb + Adverb* lexical pattern in the corpora of editorials. They found 45 types of lexical patterns in COPLNE and 76 in COBLENE respectively. However, while comparing and matching the patterns used by non-native with the native ones while consulting BNC the following mis-collocations are extracted in table 3.17:

Table 3.17

Mis-collocations of Verb + Adverb Combination

Serial Number	Frequency	Mis-collocation
1	2	Celebrated enough
2	2	Discussed intellectually

There were only two hits that did not match BCN and the native corpus. Oxford collocation dictionary provided the collocation of ‘celebrate’ with the adverbs of justly and rightly. Whereas, the root word ‘discussed’ had collocations with adverbs: exhaustively, fully, in detail, thoroughly, endlessly, extensively, informally, critically and intelligently.

Noun + Verb

The sub-group C was assigned to look for the lexical pattern of *verb + noun* in the compiled data. They found 39 types of this lexical collocation in COPLANE and 51 in COBLENE. The native speaker showed a rich variety of this combination as compared to the non-native Pakistani authors. However, the participants found certain mismatches in this pattern hunting which have been shown in the following table 3.18:

Table 3.18

Mis-collocations found in Noun + Verb Combination

Serial Number	Frequency	Mis-collocation
1	3	Words kill

2	3	Verdict exposed
3	2	Beacon sighed
4	2	Forebears crushed

The students told that while looking for the collocations of the node in Oxford Collocation Dictionary and BNC no hits were found with the combination as extracted by the students. This mis-collocation occurred due to the influence of the L-1 on the usage of a foreign language by non-native speakers.

Noun + Noun

There were 231 hits of *noun + noun* combination types in English editorials whereas, CPLENE showed 224. The participants did the pattern hunting of certain mismatches while comparing them with the relevant data and BNC data. They found the following mis-collocations as shown in table 3.19:

Table 3.19

Noun + Noun Mis-collocation

Serial Number	Frequency	Mis-collocation
1	3	Finder tool
2	3	News desk
3	3	Belly piece
4	3	Business captain
5	3	Case response

6	3	Citizen amendments
7	3	Marriage restraint
8	3	Mosque verdict
9	3	Spike year
10	2	Claim fake
11	2	Corner bursts
12	2	Cream soldiers
13	2	Suspension claims
14	2	Operation report

Sub-group D of the experimental group worked on *Noun + Noun* pattern-hunting. They founded the above-cited mis-collocations and presented among other participants of the experimental group. The major reasons of the mismatches were counted as the cultural and, other tongue influences which remained dominated in the language usages. The whole group made aware of theses mismatches which were not accepted by the native speakers. They were also invited to do further pattern-hunting for more ‘noticing hypothesis’ as propounded by Papp (2007).

Adjective + Noun

Sub-group E of the experimental group very earnestly found the pattern of *Adjective + Noun*. There was a maximum number of collocation extracted through the n-gram of AntConc tool in both corpora. The maximum number of ms-collocations was hit during the resemblance of the unusual combinations with the native corpus and BNC. The

following table 3.20 shows the mismatches found in non-native Pakistani corpus COPNELE:

Table 3.20

Adjective + Noun Mis-collocation

Serial Number	Frequency	Mis-collocation
1	13	Social media
2	11	Chinese investment
3	8	Ukranian city
4	8	Supplementary reference
5	7	Concerned authorities
6	6	Government protest
7	6	Private partnership
8	4	Indonesian tourism
9	4	Russian dog
10	3	Countrywide protest
11	3	Encroachment drive
12	3	Financial inclusion
13	3	Indian media
14	3	Light warning
15	3	Merged area
16	3	National accountability
17	3	Pakistani drama
18	3	Secular face
19	3	Strategic culture

20	2	Autonomous administration
21	2	Big beneficiaries
22	2	Century forbearer
23	2	Complaint politicians
24	2	Concerned authorities
25	2	Credible security
26	2	Digital space
27	2	Domestic tourist
28	2	Electoral exercise
29	2	Expired medicine
30	2	Forced displacement
31	2	Hegemonic ambition
32	2	Hybrid war
33	2	Hypersonic missile
35	2	Ideological distress
36	2	Indian design
37	2	Kinetic means
38	2	Mean Doberman
39	2	Mega corruption
40	2	Modern apartheid
41	2	Multiple changes
42	2	Mysterious gas
43	2	Natural agriculture
44	2	Nazi agenda
45	2	Political victimization

46	2	Rudderless journey
47	2	Slanted coverage
48	2	Spiritual pantheon
49	2	Torn country
50	2	Underage mirage

The above-cited series of mis-collocations were presented by the group in three presentations among the other participants. Among these mismatches, there were the coined collocations which were not used by the native speakers but excessively used by non-native speakers as ‘social media’, ‘concerned authority’, and ‘underage marriage’. Such combinations have been emerged in the language usage of non-native speakers due to semantic and pragmatic restriction of regional and cultural correlations in native languages. However, the whole participants were made aware of the pattern defining of correct combinations according to the Oxford Collocation Dictionary and BNC. Mostly, the natives use ‘mass media’ instead of ‘social media’; ‘relevant authority’ instead of ‘concerned authority’ and ‘early marriage’ instead of ‘underage marriage’ etc. All of these mis-collocations were described one by one by the presenters extracted from the corpus.

Adverb + Adjective

The sixth lexical pattern was explored by sub-group F of the experimental group was *adverb + adjective* combinations. Only one hit among 28 collocations of *Adverb + Adjective* was not matched with the lexical combinations of BNC and the relevant corpus

as shown in table 3.21. Whereas, there was a lexical collocation of ‘deeply conservative’ in the data of COBLANE which was acceptable in BNC.

Table 3.21

Adverb + Adjective Mis-collocation

Serial Number	Frequency	Mis-collocation
1	2	Deeply philosophical

Adverb + Verb

The seventh group G was assigned to do pattern-hunting of the lexical combination *adverb* + *verb*. They found the following mis-collocations presented in table 3.22 which were not matched with the relevant corpus COBLANE and also in BNC.

Table 3.22

Adverb + Verb Mis-collocation

Serial Number	Frequency	Mis-collocation
1	2	Miserably failed
2	2	Necessarily matched
3	2	Reluctantly take

The participants of the experimental group presented all those lexical collocations they derived from the data of corpora under investigation. After having a careful pattern-hunting, each group defined the patterns not matched with the corpora of native speakers. They all did learning under the theories of noting-hypothesis and constructive theory of

DDL and also shared their views under the paradigm of Vygotsky's sociocultural theory. The experimental treatment continued for the whole semester for 16 weeks.

Meanwhile, the controlled group was guided with the traditional way by using dictionaries and other hard material for the realization of the correct use of collocations according to the native speaker. A post-test was conducted at the end of the semester that has been analyzed in the following segment of this chapter.

3.8.4 Post-test

In the phase of post-test, the participants of the study were administrated upon certain standardized test items (Appendix-D) regarding the knowledge of lexical collocations and their output was compiled in the form of numerals which was latterly analyzed through descriptive statics and the comparison between the results of both groups was analyzed through t-test of inferential statistics(Costa et al., 2018) as shown in the following table 3.23 which showed the significant effect of the treatment of the experimentation.

Table 3.23

Statistics of the Results of Post-test

Groups	Number of Responses	Number of Blanks	Total	Mean	SD	T-test P-value
Controlled	46	4	50	13.67	2.74	
Experimental	49	1	50	22.41	2.09	1.2923E-29

3.9 Data Analysis Strategy

Testing in the education system acquires great importance; the results from the tests can be used to make decisions. To effectively utilize the tests in education/research dimensions and quality assurance their validity and reliability estimates are necessary (Bichi & Talib, 2018). In the present research study, The items selected for the tests were selected through item analysis (Arora, 2018). Then, the responses were analyzed based on Item Response Theory which has been described in detail in the following segments.

3.9.1 Identification of Lexical patterns

This portion describes the method used in the preliminary investigation. The statistical information on both corpora was obtained by AntConc 3.4.1, a famous software that showed the number of tokens and types in both corpora. Although the selected material was not too large to examine the overall behavior of the language learners yet the same provided the logical information to meet the required criterion. For the comparative measurements of corpora, the figure of type/token ratio, collected through the word list tool. The data was tagged by the PoS (CLAWS4) tagger with PoS marks in tagset C5 which means each word in both corpora has a PoS tag after it processed by CLAWS4.

After PoS tagging of the two corpora, COPLNE & COBLENE, the software ActCons with n-gram tool was used to extract seven types of lexical collocations and examine concerning their co-occurrence restrictions, For example, the word ‘choice’ was tagged as “choice_NN1” and the word ‘good’ is tagged as “good_AJ0”. In this way the lexical

patterns “good choice” could easily be extracted from n-gram of AntConc software after selecting the file “*_AJ* *_NN*” was written in the command key with the cluster size min.4 to max.6, frequency 3 and range 1 to have the list of all ‘Noun-Noun’ lexical patterns and similarly, “*_AJ* *_NN*” was commended for ‘Adjective-Noun’ pattern and the other lexical patterns like “Noun-Noun”. Noun-Verb”, and “Verb-Adverb” were extracted with the same technique.

Two corpora was compared in terms of their genre of writing i-e newspaper editorials.

Although the topics are not the same in both corpora, it will not affect the required results as the analysis of the collocational patterns such as ‘Noun-Noun’ etc. is not affected significantly by the frequency of occurrence of specific words. The frequency of each lexical pattern was extracted through the n-gram tool of AntConc software. Then, the obtained figures were noted for further analysis. Certain patterns that are not correctly tagged by the tagger, these patterns will be manually excluded from the result. Such kind of anomalies will be removed carefully for accuracy. The total number of collocations present in each corpus was then normalized @ 10000 words due to different sizes of both corpora. The total number of collocations in each lexical pattern was treated as the token and the total number of different collocations be treated as types. To find out lexical diversity between both corpora, the type/token ratios were evaluated according to Guiraud's (1954) mathematical transformation of TTR.

The unexpected patterns from COPLANE were compared with the collocations with the same node present in COBLANE and then with COCA by jotting down the code word in the search bar to have all the relevant collocations available online. In this way, certain mal-formed collocations were extracted. These kinds of mis-collocations used by, particularly, non-native learners were then be treated in the concordance tool of AntConc software from raw data of COBLANE to find the resemblance. These kinds of mis-collocations were categorized for further analysis of the factors behind the coinage of certain combinations which asserts the foreignness of non-native language users (Shah, Singh & Kaur, 2020):

- Collocation coined by non-native Pakistani authors
- Collocation used with cultural connotations
- Collocation used with stylistic features

The above-mentioned categories then were recalled while teaching the importance of collocation in language learning to assure the selected sample for further treatment of experimentation.

3.9.2 Thematic Analysis of Interviews

To listen to the voice of the participants regarding the effectiveness of DDLL and to answer the 2nd question of the research study, students of the experimental group participated in a focused interview conducted at the end of the semester. These were the participants who volunteered to be interviewed; they were contacted by e-mail and provided with time to think about the questions in advance. The questions were organized keeping because of the familiarity with the participants as shown in Appendix-F. The replies of the participants

were treated with Nvivo-12 to process the thematic analysis of the answers of the interview questions.

Similarly, to answer the third question of the research study, a focused interview was also administrated among the faculty members of the department of applied linguistics GC University to seek their opinion regarding the effectiveness of CDL. This semi-structured interview aimed to elicit participants' attitudes towards corpus-driven language learning (CDLL) in terms of language learning in Pakistan with different approaches. The interview questions were intentionally arranged according to Steinar Kvale (2007) sequence as mentioned below:

Phases of interview questions by Steinar Kvale in Doing Interview

- The introductory question
- The follow-up question
- Probing question
- Specifying question
- Direct question

Their views of participants were obtained in the form of written text and then processed for content/thematic analysis with the help of Nvivo-12 software to thematize the opinions sought from the interview questionnaire (please refer to Figure 3.10).

3.10 Implication from Pilot Study

The reliability and validity of the test results were based on the authenticity of pilot study. The procedure of item analysis for both pre and post tests based on classical test theory which was used for the selection of items for test composition. It is a statistical technique that is used in selecting or rejecting the items of the test based on their difficulty value and discriminated power. This approach was common to check the validity and reliability of the development and administration of language tests for English (Raquel, 2019). Data resulting from the field trial that was administrated at the initial stage of the study was analyzed to check the facility value, discrimination, and internal consistency properties before determining whether these could be used for the development of pre and post-tests. I intentionally kept in view, the following questions, while developing the tests before creating the answer key to develop the familiarity with its strength and weaknesses and to have acquaintance with the target in terms of construct and the extent to which test-takers might be successful.

- Did the items appear to be at the appropriate level of difficulty?
- How well did the distracters fulfill their function?
- How well the individual did items discriminate?
- Did every item contribute to the internal consistency of the task?

The test method used to collect data was made likely to be familiar with the target population. To maintain reliability, following the pilot study, multiple-choice and matching task was selected to get the responses from the selected sample. The key was reviewed to ensure that it was correct which meant that only one possible answer was correct.

The tests were scored by using a conventional Number Right (NR) scoring method (Kurz, 1999). Items of the test were dichotomously scored, correct answers were scored with a positive value, incorrect answers or omitted answers with a value of zero. The sum of scores for test responses was the test scores. The mean and central deviations were calculated based on the numerical (interval) scale to find out the exact difference between the values obtained from data.

Earlier 50 items were selected and administered upon the 40 students of 4th semester of Applied Linguistics studying at Kaims International Institution Multan Pakistan. To find difficulty value (D.V); first, the scores were arranged in descending order then took 27% of the scores out of the higher score, and 27% of the score falling at the bottom. It means that the students who obtained higher scores and those who obtained lower ones were selected for item analysis. Then the number of the students was also counted who obtained the higher scores (N.H) and lower scores (N.L). For each item counted the number of right answers (R.H) from the higher group and counted the number of right answers from the lower group (R.L). at the same time, the number of non-responses was also counted (N.R). Difficulty value is the proportion of the percentage of an examinee who attempted the correct answer; D.V of each item was calculated with the help of following formula:

$$D.V = \frac{(R.H + R.L)}{(N.H + N.L) - N.R}$$

According to Blood and Bud – the index of discrimination is the ability of an Item based on which discrimination is made between superior and inferior. It ranges from +1 to -1.

Each of the field items was operated upon the following formula to find the discrimination index (D.I):

$$D.I = (R.H. - R.L.) / N.H \text{ or } N.L$$

Among 50 items 28 were selected for the pre and post-tests as their D.V ranged from .30 to .70 which fell under the standard values and with D.I more than 3.5.

3.11 Summary

The present study is very much influenced by the symmetrical philosophical stance: ontologically and epistemologically committed to postpositivism school of thought. As it is an exploratory research study the exploration of the effectiveness of CDL approach in teaching collocations will be made based on a preliminary investigation that was endorsed with the experimentation method including pre and post-tests conducted by the researcher on the selected participants of GCUF students. To reassure the validity and reliability of the results, the test items were selected according to the classical theories and models. later, focused interviews were conducted among students and teachers to observe the pros and cons of the CDLL approach in language learning/teaching to assure the authentic implications of the research study. However, the whole framework of the study has been summarized in the following table 3.24

Table 3.24

Summary of the Methodology

Research Objective	Research Questions	Research Tools	Data Analysis
To investigate the effectiveness of CDLL in teaching collocation to Pakistani undergraduate university students.	To what extent does the use of a CDL approach in the classroom is effective in the teaching and learning collocations among Pakistani undergraduates?	Pre-Test (without CDL) and Post-Test (After CDL) (Experimental Design)	Descriptive & Inferential statistics
To seek the Pakistani university undergraduates' opinio towards learning English language learning through CDLL.	What are the Pakistani university undergraduates' attitudes towards learning English language learning through CDLL	Focused Interviews	Thematic Analysis with NVivo 12
To seek the Pakistani university teachers' opinion towards CDLL in English language teaching/learning	What are the Pakistani university teachers' attitudes towards the effectiveness of CDLL in English language teaching/learning?	Focused interviews	Thematic Analysis with NVivo 12

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Introduction

In an explanatory sequential mixed-method research design, a researcher first collects and analyzes the quantitative (numeric) data then he/she collects and analyzes the qualitative data in the form of textual material which helps to elaborate the results obtained from quantitative data phase (Ivankova, Creswell, & Stick, 2006). The data collected for this research study has also been classified into two categories: the quantitative data collected through the two tests administrated at the beginning and at the end of a complete semester comprising of 16 weeks for experimental study, that were pre-test, and post-test. quantitative research usually involves numerals, quantifications, and statistics to reach out to the solutions of research questions (Costa et al., 2018). The qualitative data which included the semi-structured interview questionnaires e-mailed to the participants. The first type of data was analyzed using certain statistical operations. These statistical calculations aimed to reach appropriate answers to the first question of the study, The second type of data was analyzed through Nvivo.12 to seek the opinion of the participants and to answer the third and fourth questions of the research study, The last portion of the chapter discusses the strategies adopted for data analysis.

4.2 Quantitative data Analysis

In the first phase of data analysis, the data was analyzed quantitatively to reach the appropriate answer to the first questions of the research study:

To what extent does CDLL contribute to the development of the English collocational knowledge of Pakistani university undergraduates?

As the sample selected for the research study were the undergraduates of the 4th semester (Morning & Evening session) of GC University Faisalabad Pakistan. Earlier, the sample of the study was administrated upon a as pre-test and then after treatment a post-test was conducted. To obtain the feedback and to have statistics for the research work an objective type test was chosen to be administrated upon the sample of the study. The declarative knowledge of the participants was evaluated by the multiple-choice questions whereas the procedural knowledge in the case of language learning was assessed by matching the column portion of the test. The results from the data obtained was analysed by descriptive statistics. The quantification of data has been described in detail in the following phases.

4.2.1 Data Analysis Method of Research Question 1

To quantify the participants' performances on pre-test and post-test which were administrated upon the selected sample. For the quantification of the pre-test, a three-part of the test is composed to measure the performance of participants regarding the choice of lexical collocations. The first part of the test was comprised of assessment about matching the column, the second part consisted of correct use of lexical collocation in sentences and the third part comprises of fill in the blank. The focus of the attention on all these questions was the usages of seven types of lexical collocations as shown in the following table 4.1:

Table 4.1

Types of Lexical Collocations

Sr No	lexical pattern	examples
1	<i>Verb + Noun</i>	<i>(make a mistake)</i>
2	<i>Verb + Adverb/Adjective</i>	<i>(to show clearly,)</i>
3	<i>Noun + Verb</i>	<i>(results proved)</i>
4	<i>Noun + Noun</i>	<i>(a ceasefire agreement)</i>
5	<i>Adjective + Noun</i>	<i>(heavy traffic)</i>
6	<i>Adverb + Adjective</i>	<i>(extremely generous)</i>
7	<i>Adverb + verb</i>	<i>(simply show)</i>

Data were collected from 100 participants including 50 in the experimental group and 50 in controlled groups. The results of the questionnaire were operated upon the tabulation with the statistical process of an excel sheet to find the average and standard deviation for the t-score of the obtained results as shown in table 4.2:

Table 4.2

Statistical Measure of Pre-test Data

Participants	Number of responses	Number of Blanks	Total	Aerage	SD	T-score
Experimental group	40.00	10.00	50.00	13.95	3.99	P value 0.21
Controlled Group	39.00	11.00	50.00	14.03	3.34	

The p-value of the above-mentioned results of table 4.1 indicates that there is no significant difference between the results of the collocational knowledge of both groups before experimentation as the p-value is higher than 0.05. The average score of the experimental group is 13.95 out of 25 and the average score of the controlled group is 14.03 out of 25. Most of the participants have selected mis-collocation which was not accepted by the Oxford dictionary of collocation as available online and the concordance of COCA also rejected this choice. A detailed discussion of the results of three parts of the questionnaire has been described in the following segments.

4.2.1.1 Analysis Strategy of the Part-1 of Pre-test

In the first part of the test, there was a list of seven lexicons in the left column to be matched with the lexicons of the right column keeping because of the natural combination of the words.

Table 4.3

Part-1 of Pre-test

Sr No	Lexicon	Collocate
1	Help	View
2	Gorgeous	a fuss
3	soar	Sorry
4	aloof	Rain
5	extremely	upward
6	ceaseless	From
7	make	Desk

61 out of 79 participants wrongly matched ‘help’ with ‘from’ because of the L-1 influence as the first language expression demanded this combination. 54 out of 79 participants

rightly selected the second item ‘gorgeous’ collocation with ‘view’ as influenced by the mutual semantic compatibility that stood in “semantic restriction” propounded by Sinclair & Allerton (Lu, 2016). 95 percent of the participants have correctly matched ‘soar’ with ‘upward’ due to semantic prosody which shows the speaker’s attitude towards what is being talked about and is the pragmatic level of language (Sinclair, 1991). 65% of students have correctly attached ‘aloof’ with ‘from’ as the result of “locational restriction” propounded by Allerton (1984). ‘Extremely’ was also rightly attached with ‘sorry’ by 71% of the participants but ‘ceaseless’ was wrongly attached by 76% with ‘a fuss’ due to cultural influence and ‘make’ was also combined wrongly with ‘desk’ due to national language dominance.

4.2.1.2 Analysis Strategy of the Part-2 of Pre-test

The second part of the test comprised of ten items of equal semantically statements with different lexical collocations. The participants were asked to select the correct expression with a natural combination of lexicons. As there is no significant difference between both of the groups (experimental & controlled) the overall result is being analyzed in the following paragraph.

Table 4.4

Part-2 of Pre-test

Q1 - Which is correct?
I have done a great deal of work.
I have done a big deal of work.
Q2 - Which is correct?
We're in severe straits.
We're in dire straits.
Q3 - Which is correct?
I've got a deafening headache.
I've got a blinding headache.
Q4 - Which is correct?
It wasn't much trouble- it was no great deal.
It wasn't much trouble- it was no big deal.
Q5 - Which is correct?
It is a fatal disease.
It is a mortal disease.
Q6 - Which is correct?
He's terminally ill.
He's mortally ill.
Q7 - Which is correct?
It's a small priority issue.
It's a low priority issue.
Q8 - Which is correct?

I like the environment of the restaurant.

I like the ambiance of the restaurant.

Q9 - Which is correct?

It was total chaos.

It was great chaos.

Q10 - Which is correct?

She was highly drunk.

She was extremely drunk.

The results of the part-2 of question-set has been presented through excel tabulation with the frequency distribution of the percentage of the answers given by the respondents as shown in table 4.5:

Table 4.5

Results of Part-2 of Pre-test

Participants	Responses	not responded	Total	Correct	incorrect	correct %	incorrect %
Q 1	99	1	100	25	74	25%	75%
Q 2	100	0	100	55	45	55%	45%
Q 3	97	3	100	22	75	23%	77%
Q 4	94	6	100	70	24	74%	26%
Q 5	100	0	100	80	20	80%	20%
Q 6	95	5	100	20	75	21%	79%
Q 7	100	0	100	80	20	80%	20%
Q 8	98	2	100	43	55	44%	56%
Q 9	96	4	100	75	21	78%	22%
Q 10	97	3	100	37	60	38%	62%

The result of part-2 has been figuratively presented with the columns as shown in the figure 4.1 the blue colour column presents the correct answers and brown the incorrects.



Figure 4.1. Results of part two of the pre-test questionnaire

The graphic representation of the results of the second part of the test as shown in figure 4.1 shows the overall performance of the 100 selected participants. The items attempted by the participants have been discussed in a detailed analysis in the following paragraph.

75% of the population of the selected sample have marked the wrong combination of the first item of 2nd part of the test due to the semantic correlative restriction of the L-1. The 2nd item was probably attempted half of the population right as because of the syntactic restriction. The third item was mostly ticked wrong “blinding headache” due to the socio-cultural effects on language. The fourth and fifth items were almost selected correctly due to idiomatic expression of pragmatic restriction of correlative lexicons. 79% of the sixth

item was marked incorrect due to the influence of L-1 and unawareness of collocational knowledge. The participants marked ‘mortally ill’ mostly instead of ‘terminally ill’ which is the strong collocation of the English language. Seventh and ninth items are marked correct by most of the participants due to the structural association of the lexical items. Whereas, eighth and tenth items were wrongly marked by most of the participants due to correlative semantic restrictions of collocations.

4.2.1.3 Analysis Strategy of the Part-3 of the Pre-test

The third part of the test comprised of multiple-choice items in which the participants were asked to select a correct collocation of the given lexicon as mentioned in table 4.6

Table 4.6

Item-3 of Pre-test (Multiple Choice Questions)

Tick the appropriate word suitable for the blank region of the sentence:

1. An open and friendly manner is ----- of this region.

(a) Personality (b) characteristic (c) individuality (d) Character

2. The children have been as good as gold. So well-----

(a) Adjusted (b) intentioned (c) dressed (d) behaved

3. I thought his comment about my putting on too much weight was quite -----.

(a) Tactless	(b) cynical	(c) sensitive	(d) egotistical
4. My sister is really -----t. She can join in a group without any introductions.			
(a) Confident	(b) controlled	(c) absorbed	(d) Centered
5. I read the newspapers every day because I like to be ----- about what's going on in the world.			
(a) Known	(b) Informed	(c) Educator	(d) brought-up
6. The ----- paragraph gives the background to their research project.			
(a) preceding	(b) bygone	(c) retrospect	(d) beforehand
7. These are very ----- times, and it may be a while before everything returns to normal.			
(a) temporary	(b) turbulent	(c) contemporary	(d) rapid
8. There has been a ----- variation in the way we see modern society.			
(a) shift	(b) amendment	(c) alteration	(d) variation

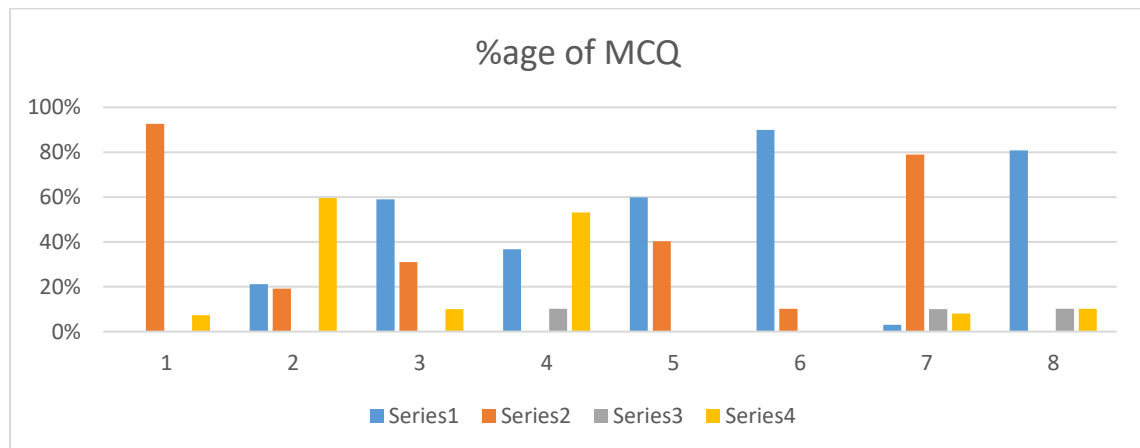
The results of the item-3 have been quantified with the help of descriptive statistics on excel tabulation in the following table 4.7:

Table 4.7

Tabulation of the Results of Item-3

Participants	Responces	Blanks	Total	% of 'a'	% of 'b'	% of 'c'	% of 'd'
Q 1	96	4	100	0%	93%	0%	7%
Q 2	99	1	100	21%	0%	19%	60%
Q 3	100	0	100	59%	31%	0%	10%
Q 4	98	2	100	37%	0%	10%	53%
Q 5	97	3	100	60%	40%	0%	0%
Q 6	99	1	100	90%	10%	0%	0%
Q 7	100	0	100	3%	79%	10%	8%
Q 8	99	1	100	81%	0%	10%	10%

Further, the results of third part of question-set is presented in the graphic form in the following figure 4.2:

*Figure 4.2. Results of the item 3 of pre-test*

93% of the participants successfully selected the right option of the first question of item-3 of the questionnaire because of semantic restriction. Most of the participants rightly ticked the correct option of question 2 some did wrong due to cultural influence on the language. Nobody attempted the correct option of the third question rather they opted mis-collocation for the adjective 'quite' due to pragmatic restriction of their socio-cultural influence. Few opted (37%) the correct collocation in question 4 whereas most of the participants selected the option 'd' which was mis-collocation due to semantic correlative restriction. Question 5 was also attempted wrong by a large number of participants. Question 6 was mostly opted successfully because of the semantic restriction. Question 7 was successfully opted correctly by 79% of the selected population due to semantic restriction of lexical collocation. The last question number 8 was also opted correctly by 81 % of the participants.

The overall analysis of the results showed that the participants of the selected population did not have sufficient collocational knowledge; they mostly mismatched the lexicons of the item-1. Mostly they selected the erroneous-collocations while attempting the other two items of the pre-test as influenced by the L-1 and their socio-cultural perspectives. Later, the selected sample was bifurcated into the experimental group and the controlled group to explore the authenticity of the hypothesis of the research study.

4.2.2 Analysis Strategy of the Post-test

After an experimental treatment of 16 weeks the both groups, experimental & controlled were administered upon post-test which comprised 25 Items in three sections (see Appendix D). The results were processed on the descriptive statistics as shown in the following table 4.8. The use of mean (M) and standard deviation (SD) were applied to compare the knowledge of lexical collocation gained by the experimental and controlled group through CDLL and traditional approaches of language learning respectively. Additionally, a t-test from inferential statistics was also applied to determine if the difference in the results of the test was significant. The level of significance was determined to be at the level of 0.05 and below.

Table 4.8

Statistics of the Results of Post-test

Groups	Number of Responses	Number of Blanks	Total	Mean	SD	T-test
						P-value
Controlled	46	4	50	13.67	2.74	
Experimental	49	1	50	22.41	2.09	1.2923E-29

The below-cited figure 4.3 shows a significant difference between the results of both groups. The mean value of the results of the experimental group is greater than the controlled group. Moreover, the p-value of the t-test which is 1.29 E-29 that is very very low indicated that the difference between the results was significant.

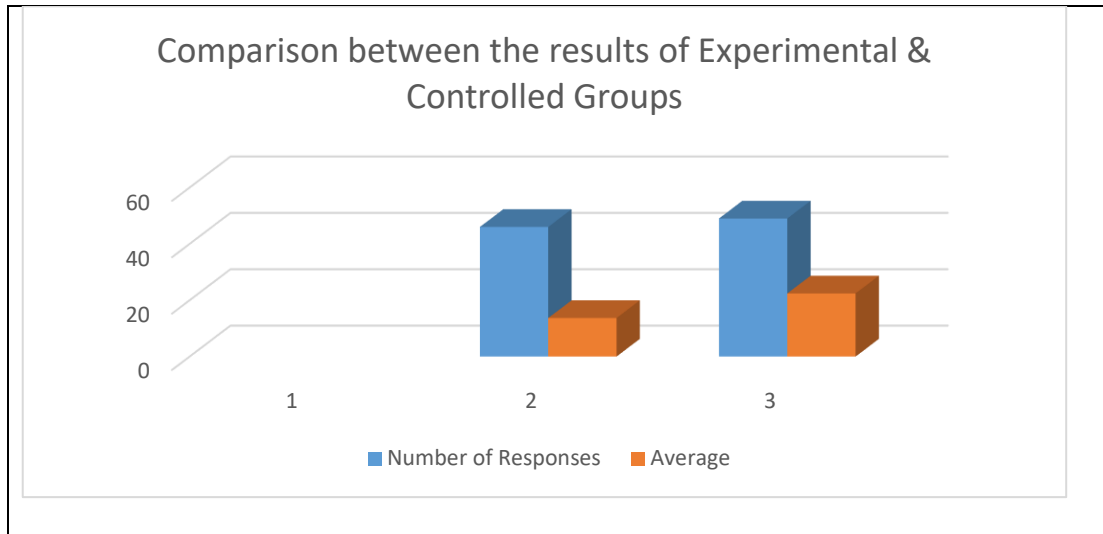


Figure 4.3. Comparison of the results of post-test

A t-test is used to compare the means to each other of the results of two groups. Herein this study the means of the participants' scores obtained in the pre and post-tests were compared with each other by t-test. To find a significant difference between two means, I found the p-value on the excel sheet. As the p-value determines how much significant the difference between the two results. If the probability value is less than 0.05 it indicates that the difference between the results is significant. The statistics of both tests are given below in table 4.9:

Table 4.9

Comparison between the Scores of Control Group and Experimental Group in Pre and Post tests

Participants	Non CDLL Group N=46		CDLL Group N=49		P-value
	Mean	SD	Mean	SD	
Pre-test	14.03	3.34	13.95	3.99	0.21
Post-test	13.67	2.74	22.41	2.09	1.29E-29

To obtain the feedback and to have statistics for the research work an objective type test was chosen to be administrated upon the sample of the study. The knowledge of the participants was evaluated by the spec-driven multiple-choice questions and matching the column portion of the test. The statistics of both types have been shown in the following table 4.10:

Table 4.10

Comparison between the Declarative and Procedural Parts of the Tests of Both Groups

Test type	Non CDLL Group N=46		CDLL Group N=49		P-value
	Mean	SD	Mean	SD	
MCQ 18	11.50	2.00	13.37	1.37	8.07E-07
Matching 10	2.17	1.05	9.04	1.41	2.50E-28

4.2.3 Descriptive Statistics

To find out the significant difference between the results of the pre-test and post-test the difference between the results of the individual participant was calculated on the excel spreadsheet then the columns of the differences were operated upon ANOVA (single factor) operational tool to measure the significant difference. The ANOVA results have been shown in the following table 4.11:

Table 4.11

ANOVA (single Factor) Results of the Difference between the Results of Control Group & Experimental Group

<i>SUMMARY</i>				
<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Controlled Group	50	82	1.64	25.94939
Experimental Group	50	526	10.52	36.90776

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1971.36	1	1971.36	62.72509	3.76E-12	3.938111
Within Groups	3080	98	31.42857			
Total	5051.36	99				

The results show that the sample size of both the control group and the experimental group is identical. There was an improvement in the results of the controlled group as 82 is a positive number but the sum of the experimental group (526) is much larger than the controlled group. Now ANOVA test verified the significant difference between the results of both groups after experimental treatment. The *p-value* is much lower than the alpha said i-e .05. It indicated a significant difference between the controlled group and the experimental group.

The validity of the experimentation was maintained as:

- The factor of history was controlled by taking the tests of both the groups simultaneously.
- Instrumentation was controlled with fixed test based on item analysis theory
- The factor of regression was controlled by the mean differences regardless of the extremity of scores of the characteristics
- However, the factor of morality might not be controlled properly. There remained a possibility of complex interactions which affected the results. Although the conditions for both groups were tried to be similar yet the both groups could not have similar sessions with the teachers.

4.3 Qualitative Data Analysis

This phase of the chapter 4th endeavors to explore the answers of 2nd and 3rd questions raised in the first chapter:

- What are the Pakistani university undergraduates' opinion about learning English collocation through CDLL?
- What are the Pakistani university teachers' opinion about the effectiveness of CDLL in English language teaching/learning?

This phase of the research study comprised of two segments according to the research questions. A qualitative approach was followed to collect data through interviews. The qualitative interview is a key value for exploring how the selected participants of the research study experience and understand the phenomenon under investigation (Kvale, 2007). Earlier, ten participants who ventured themselves voluntarily, of the study were approached through email and asked to answer the open-ended questions (see Appendix M). The aim of having data in the form of text of the answers of interview questions was to elicit Pakistani university undergraduates' attitudes towards language learning through corpus-driven language learning (CDLL). In the context of the mixed method of the present research study, the qualitative data collection after the quantitative operation, attempts to integrate the two approaches. The purpose of this method, according to Sandlelowski (2003) is to maintain reliability and validity through one set of finding against others. The purpose of this method is to achieve the traditional goal of triangulation, namely to validate the conclusion by presenting the converging results obtained through different methods. Therefore, the results of the experimentation triangulated against the results of focused

interviews to seek the opinions of the teachers & students in favor of CDDL so that a comprehensive understanding of the phenomenon in question could be reached.

4.3.1 Analysis Strategy of Research Question 2

The content analysis of the textual material was carried out through coding and classification of the received data from the interviews. Coding involves attaching one or more keywords of the text segment which later serves as the identification of a statement, whereas, categorization comprises of the assembling of code of the same category to generate a theme to elicit the general opinion of the subject about the phenomenon (Kvale, 2007).

As far as the present study is concerned, a sample of stratified randomly selected participants from the experimental group was asked to share their views in a focused interview question set right after the post-test session. Ten out of fifty students were approached through focused interview question-set (see Appendix-K), conveyed by e-mail. The responses were processed through NVivo 12 (software for content analysis) and the detail of the results have been discussed in the following paragraphs.

This focused interview aimed to elicit participants' attitudes towards corpus-driven language learning (CDLL) in terms of language learning in Pakistan with different approaches. The interview questions were intentionally arranged according to Steinar Kvale (2007) sequence as mentioned below:

Phases of interview questions by Steinar Kvale in Doing Interview

- The introductory question
- The follow-up question
- Probing question
- Specifying question
- Direct question

The themes of the study were clarified and categorized from the software NVivo 12. Earlier the textual data was imported to the software, then the files were processed to explore the word frequencies and a word cloud was generated to have a visual representation of the keywords used by the respondents while answering the interview questions as shown in figure 4.1 (DePaolo & Wilkinson, 2014).



Figure 4.4. Participants response to the interview based on the investigation of the effectiveness of CDLL.

The word cloud helped to gauge the use of keywords in students' responses regarding the effectiveness of CDLL in English language teaching/learning in Pakistan; these keywords helped to generate nodes in the phase of codification of textual material while processing NVivo12. Then the nodes were categorized to generate common themes. Three common themes were generated from 30 nodes according to the word cloud which is also relevant to the schema of interview questions as shown in the following table 4.12:

Table 4.12

Coding Summary of the Interview of the Students by NVivo 12

Name of Node	Description
English Language in Pakistan	Thematic node indicates the status of English in Pakistan
Encouragement of Peers in Speaking Eng	Sub-node under the umbrella of thematic node Peers of the learner appreciate him
Learning English	Learning English is compulsory for doing a good job in the country
Listening & Speaking of University students	speaking and listening capability in students becomes weak due to this at university level students hesitate to speak English and give most of their time to correct their Grammar pronunciation

Personal Stance about ELT in Pak	Perticipants' personal stance regarding English teaching/learning in the country
Poor Teaching System	There is a poor teaching system in our country especially for English language teaching or learning
Primary Level Teachers in Pakistan	in villages, teachers are matric pass and they also teach English in a cramming way and they don't create interest of students in This language.
Source of English Learning	The major source of English language learning is classrooms where English is not used as the target language, rather we perform boring English spelling drills, some formal grammatical rules and practice and precise definitions of English terms
Students of Underprivileged areas	students who are from lower areas don't speak English in class due to having low confidence. They can not properly learn it as language
Usage of English in Routine	Although I don't use it very often in daily communication, I use it a lot especially in more formal situations but in an informal situation that uses is limited.
ELT Approaches	ELT approaches need to be upgraded according to modern techniques

ALM Approach	ALM is being widely used in our country. Audio Lingual Method became popular during WW1. It is drilling under the supervision of the instructor
Better Approach of ELT	, both the approaches have their pros and cons and used accordingly. The inductive approach takes a lot of practice & skill that newcomer would like but it is also more fruitful
DDL Approach	DDL is a modern technique in which language becomes data and learner becomes a researcher to investigate linguistic patterns.
Deductive Approach in ELT	The learners are trained with certain rules then they use the rules to study language
GTM Approach	GTM is old fashion and failed to enhance the proficiency of language learners
Inductive Approach	The inductive approach indicates the learners explores the phenomenon at their end. It is better than deductive
opinion about CLT approach	CLT is a common language learning approach in which language is learned through communication.
Corpus Driven Language Learning	Corpus Driven Language Learning is an inductive approach learning language in which a large number of corpora is utilized and learners are asked to do pattern hunting that helps them in

identifying linguistics pattern. I came to know this term while reading about DDL

Autonomous Learning	Learning independently with the inductive approach
CDLL Preference	Participants prefer Corpus driven language Learning
Corpus Information	The participants have corpus information from their classroom
Features of Language in Corpus	Learning any language with the help of prepared corpus (a collection of written texts, especially the entire works of a particular author or a body of writing on a particular subject) while understanding, analyzing and then using the features of that corpus in your communication
Generalizing in CDLL	CDDL helps in generalizing a set of rules or particular theory
Idea of CDL	Idea of corpus driven learning was generated during ELT methods and approaches.
Line in concordance	In concordance we focus on line and on that line we find the simple meanings of words, by finding the sense of a line we focus on words which gives proper meanings
Meaning in Concordance	Students focus on meaning during concordance

Multiple exposures of words	By operating on the corpus, you get the multiple exposures of a word you search in the target language
Process of concordance	The participants did concordance using AntConc, COCA and BNC

The interview question-set was comprised of 7 questions that were administrated upon 10 participants who voluntarily vouchsafed to respond through e-mail. The question I & 2 were introductory question which sought the general views of the participants about the English language in Pakistan. Most of the participants did not support the better condition of the level of English in the country due to several factors as mentioned in the summary of the above-cited table 4.20. While answering the 2nd question, the participants who are university undergraduates claimed that they could not have the proper opportunity to use the English language in their routine that indicated English was not the second language of the country.

The followup question was composed to seek the opinion of the participants about the approaches of language teaching/learning. They were of the view that mostly GTM and deductive approaches were used in the English classroom which had not successful results.

In the case of probing questions, the participants were asked which approach was better in English language teaching/learning. They unanimously responded in favor of the inductive

approach through procedural knowledge would be the best one for language learning like data-driven learning.

While in the specifying phase, the participants were specified with the learning through a concordance. They preferred concordance lines and language learning in CDLL. As far as the direct question phase was concerned, the participants responded that the CDLL approach in language learning was effective to a great extent with the assistance and supervision of trained teachers; which confirmed the reliability of the hypothesis already confirmed in the phase of quantitative analysis.



Figure 4.5. Hierarchy of nodes of the interview text

4.3.2 Analysis Strategy of Research Question 3

This phase of analysis helps to answer the third question of the research study which asks:

- What are the Pakistani university teachers' opinion about the effectiveness of CDLL in English language teaching/learning?

A focused interview question-set (see Appendix-M) was prepared (keeping in view the validity of the study) to seek the opinion of the faculty members of the Department of Applied Linguistics GC University Faisalabad. 10 faculty members including two male and eight female teachers of the department were approached through e-mail and they responded positively as per the requirement of the study. The text of the responses was compiled and processed with the help of NVivo 12 based on analysis focusing on meaning.

This analysis involved five steps:

1. Read through the text of interview responses completely
2. Import the text files to the NVivo 12 software and explored the word cloud
3. Generated the nodes based on keywords highlighted in the word cloud
4. Classified the nodes to create themes
5. Themes of the entire text were tied together into a descriptive statement

The respondents replied to the interview question within one week. They were not compelled to respond earlier to maintain the reliability of the research study. All the received emails were studied thoroughly to clear the sense of the voice of the respondents. The text files were imported in the NVivo 12 software to explore the word frequency and word cloud as shown in figure 4.4. Word cloud helped the researcher to focus on the keyword of the text generated by the respondents.



Figure 4.6. Analysis of the entire text of the responses of faculty members regarding the effectiveness of CDLL

Figure 4.5 represents the visual image of the views of the faculty members regarding English language learning through corpus methods. The text was then codified to generate nodes through NVivo 12 based on the keywords. The summary of the nodes was presented in the following table 4. 12. There were 27 nodes generated while codifying the text of the interview responses. Moreover, the nodes were classified into specific themes and categorized into three major themes. These themes then were integrated to describe the inter-view of the opinion of the respondents about the hypothesis of the study.

Table 4.13

Coding Summary of the Text of Responses of Faculty Interviews

Name	Description
English language in Pakistan	It encompasses the importance of the English language in the socio-political scenario of the country as well as the status of the English language in the country. The teaching/learning conditions regarding EFL/ESL.
Desire to learn English	There are a desire and an effort to improve English as a second language in Pakistan
English as Second language	English is the language of law and parliament in the country
English Teachers' incompetency	English Language teachers are not well trained
Factors affecting Language Learning	Socio-economic status of the learners effect they are learning the language
Functional approach of English in Socio-political scenario	English in its functional approach has dominated the official and political dimensions of the country
learners' inconvenience	The learner feels inconvenient to learn the foreign language in an indigenous environment

Name	Description
Medium of teaching	Mostly the teachers use a deductive approach with the usage of the mother or regional languages in the classroom
requirements of EFL & ESL in Pakistan	Obviously, English is the need of the hour as it has become an international language
Social disparities in English Language Learning	The elite class learners have more opportunity to learn English than medium or poor class
Status of English as an Official Language	English is the official language in Pakistan
Student background	Students who belong to a rural area and illiterate background are less proficient in English
Language Learning Approaches	There are various Language teaching/learning approaches in Pakistan
GTM with CLT approach	Mostly the teachers prefer CLT with GTM approaches
The idea of Corpus Driven Language Learning	It is an innovative approach in language learning. teachers show their consent to use this approach

Name	Description
Mixed-Method Approach	The eclectic approach involving GTM, ALM, CLT with modern approaches
Personal stance	Teachers' personal view regarding language learning approaches
Preference of Procedural Knowledge	All the respondents preferred the importance of procedural knowledge
Effectiveness of CDLL	Respondents showed the effectiveness of CDLL positively
Awareness of Corpus Linguistics	Being the faculty members of the Applied Linguistics the respondents were well aware of the corpus linguistics
CDLL meets the requirements of International level	Corpus driven language learning meets the requirement of the international level of language learning among university undergraduates.
Corpus tools for Pedagogical purposes	Most of the participants were of the vies that corpus tools could be helpful in language teaching/learning
Awareness of Corpus Linguistics	The user of the corpus tool must have the mastery of corpus linguistics

Name	Description
syllabus contents & DDL	The respondents were of the view that Data-Driven Learning should be included in English Language Teaching Methods and approaches
Teaching Grammar & Data-Driven Learning	Grammatical items could be well taught through DDL
Idea	The participants strongly believed that it might prove a great aid to learning among university learners. In addition to language learning, it might have added benefits of sharpening learners' exploratory skills and their skills of inductive generalization.
Teachers' Training and CDLL	For Corpus Driven Language Learning approach demands well-trained teachers equipped with computer software programs

The hierarchy of the coding of text was explored from NVivo 12 as shown in figure 4.6.



Figure 4.7. Hierarchy of the coded text of the responses of the faculty interview

As per appendix the summary of the coded text of the responses obtained from the faculty of the Department of Applied Linguistics G C University Faisalabad, three major themes were generated from the views of the participants.

In the context of the English language learning in Pakistan, teachers were of the view that it varied from region to region. The learners belong to the remote and underprivileged areas have less opportunity to improve the English language as compare to the urban areas. Moreover, the English teachers of the country are not well trained to cope with the demand of the modern age. They generally use regional language in the English classroom focusing on the contents of the syllabi rather than the objective of the curricula. No doubt, the learners have an urge to have proficiency in this international language. Overall, the status of English in Pakistan needs to be upgraded with modern techniques.

As far as the second theme of the text is concerned, that is the language learning approaches, the participants were of the view that since the emergence of Pakistan, teachers had been using GTM followed by CLT while teaching English which was the major cause of the lack of fluency among the learners. They supported the inductive method for the promotion of procedural knowledge with modern approaches like data-driven learning.

Finally, the participants gave their view of the core issue of the study regarding corpus-driven language learning. The text search query revealed this core theme as shown in figure 4.5. All the participants who were approached for the interview had well awareness of corpus linguistics. They claimed to have compiled corpora during the teaching/learning process. They supported corpus tools for pedagogical purposes but no one claimed to apply this method while teaching in the classroom. However, they were appreciated to revise the course syllabus of ELT method & approaches with the inclusion of CDLL in the course contents.

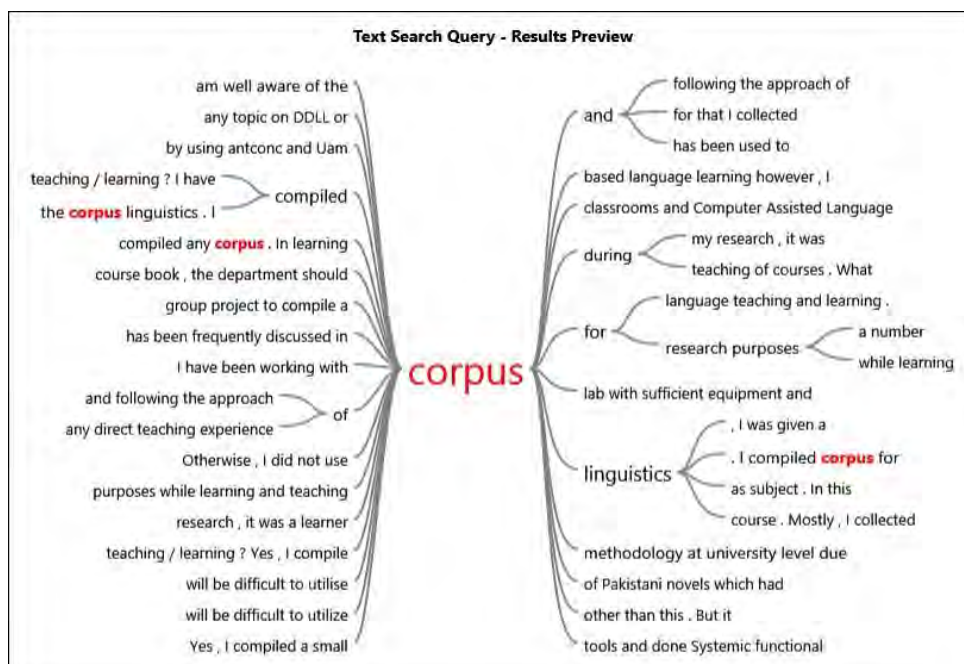


Figure 4.8. Text search query of the data obtained from the responses of faculty members.

The clustered items were also explored from the software and presented in the following figure 4.8 which shows the respondents' views regarding the strong association of language learning approaches with corpus-driven language learning in the country.

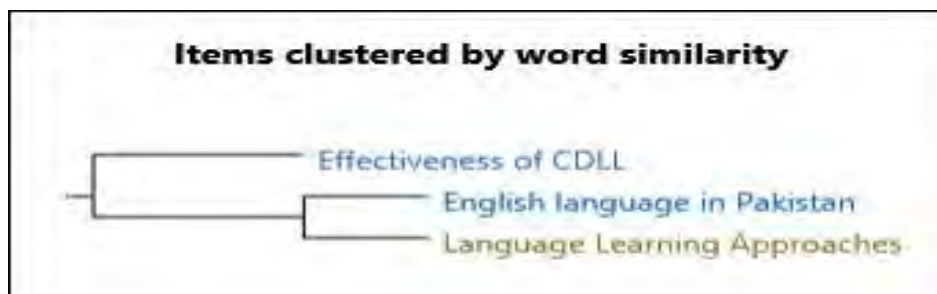


Figure 4.9. Item clustered with word similarity of the text of interviews of the faculty member

4.3.4. Thematic Analysis of Interviews

The focused interviews were conducted among the participants and the faculty members of the Department of Applied Linguistics GC University Faisalabad Pakistan. Then the text of the interviews was operated upon NVivo12 to generate codes and then codes were classified into themes.

Common themes were generated from the data of both interview groups as shown in table 4.8 and 4.9. as well as the figures generated by the software were also identified that indicated the favorable opinion of the interviewees regarding corpus-driven language learning. The learners, as well as teachers, showed their dissatisfaction with English language learning in Pakistan; however, they were of the view that procedural knowledge was more effective in the case of foreign language learning. Moreover, all of the interviewees unanimously argued that CDLL was an effective approach in English language teaching/learning with the well-equipped institution under the supervision of corpus expert teachers.

4.4 Discussion on the Results

The main findings from the study can be summed up into two areas: Effectiveness of CDLL on language teaching/learning and teachers' as well as learners' attitude towards corpus-driven language learning. To discuss the effectiveness of CDLL on language teaching/learning, the declarative and procedural knowledge of the research participant based on the expertise of lexical collocation using the corpus-driven language learning was

compared with the traditional method of seeking knowledge while using text and dictionaries.

4.4.1 Declarative Knowledge of the Participants

The participants' declarative knowledge of the language chunks, which is familiarity with the lexical combination (collocations) in general and frequently used in language by native and nonnative speakers was evaluated in the multiple-choice and fill in the gap parts of the pre and posttests. The statistical analysis of these parts of the tests showed that both groups enjoyed some improvement which might be interpreted as an instructional influence during the semester.

Table 4.10 shows the comparison between the results of the MCQ items attempted by the control and experimental group after pre and post-tests. The significant difference between both of the groups shows that the students of CDLL presented more scores which indicated the positive effects of the emerging approach of corpus-driven language learning in language teaching/learning.

What I have concluded at this point is that even if the teacher of language class or syllabus designer aims at familiarizing the learners with a lexical pattern like collocations to certain language points during teaching, the CDLL technique and material can be considered at least a viable alternative instructional method.

4.4.2 Procedural Knowledge of the Participants

Procedural knowledge is an adequate measure to empower the students to convert their interest into knowledge (Fakharzadeh, Youhanaee, & Nejadansari, 2013). In this study, CDLL was demonstrated to be an effective means of activating and converting interest into performances of understanding. As for the procedural aspect of research participants' language development, they have undergone the process of pattern-hunting and pattern-defining phase; both have been discussed in the following paragraphs

4.4.2.1 Pattern-hunting

Seven sub-groups of the experimental group did the pattern-hunting of seven types of lexical collocations respectively. Each group looked at the prescribed patterns of the lexical combination from both of corpora under investigations as described in the phase of the methodology. They congenially did the pattern hunting based on noticing hypothesis theory (Flowerdew, 2015). It has been noted that in this procedural phase students developed a certain interest which converted into the performance of understanding. The patterns of lexical collocations which were used by the non-native speakers/writers and were not accepted by the native speakers were embossed into the permanent memory of the participants as they processed upon those lexical combinations which were latterly described by them in the phase of pattern defining.

4.4.2.2 Pattern -defining

After having pattern-hunting of specific lexical collocations, the participants subsequently

discussed the uneven patterns among the whole class in the form of presentation. They successfully did pattern-defining. Earlier, the participants mentioned the reasons of mis-collocation based on the model of co-occurrence of restrictions propounded by Sinclair and Allerton (1984); later, they exemplified these mis-collocations concerning the influence of cultural and regional languages as caused those occurrences which led to the element of foreignness in the language usage of non-native English learners.

The impact of the procedural knowledge upon the participants has been revealed in the results of the post-tests the participants of the experimental group showed good performances as compared to the control group who sought the knowledge of collocations conventionally from books and dictionaries along with other written materials. Moreover, when data was collected qualitatively through semi-structured interviews, all the participants strongly recommended knowledge-seeking through the procedural way. The faculty members and the students both favored that process of learning as it was the best method with which knowledge could be retained permanently by the learner himself.

It has been sharply observed that the participants took keen interest while processing the corpora on software. They developed a taste of learning formulaic chunks of the foreign language. The sentence structure of the text of interview responses showed the usage of combined lexicons with longer expressions. The feature was discussed with the faculty of the department of applied linguistics GC University Faisalabad Pakistan; they not only

appreciated the CDLL approach but also recommended to revise the course outline of ELT Method & approaches being taught in the department.

4.4.3 Learners' Attitude towards CDLL

The second question in the research study sought the learners' attitude towards corpus-driven language learning as an instructional method in the case of English language teaching/learning. An overall analysis of the feedback of the responses received from the participants revealed a positive attitude towards the effectiveness of CDLL in foreign language teaching/learning. The text of the responses of the interview questionnaire showed that the participants were new to the CDLL. Although they were getting introduced to the new approach of inductive learning, yet they showed an acceptable level of satisfaction with the method, they involved in seeking knowledge of lexical collocations through CDLL.

The attitude of the learners was tracked down in analyzing the interviews of 10 randomly selected participants among the experimental group. As it was shown in the summary section of the analysis table 4.11; the participants expressed their dissatisfaction with the English language teaching/learning approaches being used in Pakistan. They belonged to various regions of the country therefore, the participants of remote and underprivileged areas seemed quite dissatisfied with the teaching/learning approaches.

NVivo-12 software was used to generate the codes from the text of the interview responses. Seven questions were formulated according to the semi-structured interview design propounded by Stienar Kvale (2007). 31 codes were generated which later on classified into 3 themes including the level of English in Pakistan, ELT approaches, and Effectiveness of CDLL in English language teaching/learning. Moreover, the word cloud obtained from the text of the responses of the interview showed the maximum frequency of Corpus Driven language learning with Language learning in Pakistan that indicated the preference of this new method on the part of learners.

All of the participants, belonging to urban and rural areas, opined that whether they attended English medium schooling or Urdu medium they never felt like knew English the way it should be. They have never had the opportunity to communicate in English their routine affairs. However, they had used the English Language in the academic perspective under stressed and restricted conditions.

After having a detailed discussion of inductive learning with modern approaches like CDLL the participants were provided with the opportunity to process and derive knowledge from Adhoc corpora, all the participants strongly recommended the inductive approach for promotion of procedural knowledge and preferred the CDLL. The participants claimed that the corpus gave them the exposure of authentic language, the language used by a native speaker in natural context; they could develop a natural familiarity with the multiple exposures of one word-combining with other words. Hence, they could learn such

word combinations (collocations) used by the native speakers and promote their fluency along with accuracy.

Few participants criticized by laying much emphasis upon the training of ELT teachers and non-availability of language furnished labs in the institutions particularly, at the intermediate and university level in Pakistan. As the participants expressed their dissatisfaction with the traditional method of English language teaching/learning, they demanded the refresher courses for the language teachers who must be acquainted with the modern approaches and software for the promotion of procedural knowledge with inductive approaches. Moreover, It has also been found that the syllabus contents of the ELT courses needed to be revised with the exclusion of traditional methods and inclusion of modern ones with more emphasis upon the inductive learning with student-centered activities.

4.4.4 Teachers' Attitude towards CDLL

To look for the answer to the third question of this research study, the opinion of 10 faculty members of the department of applied linguistics GC University Pakistan was sought through online semi-structural interview. Although the faculty members of the department were already equipped with the knowledge of corpus linguistics and data-driven learning yet they were not habitual to practice CDLL approach in the foreign language pedagogy.

As per the general view of the faculty members regarding English language learning status in Pakistan, although the English language held a special prestige, hardly enjoyed by any of the other languages spoken by the citizen of the country yet the non-native Pakistani speakers did not use English in their routine matters. Those who belonged to upper strata had the opportunity to attend elite class schooling might use English in their academic affairs the rest of middle and poor class learners were deprived of this opportunity altogether.

All of the participants unanimously preferred CLT method in English teaching/learning. However one of the faculty members claimed to prefer an eclectic approach with mixed methods of language teaching. Similarly, they believed in the development of procedural knowledge instead of declarative knowledge in the case of ELT. At the same time, few criticized that in the existing scenario the contents of the syllabus provided with the students, and the examination pattern did not support the procedural knowledge. It has shown a sense of dissatisfaction with the existing syllabus contents which need to be revised with the incorporation of certain items that help to promote the procedural knowledge.

As far as the direct question of the interview was concerned, the participants of the interview showed their consent in favor of data-driven learning which led to CDLL. They were of the view that DDL was an excellent method for inductive teaching/learning and a better way for engaging learners' mind and triggered critical thinking. The faculty members

agreed to incorporate CDLL in the syllabus contents of ELT Method & Approaches being introduced in the Department of Applied Linguistics GC University Faisalabad. However, they also demanded a well-furnished language lab along with computer proficient staff for smooth pedagogical intake and outlet.

4.5 Summary

This chapter addresses the research questions. In 1st question effect of corpus driven language learning (CDLL) was investigated and in 2nd and 3rd questions the opinion of the participants was sought in this regard. The 1st question was addressed through experimental method, significant differences between the results of both experimental and control groups before treatment (pre-test) and after treatment (post-test) were called for ANOVA in comparing certain features to prove the hypothesis. The items selected for pre and post tests were based on item analysis theory. The second and third questions of the research study were addressed by qualitative approach; focused interviews were organized to seek the opinion of the selected participants about CDLL. The text of the answers of the questions were processed on NVivo to generate word cloud which helped to form nodes and then these nodes were classified to generate three major common themes regarding language learning in Pakistan and approaches as well as the effect of CDLL.

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter begins with the summary of the findings of this research study in relation to the three research questions. The study conducted on the effectiveness of corpus-driven language learning on teaching collocation to university undergraduates of Pakistan. In addition to quantitative analysis of the pre and post-tests regarding the evaluation of collocational knowledge of experimental and controlled groups, the learners, as well as teachers' attitudes towards the CDLL approach used in the treatment of experimental group, was also observed through focused interviews. Simultaneously, the recommendations on the findings are also elaborated. Then, the implications of the research study on various aspects of teaching and learning has been described. Finally, the limitations, confronted during the research work has been discussed at the end of the chapter.

5.2 Findings for Research Question 1

It was found that the participants who have undergone the treatment of CDLL approach were in a better position in developing their knowledge in the context of lexical collocation than their fellow beings in non-CDLL treatment. The results generated before experimentation and after experimental treatment has shown significant difference, based on inferential statistics, as discussed in the chapter four. It shows that the university undergraduates using CDLL approach can develop their knowledge of language elements

like collocations, frequently used in connecting ideas explicitly and implicitly, easier than their counterpart learners using the conventional material and teachers' presentations.

The second finding concerns the development of procedural knowledge of the learners on the grounds of the usage of correct collocations as accepted by the native speakers. The results show that CDLL participants have improved their 'language use' features more than that of non-CDLL group. 'Language use' features are defined as linguistic constructions, word order, and linguistic accuracy. These findings has been supported by the results obtained from the post-test of the experimental treatment.

Thirdly, the participants of the experimental group showed their keen interest while seeking procedural knowledge through the inductive approach of language learning. They did pattern-hunting and then pattern-defining to maintain the durability of knowledge. Therefore, there is a significant difference between the results of the part of procedural knowledge between the experimental and controlled groups as compared to the declarative part of the tests.

I believe one of the goals of the course (ELT Method & Approaches) which seems to be the development of learners' micro-level skills in lexical knowledge has been met. They have developed a taste of vocabulary enhancement through CDLL approach. The learners as well as teachers have recommended the need for procedural knowledge in the case of EFL/ESL teaching/learning.

5.3 Findings for Research Question 2

The second research question was to elicit learners' attitudes towards CDLL approach, It was shown in 4.4.1 there exists a generally positive attitude towards the CDLL approach in teaching and learning a foreign language. The participants seem to be contented while participating in CDLL classroom activities, satisfied with the amount of learned material, showed their curiosity in playing with the software to do pattern-hunting, and thought the method was new and innovative. None of the interviewees expressed their lack of interest in being a part of CDLL classroom activities and they did not know anyone of their classmates who might have complained against the new CDLL approach. The answers of the interview questions revealed that CDLL approach can have at least one advantage over the method of teachers' presentation to inculcate knowledge and conventional method of consulting textbooks, to provide the learners with direct exposure of language and get them involved with learning pattern of language through autonomous independent learning. Therefore, I would like to suggest, in ELT classroom teacher should use CDLL approach with appropriate instructional materials and after familiarizing the learners with the way(s) of pattern-hunting and pattern-defining, their visual intelligence can be exploited to enhance their vocabulary knowledge. The linguistic chunks out of concordance or n-gram extractions and another format of CDLL-based material, can easily be made a part of learners' memory with the little assistance of teachers as facilitators. I think teachers who are using the CDLL approach should function as learning consultants in the classroom.

5.4 Findings for Research Question 3

Likewise, the faculty of the Department of Applied Linguistics has also strongly recommended the incorporation of CDLL approach among the course contents of ELT Method and Approaches being taught in the department. The board of study unanimously approved the course contents based on CDLL approach to be introduced in language teaching/learning activities.

A few criticisms, however, were made by one or more interviewees, mentioning the non-availability of the furnished language labs in the university as well as lack of computer expertise might prove a handicap in the promotion of CDLL approach in the language classroom. Someone has also complained against the lengthy course contents which need more time to be covered and CDLL approach itself is time-consuming. One among the points of concern for some faculty members was the educational background of the language learners. The point, they were trying to make, was the cultural tendency of the learners who were accustomed to being dependent on teachers and other sources, they were not used to learners' autonomy in the case of language learning. Nevertheless, the overall attitude of the faculty was in favor of the development of procedural knowledge with an inductive approach in the case of EFL teaching/learning as the evidences sought from the thematic analysis of the opinion generated by the faculty members and learners through interviews. Nevertheless, the study confirms certain theoretical and practice implication which have been described in the following segments.

5.5 Implications of the Research Study

The implications of the study can be examined on two facets, the theoretical as well as practical areas. At the theoretical level, some areas of corpus-driven language learning should be undergone further scrutiny and also the extent to which study could have contributed to the conceptualization involved as presented in the following segments. As far as the practical implications are concerned, there are more objective impacts of using this new approach (CDLL) in language teaching/learning at morphological, lexical, syntactical, and structural levels.

5.5.1 Authenticity of Material

One of the salient features of CDLL approach is the usage of authentic material as linguistic input in which learners are involved. This arises from the fact that CDLL typically involves, exposing learners to large quantities of authentic data so that learners can play an active role in exploring the language and looking at certain features like collocations and other language chunks in it. Gilquin and Granger (2010) have proved that ‘not only do corpora make it possible to expose learners to authentic language, but they can present them with a large number of authentic instances of a particular linguistic item and this ‘condensed exposure’ (Gabrielatos, 2005: 10) can contribute to vocabulary expansion or heightened awareness of language patterns’ (2010: 359). As it was explained in the methodology chapter, the sources of collocations investigated in this study were from two learner corpora, COPLNE & COBLENE; both corpora belonging to authentic language products not for teaching purpose but for some kind of real communication between language learners and their teachers at inter-language and intra-language levels. I decided

to choose these two corpora to take advantage of ascertained by previous researchers. It has been observed that the corpora is extremely useful for form-focused instructions like lexical combinations (collocations) because these present learners with typical inter-language features, especially, when data were produced by learners from the different mother tongue background (Gilquin & Granger, 2010b).

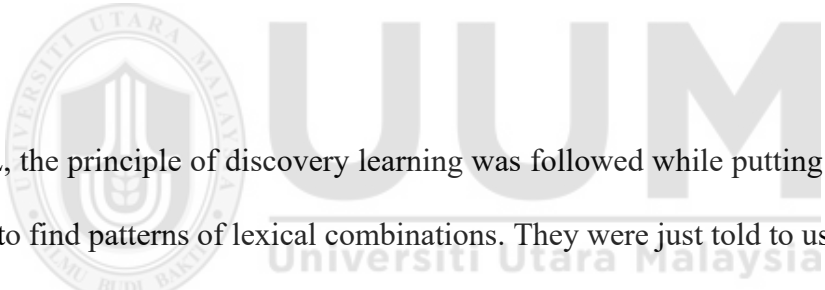
5.5.2 Discovery learning

The theoretical grounds of CDLL have paved the way for discovery learning that makes learning more motivating and stimulating as a fun, as an advantage for approach (Gilquin & Granger, 2010b). While discovering certain features, learners become researchers or detectives (Johns, 1991). However, in the presence of certain factors such as language labs, teachers' point of view, learners' level of proficiency, and the contents of CDLL, two types of CDLL activities may be used in an instructional context: teacher-oriented and learner-oriented activities. Gabrielatos (2005) makes a distinction between the two in the following paragraph:

"At the teacher-centered end, the teacher decides on the aims of the lesson, selects/designs the materials, and manages the lesson. At the learner-centered end, the learner decides on all three, with the teacher or computer program acting as facilitator and guide. Of course, there can be intermediate combinations, particularly when decisions are taken collaboratively between teacher and learners"

(Gabrielatos, 2005: p.12).

In this research study, the teacher-led end of the continuum was taken into account for CDLL based material preparation. The main reason was that the learners were not familiar with the autonomous learning procedure and strategies as required for learner-led activities. Bernardini (2004: 22) believes that the kind of activities in which learner borrows large and varied text collection in open-ended exploratory ways, would be called discovery learning. I think the learning situation in which learners think, guess and hypothesize and ultimately retrieve a rule out of the corpus they have been exposed to, could also be named 'discovery learning'. This is because even though the teacher is already familiar with the language pattern like collocation, this is the learner whom his teacher intended, should discover.



In CDLL, the principle of discovery learning was followed while putting the learners in a position to find patterns of lexical combinations. They were just told to use n-gram to find the lexical combinations; they discovered the combinations and compared with the BNC or Oxford collocation dictionary and found mis-matches. They not only came to know the concept of collocation but also developed a curiosity to learn the vocabulary in a sequential way that led towards the development of fluency.

John (1991) in his earlier work suggested that DDL attempted to cut out the middleman as far as possible and gave the learners direct access to data. Later, the researchers proved the active role of teachers which did not manifest itself necessarily in the classroom. The teacher's role got its shape specifically at material preparation or fixing the objectives of

learning. Gabrielatos (2005) suggests an intermediate combination in which decisions are taken collaboratively between teachers and learners. The CDLL approach enjoyed a new metaphor of ‘learner as junior researcher’ or ‘learner as co-detective’ as the learners’ research on language points decided by the teacher based on pre-selected material. In the present study, the teacher has decided to motivate the learners with the investigation of lexical combination (collocations) that were discovered by the participants themselves and later on described their features accordingly.

5.5.3 Noticing Hypothesis

The noticing hypothesis has been applied in the present study while learning language is of two folds:

- I. Learners’ identification of lexical combination (pattern-hunting)
- II. Learners’ comparison between their inter-language and model target language (pattern-defining).

As for the first area the learners’ attention was drawn to linguistic features as input. The participants were asked to prepare adhoc corpora of newspaper editorials from British and Pakistani leading newspapers. Then they were divided into groups to find the relevant patterns of lexical combinations assign to them. In this phase, they worked based on the noticing hypothesis (Schmidt, 2010). The immediate aim of CDLL as noticing activity in the n-gram tool of AntConc software was the identification of lexical patterns (collocations). Learners’ attention was drawn to the lexical pattern of the linguistic feature of the language and enabled them to use them correctly (as tested in the testes).

The results of comparing the groups' progress in the first part of post-test showed that CDLL could help the learners better than those who did not follow CDLL approach while identifying target linguistic features. Besides, we saw that the experimental group got more proficiency in the lexical patterns during the phase of pattern-defining. The sub-groups of the experimental group well defined the lexical collocation and described the reason for mis-collocation produced by the non-native language users.

The obtained results, I think, are in line with the results obtained by Schmidt & Frota (1986) and Doughty (1991) providing more practice evidence for the role of noticing in learners' production. Doughty (1991) used the high lightening technique and capitalization of text on the computer in the saliency of target forms, whereas, in this study, the extraction of lexical patterns from corpora through n-gram tool of AntConc presented the hidden target lexical patterns explicitly in helping learners to acquire the habitual juxtaposition of one word with respect to the other.

5.6 Practical Implications of the Research Study

As far as the practical implications of this research is concerned, it can be discussed in the following sessions:

- Development of Collocational Knowledge
- Using Corpora
- Training Learners
- Preparing CDLL materials

5.6.1 Development of Collocational Knowledge

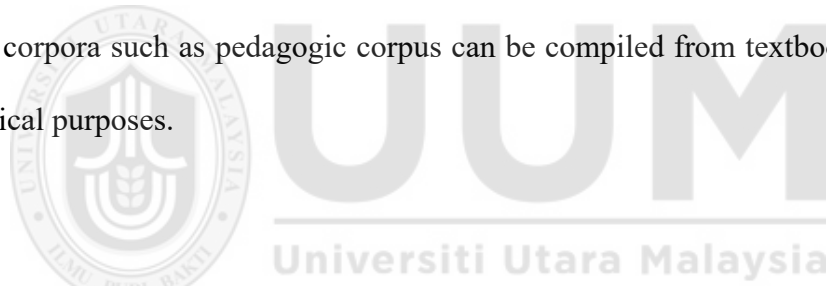
The present study aimed at examining students' performance regarding their knowledge about lexical collocation with the comparison between two groups. Having reviewed different theories of collocation and approaches to language learning particularly, data-driven language learning, a structural approach which was adopted to guide the students to look into lexical collocation from the corpora.

To evaluate the effectiveness of CDLL approach in language learning, the learners' performance was evaluated in terms of declarative as well as procedural knowledge about lexical collocations. Regarding declarative knowledge, data derived from corpora could have helped the learners to develop familiarity with certain structures as pointed out by the teachers to the students. As far as the procedural knowledge is concerned the learners were provided with the choices to match the exact collocation with the given words. As during the process of knowledge they sought from corpora, the learners recalled the process of CDLL and select the appropriate collocation with the required word.

In short, one of the goals of the course which is the development of learners' micro-level skills in knowing collocations has been achieved. Whereas, the learners have also developed a taste of vocabulary enhancement at the macro level which engenders their fluency as well. The results would reveal that CDLL would be viewed a teaching technique that language teachers can apply along with appropriate tasks to teach the foreign language so that students become active and autonomous learners in the classroom.

5.6.2 Using Corpora

To adopt a CDLL methodology, one of the choices is the selection of corpora from which the linguistic features can be retrieved. According to Gilquin & Granger (2010), ‘any type of corpus can be used as CDLL material, and indeed, the literature on DDL mentions quite a large number of corpora: written or spoken or multimodal, monolingual or bilingual general or specific, native or non-native, and tagged or un-tagged, etc. However, particular corpora are best suited for a particular purpose and based on learners’ proficiency level, the goal of teaching a course, and institutional capabilities. For instance, the use of learner corpora as used in the present research, for an instructional material such as introduction to lexical collocation to improve learners’ collocational knowledge at micro-level; other types of corpora such as pedagogic corpus can be compiled from textbooks and used for pedagogical purposes.



5.6.3 Training Learners

Learners’ training is inevitable in the case of CDLL approach to foreign language teaching/learning. For the novice EFL learners who have no experience of corpora and the software to operate these corpora for the extraction of certain linguistic features, they need to be trained on how to compile, annotate and retrieve the corpora under investigation. Moreover, a computer expert must impart necessary training to overcome the barriers which may create hindrance in the way of learning. In this regard, teachers should provide necessary training to their students how looking at concordance lines and make them aware that in the case of CDLL approach, students should not read the concordance lines from left to right but it should be read from top to bottom vertically in which they should try to

search what happens around the keyword in the middle of the lines, the node. Students would be taught to focus on the word span of 4 to 5 words before and after the keywords.

In the present study, the learners were asked to prepare adhoc corpora. Their training was started with a compilation of corpus data. Then they were asked to do pattern-hunting of certain lexical collocation with the necessary training of the usage of AntConc software. Whereas in the phase of retrieval they successfully did pattern-defining regarding the use of mis-collocations by the non-native Pakistani authors. Moreover, they discussed the reasons behind the production of such mismatches in the case of lexicon.

5.6.4 Preparing CDLL Material

As discussed in chapter 3, since ELT Method and Approaches course is 3 credit hour mandatory course for the undergraduates of the department of Applied Linguistic Government College University Faisalabad Pakistan, the syllabus outline was already approved by the board of study of the department. I decided to incorporate the new approach of corpus-driven language learning in the course outline of the said syllabus which was approved by the board of studies of the department.

For the first experience, the learners were taught lexical collocation through inductive learning. The revised course contents were shared among the members of the board of studied who congenially approved this approach and encouraged to continue this process for the teaching of other language features like paragraph development and other

grammatical chunks as investigated by the other former researchers in the fields of DDL. For instance, Sepehere (2015) decided to prepare DDL units according to one of the textbooks which were widely used in the university, that is, paragraph development; by Arnaudet and Barret (1990), a list of 18 grammatical patterns was prepared after analyzing the main topics for the DDL-based units. Similarly, the students of this study were acquainted with lexical collocations through CDLL positively.

Classroom concordancing formats, the mostly used, were intended to present the lexical relations like collocations with wide language exposure with immediate contexts to show language patterning. However, the longer language structure at a clause or even sentence level can also be worked out if the appropriate formatting is chosen. The CDLL based tasks prepare for this study aimed at the development of collocational knowledge among the learners, hence, the contents, format, and classroom tasks could be specified, designed, and assigned accordingly, nonetheless, it is a rather specific use of the method. Moreover, the language teachers can prepare their CDLL-based materials in general teaching context to make learners acquaint with different language patterns. I believe that appropriate materials can be prepared at the word and phrase levels and even at clause, sentence, and discourse levels as Sepehere (2015) used in his study. So, it can be suggested that CDLL materials could be used as supplementary exercises to the main textbook used in each course, or can be used as a general grammar and vocabulary exercises as part of a lesson, or in a self-access language Centre (Sepehri, 2015).

5.7 Limitations of the Study

Undoubtedly, every research work has its specific limitations, this one is no exception. One of the limitations with which I was dealing, was the sample of the study, the participants. Since it was impossible to have two groups of students from different universities who had the same level of English proficiency, I had to take the groups from the same institution because of the same course contents they were taught in the course of their study. However, the sampling belongs to the different regions and background could help to generalize the results of the study and maintain the reliability of the research study.

This research was conducted in the context of ELT approaches and methods, a course being taught at public sector university of Pakistan. The primary function of instructor was to teach and cover the syllabus contents and collecting data was the secondary function. On the other hand, students were attending a three-credit-hour course for which they had to appear for the final exam at the end of the term. The course syllabus had to be covered in both groups. And therefore it was impossible to compare the whole CDLL material with the non CDLL group. Only one aspect of the linguistic feature that was lexical collocations was introduced and practiced through CDLL approach because of the time limitations and the coverage of the other syllabus contents of the course of ELT Method and Approaches.

Moreover, the factor of morality in this experimental method was not properly controlled. There were complex interactions among the participants of the research; both of the groups could not have similar sessions with the teachers.

5.8 Summary

It can be concluded that the hypothesis, conceptualized on strong theoretical frame work has been scientifically and systematically proved that the modern technique like corpus driven language learning (CDLL) has positively affected the the language learning of EFL Pakistani undergraduates. Moreover, both the teachers and learners have shown their interest in the incorporation of CDLL in the course of ELT Methods and Approaches, recommended for the degree program of EFL university undergraduates. The research work has a positive theoretical and practical contributions in the research area of Applied Linguistics as it has endorsed the theories of modern linguists like Smith, Flowerdew, Boulton who have supported the autonomous learning of the foreign language learners as well as it has opened new horizons for the future researchers to apply the modern technique like CDLL for the development of larger language units as proved in this research work, the development of collocational knowledge among EFL/ESL learners. The researcher has met a primary objective to assure the pedagogical implication of his study among academia including syllabi designers and policy makers. They have been realized the worth of modern techniques (CDLL) in teaching and learning of the foreign language.

REFERENCES

- Akram, M. (2017). Learning and Teaching English in Pakistan: Predicaments and Solutions. *International Journal of Educational Sciences*, 19(1), 10–14.
<https://doi.org/10.1080/09751122.2017.1377917>
- Al-mahbashi, A. (2015). The Effect of Data Driven Learning on Receptive Vocabulary Knowledge of Yemeni University Learners. *3L: Language, Linguistics, Literature The Southeast Asian Journal of English Language Studies*, 21(3), 13–24.
- Almutairi, N. D. (2016). The Effectiveness of Corpus- Based Approach to Language Description in Creating Corpus-Based Exercises to Teach Writing Personal Statements. *English Language Teaching*, 9(7), 103-109.
<https://doi.org/10.5539/elt.v9n7p103>
- Arora, S. (2018). Item Analysis. . *International Journal of Nursing Science Practice and Research*, 4(2). 6-7. Retrieved from:
<file:///C:/Users/NEC/Downloads/itemanalysis.pdf>
- Bichi, A. A., Talib, R., & Development, T. I. (2018). *Item Response Theory : An Introduction to Latent Trait Models to Test and Item Response Theory* : Available at: <https://doi.org/10.11591/ijere.v7i2.12900>
- Binkai, J. (2012). An empirical study on corpus-driven English vocabulary learning in China. *English Language Teaching*, 5(4), 131–137.
<https://doi.org/10.5539/elt.v5n4p131>
- Boontam, P., & Phoocharoensil, S. (2018). Effectiveness of English preposition learning

- through data-driven learning (DDL). *3L: Language, Linguistics, Literature*, 24(3), 125–141. <https://doi.org/10.17576/3L-2018-2403-10>
- Boulton, A. (2008). Looking for Empirical Evidence of Data-Driven Learning At Lower Levels. Retrieved from:
http://C:/Users/NEC/Downloads/boulton_2007_PALC_looking.pdf
- Boulton, A. (2012). What data for data-driven learning? *EuroCALL Review*, 20(1), 23–27.
- Boulton, A. (2017). Data-Driven Learning and Language Pedagogy, retrieved from:
Language, Education and Technology, (January 2017). <https://doi.org/10.1007/978-3-319-02237-6>
- Buriro, A. G., Awan, J. H., & Lanjwani, A. R. (2017). Interview: a Research Instrument for Social Science Researchers. *IJSSHE-International Journal of Social Sciences, Humanities and Education*, 1(4), 1–14. Retrieved from:
<https://www.researchgate.net/publication/323278410>
- Celce-Murcia, M., & McIntosh, L. (1991). *Teaching English as a second or foreign language*. Retrieved from:
<http://www.tesle.org/wordpress/issues/volume5/ej20/ej20r4/7/>
- Channa, L. A. (2017). English in Pakistani public education. *Language Problems and Language Planning*, 41(1), 1–25. <https://doi.org/10.1075/lplp.41.1.01cha>
- Collentine, J. (2000). Insights into the construction of grammatical knowledge provided by user-behavior tracking technologies. *Language Learning and Technology*, 3(2),

44–57.

Costa, P. De, Plonsky, L., & Starfield, S. (2018). Applied Linguistics Research: Current Issues, Methods, and Trends from *The Palgrave Handbook of Applied Linguistics Research Methodology*. Springer link: <https://doi.org/10.1057/978-1-137-59900-1>

Cresswell, J. (2014). *Research Methodology-Qualitative, Quantitative, and Mixed Method*. SAGE Publication India PVT Ltd.

Demir, C. (2017). Lexical collocations in English: A comparative study of native and non-native scholars of English. *Dil ve Dilbilimi Çalışmaları Dergisi*, 13(1), 75–87.

DePaolo, C. A., & Wilkinson, K. (2014). Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data. *TechTrends*, 58(3), 38–44.
<https://doi.org/10.1007/s11528-014-0750-9>

Dörnyei, Z. (2003). Questionnaires in Second Language. In *Mahwah: Lawrence Erlbaum Associates*, from: <https://doi.org/10.1017/S027226310423306X>

Ellis, R. (2015). *Understanding Second Language Acquisition*. Oxford University Press.

Fakharzadeh, M., Youhanaee, M., & Nejadansari, D. (2013). Proceduralization, transfer of training and retention of knowledge as a result of output practice. *International Journal of Research Studies in Language Learning*, 3(2), 3–16.
<https://doi.org/10.5861/ijrsl.2013.431>

Fareed, D. M., Jawed, S., & Awan, S. (2018). Teaching English Language at SSC Level in Private Non-Elite Schools in Pakistan: Practices and Problems. *Journal of Education and Educational Development*, 5(1), 80-87.

<https://doi.org/10.22555/joeed.v5i1.1756>

Farghal, M. (2007). Lexical Collocations in EFL Writing . *The Journal of Asia TEFL*, Vol. 4, No. 2, 69-94.

Finfgeld-Connett, D., Finfgeld-Connett, D., & Johnson, E. D. (2018). Data Collection and Sampling. *A Guide to Qualitative Meta-Synthesis*, (January), 18–30.

<https://doi.org/10.4324/9781351212793-3>

Flowerdew, L. (2015). *Data-driven learning and language learning theories*. John Benjamins Publishing Company.

Fuentes, A. C. (2001). Lexical behaviour in academic and technical corpora: Implications for ESP development. *Language Learning and Technology*, 5(3), 106–129.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring Learner Language Through Corpora: Comparing and Interpreting Corpus Frequency Information. *Language Learning*, 67(June), 130–154. <https://doi.org/10.1111/lang.12226>

Gabrielatos, C. (2005). Corpora and Language Teaching: Just a Fling or Wedding Bells?. *The Electronic Journal for English as a Second Language, Tesl-Ej*, 8(4), from: <https://files.eric.ed.gov/fulltext/EJ1068106.pdf>

Gardner, H. (1983). *Frames of Mind: Theory of Multiple intellegences*. Routledge, London.

Gilquin, G., & Granger, S. (2010a). How can DDL be used in language teaching? *The Routledge Handbook of Corpus Linguistics*, Routledge, London,(May), 359–369.
from:file:///C:/Users/NEC/Downloads/How_can_datadriven_learning_be_used_in_1

anguage_t.pdf

Gilquin, G., & Granger, S. (2010b). How can DDL be used in language teaching? *The Routledge Handbook of Corpus Linguistics*, Routledge, London (January 2010), 359–369.

Godau, R. (2014). NVivo qualitative data analysis software; QSR. *International*, (July).
From: <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/support-services/nvivo-downloads>

Government of Pakistan. (2018). *Pakistan - National Education Policy Framework November 2018*. 12.
https://aserpakistan.org/document/2018/National_Eductaion_Policy_Framework_2018_Final.pdf

Granger, S. (1998). *Learner English on Computer* - Sylviane Granger - Google Books.
Retrieved September 16, 2019, from Routledge website:
https://books.google.com.my/books?hl=en&lr=&id=7z_KAgAAQBAJ&oi=fnd&pg=PA1&dq=Learner+English+on+Computer&ots=XjJ4arpguJ&sig=3CTfEpgFG-IIfunPA9jM0dSe76Y#v=onepage&q=Learner+English+on+Computer&f=false

Hadley, G. (2002). An introduction to data-driven learning. *RELC Journal*, 33(2), 99–124. <https://doi.org/10.1177/003368820203300205>

Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field Methods*, 18(1), 3–20. <https://doi.org/10.1177/1525822X05282260>

- Jalil, H. A., & Mcfarlane, A. (2008). Assisted Performance – a Pragmatic Conception of Online Learning. *International Journal of Instruction*, 1(2), 57–76.
- Jiao, S. K. S. C. . J. (2019). Technology in education: Pedagogical Innovation. Springer
<https://doi.org/10.1007/978-981-13-9895-7>
- Johns, T. (1991). Should you be persuaded: two examples of data-driven learning. *ELR Journal 4: Classroom Concordancing*. Birmingham: CELS, The University of Birmingham, 4(Farrington 1989), 1–16.
- Kehoe, J. (1995). Basic item analysis for multiple-choice tests. *Practical Assessment, Research and Evaluation*, 4(10), 1994–1995.
- Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning and Technology*, 14(1), 28–44.
- Kim, H. (2012). The Relationship of Collocation Competence with Reading and Writing Skills. *English Teaching*, 67(3), 95–119.
<https://doi.org/10.15858/engtea.67.3.201209.95>
- Kuiper, K. (2004). Formulaic Language and the Lexicon (review). In *Language Linguistics Society of America* (Vol. 80). <https://doi.org/10.1353/lan.2004.0209>
- Kurz, T. B. (1999). A review of scoring algorithms for multiple-choice tests. *Annual Meeting of the Southwest Educational Research Association* (San Antonio, TX, January 21-23, 1999, (1), 1–21.
- Kvale, S. (2007). *Doing Interview*. SAGE Publication India PVT Ltd.

- Lewis, M. (1993). *The Lexical Approach: The state of ELT and a Way Forward*. Commercial Colour Press, London E7.
- Lixun, W. (2001). Exploring parallel concordancing in English and Chinese. *Language Learning and Technology*, 5(3), 174–184.
- Lu, Y. (2016). *A Corpus Study of Collocation in Chinese Learner English*. Routledge, London
- Luo, Q., & Zhou, J. (2017). Data-driven learning in second language writing class: A survey of empirical studies. *International Journal of Emerging Technologies in Learning*, 12(3), 182–196. <https://doi.org/10.3991/ijet.v12i03.6523>
- Men, H. (2018). Vocabulary Increase and Collocation Learning. In *Vocabulary Increase and Collocation Learning*. <https://doi.org/10.1007/978-981-10-5822-6>
- Nesselhauf, N. (2005). Collocations in a Learner Corpus. In *John Benjamins Publishing Company* (Vol. 14). <https://doi.org/10.1192/bjp.111.479.1009-a>
- Raquel, V. A. & M. (2019). Quantitative Data Analysis for Language Assessment Volume I. In *Quantitative Data Analysis for Language Assessment Volume I*. <https://doi.org/10.4324/9781315187815>
- Sana Nawaz, Aisha Umer, Muffasira Tabasum, Maria Zaman, Anam Batool, S. A. (2015). Difficulties facing by students of L1 in adopting L2. *European Journal of English Language, Linguistics and Literature*, 2(2), 1–6.
- Schmidt, R. (2010). Attention , Awareness , and Individual Differences in Language Learning. *Proceedings of CLaSIC 2010*, 4, 721–737.

- Sepehri, M. (2015).. *The Effects of Data Driven Learning on Iranian Efl University of Birmingham Research Archive, (July)*. Retrieved from:
https://etheses.bham.ac.uk/id/eprint/6267/1/Sepehri15PhD_Redacted.pdf
- Shah, M. I. et al. (2020). Role of Collocation in Lexical Behavior of Foreign Language Learners : A Corpus-based Study of Pakistani English Learners. *International Journal of Advanced Science and Technology*, 29(02), 1955–1963.
- Shamim, F. (2013). English as the language of development in Pakistan: Issues, challenges, and possible solutions. *Language in Society*, 42(4), 291–310.
<https://doi.org/10.1017/S0022268313000045>
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Soiferman, L. K. (2010). *Inductive and Deductive Research Approaches*. University of Manitoba (April, 2010), 1–23. From: <https://files.eric.ed.gov/fulltext/ED542066.pdf>
- Tongco, M. D. C. (2007). Purposive sampling as a tool for informant selection. *Ethnobotany Research and Applications*, 5, 147–158.
<https://doi.org/10.17348/era.5.0.147-158>
- Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research*, 14(2), 103–135. <https://doi.org/10.1191/026765898674803209>
- Tu, H.-L., & Talley, P. C. (2016). Explicit vs. implicit second language instruction: concept and paradigm. *International Journal of Humanities and Social Science*, 6(4), 204–208. Retrieved from www.ijhssnet.com
- Vygotsky, L. S. (1999). Vygotsky 's Sociocultural Theory. *Education Psychology for*

Learning, (2007), 26925. <https://doi.org/10.1007/978-94-007-4065-5>

Wanner, L., et al. (2013). Annotation of Collocations in a Learner Corpus for Building a Learning Environment. *Computer Science*, (1998), 1–10.

Williams, G., & Millon, C. (2012). Palmer, Firth and Internet: Drawing together collocational threads. *Proceedings of Corpus Linguistics 2011*, 1–34.

Yaemtui, W. (2018). The Effectiveness Of Data - Driven Learning (DDL) on Teaching English Collocations to Thai Efl Students. *PhD Thesis at Thammasat University Thailand*. Retrieved from:
http://ethesisarchive.library.tu.ac.th/thesis/2016/TU_2016_5821042040_6907_4721.

pdf



APPENDICES

Appendix-A

LNG-403

ELT Methods & Approaches

Semester: BS 3rd English (Literature & Linguistics)

Course Description: The course introduces participants to some of the major trends in methods, approaches, assumptions and issues on the nature of language and how languages can be acquired and taught. It is relevant to English Language Teaching methods. It develops the students' pedagogical knowledge and the basic skills needed to teach/learn English efficiently. The course is divided into 2 main areas:-

1. Critical survey of the modern ELT Methods such as Data Driven learning
2. A variety of teaching techniques related to teaching the four language skills.

Aims

1. To increase students' awareness of basic concepts and principles related learning and teaching English in Pakistan
2. To provide students with a brief, critical survey of the main ELT approaches.
3. To expose students to Modern English language teaching techniques appropriate to their socio-cultural context and ESL/EFL.

4. To develop participants' skills in critically reflecting on their own, and others', language education experiences, in the light of relevant research, theory and current areas of debate in the methodology of teaching languages.

Intended Learning Outcomes

On completion of this course, successful participants will be able to demonstrate:

- I. a clear understanding of basic ELT terminology.
- II. an understanding of the basic theoretical background behind teaching and learning languages.
- III. a critical understanding of modern approaches and techniques involved in teaching EFL/ESL.
- IV. the ability and skill to plan and deliver a lesson to a class of EFL/ESL students.
- V. the ability to use DDL approach of teaching language skills, grammar, vocabulary, etc.
- VI. the ability to cope with challenging learning and teaching Pakistan in the current scenario.
- VII. the ability to reflect critically on learning experiences on the course and relate them to educational contexts they are familiar with

Course Description

Evaluation System:

➤ Quizzes/Presentations/Viva Voice	10%
➤ Assignment	10%
➤ Mid Term Test	30%
➤ Final Examination(at the end of the semester)	50%

Text Book:

Material & Method in ELT by Jo McDonough & Christopher Shaw

Approaches & Methods in Language Teaching by Jack C Richard

Teaching Methodology:

- Active participation in class
- Whole class discussion
- Pair and group work
- Giving presentations

Signature of Instructor

Appendix-B

Questionnaire 1

(administered at the beginning of the course)

1. General information

First name _____ Last name _____ Gender (m / f)

Age ☐ under 19 ☐ 19-22 ☐ over 22

Year _____ Faculty _____ Major _____

GPA ☐ less than 12 ☐ 12 – 13.9 ☐ 14 -15.9 ☐ 16-18 ☐ more than 18

2. How long have you been learning English?

☐ 1-3 ☐ 4-6 ☐ 7-9 ☐ 10-12 ☐ more than 12 years

3. Why do you learn English? (You may tick more than one)

☐ Because it helps me to get a good and well-paid job

☐ Because it helps me to prepare myself for studying abroad

☐ Because I like the language

☐ Because it helps me to be able to communicate with foreigners

☐ Because it is very widely spoken in the world today

☐ Because it helps me to pass the exams

☐ Because it helps me to get information or pleasure (e.g. reading English-language newspapers, watching English films)

☐ Because it is fashionable

☐ Because it helps me to learn about other cultures through English

☐ no particular reason

☐ other reasons (Please specify _____)

4. How many times have you been to countries where you have to use English language?

☐ never ☐ 1-2 ☐ 3-4 ☐ more than 5

5. What is the longest period of your stay in number 4? _____ months/years

6. How often do you do such activities?

[5 = very often 4 = often 3 = sometimes 2 = rarely 1 = never]

Questions	1	2	3	4	5
1. Read English-language newspapers, magazines, Internet information, or other English texts, outside classroom					
2. Speak English outside classroom					
3. Watch English movies or listen to English songs					
4. Go to language lab or self-study room after class					
5. Write letters in English or chat via the Internet in English					

7. How many English courses have you taken so far at your University (NOT including This semester)?

☐ 1-3 ☐ 4-6 ☐ 7-9 ☐ 10-12 ☐ 12 -15 ☐ more than 15

8. How many hours per week are you taking English courses outside your University

this semester?

☐ not taking any ☐ 1-2 ☐ 3-4 ☐ 5-6 ☐ 7-8 ☐ 9-10 ☐ more than 10

9. Before this semester, how often did your English teachers do the following things in class?

[5 = very often 4 = often 3 = sometimes 2 = rarely 1 = never]

Questions	1	2	3	4	5
1. Ask you to find missing words from context (e.g. cloze test)					
2. Ask you to guess unfamiliar words from context					
3. Ask you to remember lists of vocabulary and/or give dictations					
4. Ask you to observe grammar points by yourself from songs, newspapers, or other given examples					
5. Ask you to do grammar exercises (e.g. change verb tenses or change active voice to passive voice)					
6. Ask you to correct language mistakes by yourself Give and explain grammar rules					
7. Ask you to do translation exercises					

Appendix-C

Pre-test

This is a purely research oriented task. You are supposed to response earnestly and honestly as the response obtained naturally would validate the results.

Item. 1 Match the column

Sr No	Lexicon	Collocate
1	Help	view
2	Gorgeous	a fuss
3	Soar	sorry
4	Aloof	rain
5	extremely	upward
6	ceaseless	from
7	Make	desk

Item 2 Choose the natural collocation.

Q1 - Which is correct?

I have done a great deal of work.

I have done a big deal of work.

Q2 - Which is correct?

We're in severe straits.

We're in dire straits.

Q3 - Which is correct?

I've got a deafening headache.

I've got a blinding headache.

Q4 - Which is correct?

It wasn't much trouble- it was no great deal.

It wasn't much trouble- it was no big deal.

Q5 - Which is correct?

It is a fatal disease.

It is a mortal disease.

Q6 - Which is correct?

He's terminally ill.

He's mortally ill.

Q7 - Which is correct?

It's a small priority issue.

It's a low priority issue.

Q8 - Which is correct?

I like the environment of the restaurant.

I like the ambiance of the restaurant.

Q9 - Which is correct?

It was total chaos.

It was great chaos.

Q10 - Which is correct?

She was highly drunk.

She was extremely drunk.



Item.3 Tick the appropriate word suitable for the blank region of the sentence:

1. An open and friendly manner is ... of this region.
(a) Personality (b) characteristic (c) individuality (d) Character
2. The children have been as good as gold. So well-...
(a) Adjusted (b) intentioned (c) dressed (d) behaved
3. I thought his comment about my putting on too much weight was quite ...
(a) Tactless (b) cynical (c) sensitive (d) egotistical
4. My sister is really self-... She can join in a group without any introductions.
(a) Confident (b) controlled (c) absorbed (d) Centered
5. I read the newspapers every day because I like to be well-... about what's going on in the world.
(a) Known (b) Informed (c) Educator (d) brought-up
6. The ... paragraph gives the background to their research project.
(a) preceding (b) bygone (c) retrospect (d) beforehand
7. These are very ... times, and it may be a while before everything returns to normal.
(a) temporary (b) turbulent (c) contemporary (d) rapid
8. There has been a dramatic ... in the way we see modern society.
(a) shift (b) amendment (c) alteration (d) variation

Participant-----

Appendix-D

Post-test

This is purely a research oriented test. Your earnest response would help to validate the results of the research study. Your response would be mentioned anonymously in the document.

Item 1 Gap Filling

Select the proper collocation from the parenthesis and fill the gap

1. It's a bit of a sticky ___. I don't know what to do. (Area, situation)
2. I think we've made a terrible ___. (Judgment, Mistake)
3. I think we've had a very lucky ___. (Chance, escape)
4. I have a gut ___ that this project will be unsuccessful. (Feeling, instinct)
5. Phew! That was a close ___ ! (Edge, shave)
6. We need to talk about the thorny ___ of restructuring the company. (Issue, danger)
7. He suffered a severe ___ in his ambitions. (Obstacle, setback)

8. You'll need a strong survival ___ to do well in this company. The office politics are terrible! (Bone, instinct)
9. There was a tense ___ between the police and the protesters. (Fight, stand-off)
10. It's not easy knowing what to do. We're in uncharted ___ here. (Land, territory)

Item 2 MCQ

Tick the correct collocation among the options given below the each sentence.

1. The _____ majority of rain in Spain falls mostly on the central plain.
- a) Big
 - b) vast
 - c) most
 - d) far
2. The fact there is a _____ gap between the rich and poor in my country is worrying
- a) Climbing
 - b) Lengthening
 - c) Widening

d) soaring

3. In _____ years we have experienced a rise in the number of cars on the road.

a) Last

b) Latest

c) Past

d) Recent

4. I was _____ disappointed when I failed my driving test.



a) Bitterly

b) Strongly

c) Forcefully

d) largely

5. Many _____ species can be found on the grasslands and in the jungles of Africa.

a) Risky

b) Imperiled

c) Endangered

d) unlucky

6. The internet added a _____ to crime.

a) New area

b) New dimension

c) New zone

d) New criminality

7. Darwin and Wallace were the first scientists to _____ the theory of evolution.

a) Predict over

b) Made up

c) Plan out

d) Put forward

8. The police said that the _____ details of what happened are still not available.

a) Precise

b) Really

c) Best

d) Exactly



Item 3. Match the column:

1	genuinely	verdict
2	effective	words
3	possible	mob
4	information	inclusion
5	significant	release
6	weasel	amendment
7	marriage	partnership
8	adverse	concerned
9	lynch	breakdown
10	news	desk

Participant-----

APPENDIX-E

Adjective+			Noun				
#Total No. of Cluster Types:			#Total No. of Cluster Types:				
5292			4469				
ENG			PAK				
#Total No. of Cluster Tokens:			#Total No. of Cluster				
6201			Tokens: 5403				
1	43	1	prime_jj minister_nn	1	51	1	prime_jj minister_nn
							human_jj
2	28	1	human_jj rights_nn	2	27	1	rights_nn
			general_jj				
3	19	1	election_nn	3	16	1	current_jj account_nn
							international_jj
4	12	1	labour_jj party_nn	4	16	1	community_nn
			prime_jj				
5	12	1	ministers_nn	5	15	1	prime_jj minister_nn
6	11	1	young_jj people_nn	6	13	1	afghan_jj taliban_nn
			minimum_jj				maternal_jj
7	10	1	wage_nn	7	13	1	mortality_nn
			prime_jj				social_jj
8	10	1	minister_nn	8	13	1	media_nn
							economic_jj
9	9	1	living_jj wage_nn	9	12	1	growth_nn
			european_jj				chinese_jj
10	8	1	union_nn	10	11	1	investment_nn

							new_jj
11	8	1	free_jj speech_nn	11	11	1	plot_nn
			greek_jj				new_jj
12	8	1	economy_nn	12	11	1	year_nnt
							ukranian_jj
13	8	1	tory_jj party_nn	13	11	1	city_nn
							indian_jj
14	7	1	gay_jj rights_nn	14	10	1	government_nn
							foreign_jj
15	7	1	private_jj sector_nn	15	9	1	policy_nn
							other_jj
16	7	1	public_jj sector_nn	16	9	1	hand_nn
							public_jj
17	7	1	public_jj services_nn	17	9	1	sector_nn
18	6	1	african_jj union_nn	18	9	1	supreme_jj court_nn
19	6	1	civil_jj liberties_nn	19	8	1	chief_jj minister_nn
							provincial_jj
20	6	1	civil_jj war_nn	20	8	1	government_nn
			criminal_jj				supplementary_jj
21	6	1	justice_nn	21	8	1	reference_nn
			economic_jj				afghan_jj
22	6	1	credibility_nn	22	7	1	government_nn
			financial_jj				concerned_jj
23	6	1	support_nn	23	7	1	authorities_nn
24	6	1	foreign_jj policy_nn	24	7	1	global_jj warming_nn

							joint_jj
25	6	1	greek_jj people_nn	25	7	1	fund_nn
26	6	1	islamic_jj state_nn	26	7	1	maternal_jj health_nn
							chief_jj
27	6	1	legal_jj aid_nn	27	6	1	justice_nn
							cold_jj
28	6	1	only_jj thing_nn	28	6	1	war_nn
							foreign_jj
29	6	1	public_jj health_nn	29	6	1	office_nn
			public_jj				govt_jj
30	6	1	spending_nn	30	6	1	protests_nn
							interim_jj
31	6	1	single_jj market_nn	31	6	1	president_nn
							national_jj
32	6	1	social_jj security_nn	32	6	1	assembly_nn
33	6	1	supreme_jj court_nn	33	6	1	national_jj saving_nn
			universal_jj				
34	6	1	credit_nn	34	6	1	political_jj parties_nn
35	5	1	black_jj people_nn	35	6	1	prime_jj ministers_nn
							private_jj
36	5	1	central_jj bank_nn	36	6	1	partnership_nn
			chief_jj				
37	5	1	executive_nn	37	6	1	private_jj schools_nn
			conservative_jj				private_jj
38	5	1	party_nn	38	6	1	sector_nn

different_jj			provincial_jj		
39	5	1 things_nn	39	6	1 governments_nn
		european_jj			
40	5	1 commission_nn	40	6	1 pti_jj government_nn
		european_jj			
41	5	1 elections_nn	41	6	1 supreme_jj court_nn
					general_jj
42	5	1 good_jj news_nn	42	5	1 elections_nn
43	5	1 good_jj reasons_nn	43	5	1 general_jj public_nn
					high_jj
44	5	1 greek_jj crisis_nn	44	5	1 court_nn
		greek_jj			high_jj
45	5	1 government_nn	45	5	1 levels_nn
46	5	1 labour_jj leader_nn	46	5	1 indian_jj muslims_nn
		labour_jj			kashmiri_jj
47	5	1 leadership_nn	47	5	1 people_nn
		national_jj			large_jj
48	5	1 income_nn	48	5	1 number_nn
49	5	1 other_jj countries_nn	49	5	1 magnetic_jj field_nn
50	5	1 other_jj parties_nn	50	5	1 new_jj elections_nn
51	5	1 private_jj sector_nn	51	5	1 other_jj countries_nn
					profit_jj
52	5	1 recent_jj years_nnt	52	5	1 organizations_nn
53	5	1 social_jj media_nn	53	4	1 afghan_jj people_nn

							civil_jj
54	5	1	united_jj nations_nn	54	4	1	war_nn
							collective_jj
55	5	1	young_jj men_nn	55	4	1	punishment_nn
							economic_jj
56	4	1	big_jj tech_nn	56	4	1	development_nn
							economic_jj
57	4	1	blind_jj eye_nn	57	4	1	progress_nn
			conservative_jj				economic_jj
58	4	1	government_nn	58	4	1	prosperity_nn
			democratic_jj				educational_jj
59	4	1	party_nn	59	4	1	institutions_nn
			developing_jj				federal_jj
60	4	1	world_nn	60	4	1	government_nn
			earlier_jjr				high_jj
61	4	1	version_nn	61	4	1	level_nn
			economic_jj				hindu_jj
62	4	1	growth_nn	62	4	1	right_nn
			electoral_jj				indian_jj
63	4	1	system_nn	63	4	1	army_nn
			european_jj				indonesian_jj
64	4	1	convention_nn	64	4	1	tourism_nn
							international_jj
65	4	1	european_jj court_nn	65	4	1	media_nn

european_jj				long_jj			
66	4	1	leaders_nn	66	4	1	term_nn
67	4	1	financial_jj crisis_nn	67	4	1	magnetic_jj fields_nn
				middle_jj			
68	4	1	greek_jj exit_nn	68	4	1	class_nn
69	4	1	greek_jj voters_nn	69	4	1	muslim_jj league_nn
70	4	1	green_jj paper_nn	70	4	1	national_jj culture_nn
				national_jj			
71	4	1	human_jj rights_nn	71	4	1	investment_nn
				national_jj			
72	4	1	human_jj beings_nn	72	4	1	security_nn
				only_jj			
73	4	1	implicit_jj bias_nn	73	4	1	way_nn
				independent_jj			
74	4	1	voices_nn	74	4	1	political_jj leaders_nn
				liberal_jj			
75	4	1	democrats_nn	75	4	1	party_nn
				real_jj			
76	4	1	long_jj term_nn	76	4	1	power_nn
				recent_jj			
77	4	1	low_jj pay_nn	77	4	1	statement_nn
				recent_jj			
78	4	1	muslim_jj world_nn	78	4	1	visit_nn
				new_jj			
79	4	1	legislation_nn	79	4	1	dog_nn
				russian_jj			

80	4	1	other_jj parts_nn political_jj	80	4	1	statutory_jj rape_nn
81	4	1	advantage_nn	81	4	1	united_jj nations_nn armed_jj
82	4	1	political_jj bias_nn religious_jj	82	3	1	forces_nn babri_jj
83	4	1	education_nn scientific_jj	83	3	1	mosque_nn
84	4	1	community_nn social_jj	84	3	1	chief_jj executive_nn civilian_jj
85	4	1	platforms_nn unstoppable_jj	85	3	1	supremacy_nn common_jj
86	4	1	rise_nn	86	3	1	man_nn countrywide_jj
87	3	1	accurate_jj news_nn american_jj	87	3	1	protests_nn democratic_jj
88	3	1	president_nn authoritative_jj	88	3	1	process_nn developing_jj
89	3	1	reporting_nn	89	3	1	countries_nn economic_jj
90	3	1	big_jj part_nn	90	3	1	activities_nn economic_jj
91	3	1	big_jj questions_nn	91	3	1	corridor_nn elected_jj
92	3	1	bitter_jj pill_nn	92	3	1	government_nn

							electromagnetic_jj
93	3	1	british_jj people_nn	93	3	1	field_nn
							encroachment_jj
94	3	1	british_jj politics_nn	94	3	1	drive_nn
			calm_jj				ethnic_jj
95	3	1	explanation_nn	95	3	1	cleansing_nn
							external_jj
96	3	1	closer_jjr union_nn	96	3	1	debt_nn
			commercial_jj				federal_jj
97	3	1	ownership_nn	97	3	1	governments_nn
			conservative_jj				financial_jj
98	3	1	manifesto_nn	98	3	1	inclusion_nn
			constitutional_jj				foreign_jj
99	3	1	convention_nn	99	3	1	exchange_nn
			critical_jj				fundamental_jj
100	3	1	decisions_nn	100	3	1	right_nn
			developing_jj				global_jj
101	3	1	regions_nn	101	3	1	level_nn
			different_jj				global_jj
102	3	1	choice_nn	102	3	1	trade_nn
			disabled_jj				golden_jj
103	3	1	people_nn	103	3	1	opportunity_nn
			economic_jj				
104	3	1	reasons_nn	104	3	1	good_jj reputation_nn

			editorial_jj				high_jj
105	3	1	independence_nn	105	3	1	time_nnt
							hot_jj
106	3	1	equal_jj access_nn	106	3	1	seat_nn
			existential_jj				humanitarian_jj
107	3	1	threat_nn	107	3	1	law_nn
			extraordinary_jj				
108	3	1	times_nnt	108	3	1	hybrid_jj system_nn
							incumbent_jj
109	3	1	final_jj outcome_nn	109	3	1	government_nn
							indian_jj
110	3	1	fiscal_jj union_nn	110	3	1	forces_nn
							indian_jj
111	3	1	free_jj time_nnt	111	3	1	media_nn
			general_jj				indian_jj
112	3	1	elections_nn	112	3	1	parliament_nn
							international_jj
113	3	1	global_jj heating_nn	113	3	1	law_nn
							kashmiri_jj
114	3	1	good_jj reason_nn	114	3	1	people_nn
115	3	1	greek_jj peoples_nn	115	3	1	larger_jjr picture_nn
			greek_jj				light_jj
116	3	1	referendum_nn	116	3	1	warning_nn
							live_jj
117	3	1	green_jj levies_nn	117	3	1	births_nn

							mainstream_jj
118	3	1	human_jj nature_nn	118	3	1	media_nn
			illegal_jj				mass_jj
119	3	1	migration_nn	119	3	1	hysteria_nn
			investigatory_jj				massive_jj
120	3	1	powers_nn	120	3	1	devaluation_nn
			labour_jj				
121	3	1	government_nn	121	3	1	mental_jj health_nn
			legislative_jj				merged_jj
122	3	1	council_nn	122	3	1	areas_nn
							military_jj
123	3	1	liberal_jj values_nn	123	3	1	expenditure_nn
							monetary_jj
124	3	1	long_jj way_nn	124	3	1	support_nn
							national_jj
125	3	1	main_jj parties_nn	125	3	1	accountability_nn
							national_jj
126	3	1	modern_jj times_nnt	126	3	1	flag_nn
			monetary_jj				national_jj
127	3	1	union_nn	127	3	1	register_nn
128	3	1	new_jj book_nn	128	3	1	negative_jj effects_nn
							neighbouring_jj
129	3	1	new_jj data_nn	129	3	1	country_nn
							new_jj
130	3	1	new_jj law_nn	130	3	1	development_nn

131	3	1	new_jj report_nn northern_jj	131	3	1	other_jj countries_nn other_jj
132	3	1	powerhouse_nn ordinary_jj	132	3	1	day_nnt other_jj
133	3	1	people_nn	133	3	1	factors_nn other_jj
134	3	1	other_jj hand_nn	134	3	1	forms_nn
135	3	1	other_jj words_nn	135	3	1	other_jj provinces_nn other_jj
136	3	1	past_jj decade_nnt	136	3	1	way_nn other_jj
137	3	1	past_jj year_nnt political_jj	137	3	1	words_nn pakistani_jj
138	3	1	capital_nn political_jj	138	3	1	dramas_nn political_jj
139	3	1	consequences_nn political_jj	139	3	1	parties_nn political_jj
140	3	1	leaders_nn political_jj	140	3	1	element_nn
141	3	1	parties_nn	141	3	1	pti_jj government_nn recent_jj
142	3	1	political_jj power_nn	142	3	1	past_nn recent_jj
143	3	1	political_jj union_nn	143	3	1	years_nnt

							red_jj
144	3	1	pressing_jj issue_nn	144	3	1	beacon_nn
			public_jj				ruling_jj
145	3	1	services_nn	145	3	1	party_nn
							secular_jj
146	3	1	public_jj funds_nn	146	3	1	face_nn
							separatist_jj
147	3	1	public_jj life_nn	147	3	1	movement_nn
			public_jj				sexual_jj
148	3	1	prosecutions_nn	148	3	1	abuse_nn
							short_jj
149	3	1	racial_jj hatred_nn	149	3	1	term_nn
							social_jj
150	3	1	real_jj terms_nn	150	3	1	justice_nn
							strategic_jj
151	3	1	recent_jj months_nnt	151	3	1	culture_nn
			religious_jj				successive_jj
152	3	1	hatred_nn	152	3	1	governments_nn
			responsible_jj				
153	3	1	journalism_nn	153	3	1	western_jj media_nn
							accused_jj
154	3	1	social_jj policy_nn	154	2	1	persons_nn
155	3	1	syrian_jj kurds_nn	155	2	1	affected_jj areas_nn
			traditional_jj				afghan_jj
156	3	1	media_nn	156	2	1	government_nn

			violent_jj				afghan_jj
157	3	1	jihadism_nn	157	2	1	dialogue_nn
			western_jj				american_jj
158	3	1	governments_nn	158	2	1	policy_nn
			wholesale_jj				anaerobic_jj
159	3	1	energy_nn	159	2	1	bacteria_nn
			working_jj				anxious_jj
160	3	1	families_nn	160	2	1	looks_nn
			affordable_jj				asian_jj
161	2	1	housing_nn	161	2	1	region_nn
			african_jj				asian_jj
162	2	1	activists_nn	162	2	1	women_nn
			african_jj				autonomous_jj
163	2	1	americans_nn	163	2	1	administration_nn
			african_jj				basic_jj
164	2	1	government_nn	164	2	1	rights_nn
			american_jj				
165	2	1	society_nn	165	2	1	better_jjr society_nn
			ample_jj				
166	2	1	evidence_nn	166	2	1	better_jjr strategy_nn
			average_jj				biggest_jjt
167	2	1	earnings_nn	167	2	1	beneficiaries_nn
							biggest_jjt
168	2	1	baltic_jj states_nn	168	2	1	problem_nn

			basic_jj				bilateral_jj
169	2	1	principles_nn	169	2	1	visit_nn
							called_jj
170	2	1	basic_jj rate_nn	170	2	1	accountability_nn
							cardiovascular_jj
171	2	1	be_jj migrants_nn	171	2	1	diseases_nn
							central_jj
172	2	1	best_jjt chance_nn	172	2	1	bank_nn
			british_jj				century_jj
173	2	1	government_nn	173	2	1	forebears_nn
							certain_jj
174	2	1	british_jj history_nn	174	2	1	areas_nn
							chinese_jj
175	2	1	british_jj society_nn	175	2	1	state_nn
							civil_jj
176	2	1	chief_jj adviser_nn	176	2	1	servants_nn
			chief_jj				civil_jj
177	2	1	executives_nn	177	2	1	society_nn
							civilian_jj
178	2	1	chief_jj inspector_nn	178	2	1	government_nn
			chinese_jj				
179	2	1	economy_nn	179	2	1	civilian_jj leader_nn
			christian_jj				
180	2	1	character_nn	180	2	1	civilian_jj objects_nn
181	2	1	civil_jj war_nn	181	2	1	classic_jj example_nn

							colonial_jj
182	2	1	civil_jj society_nn	182	2	1	societies_nn
							committed_jj
183	2	1	clear_jj heads_nn	183	2	1	suicide_nn
			collective_jj				compliant_jj
184	2	1	bargaining_nn	184	2	1	politician_nn
							concerned_jj
185	2	1	colored_jj people_nn	185	2	1	authorities_nn
			common_jj				
186	2	1	market_nn	186	2	1	concrete_jj steps_nn
			compelling_jj				constitutional_jj
187	2	1	narrative_nn	187	2	1	amendment_nn
			competitive_jj				cost_jj
188	2	1	edge_nn	188	2	1	solution_nn
			conservative_jj				credible_jj
189	2	1	manifesto_nn	189	2	1	security_nn
			conservative_jj				cultural_jj
190	2	1	chancellor_nn	190	2	1	renaissance_nn
			conservative_jj				current_jj
191	2	1	majority_nn	191	2	1	account_nn
			conservative_jj				daily_jj
192	2	1	thinking_nn	192	2	1	basis_nn
			constitutional_jj				daily_jj
193	2	1	changes_nn	193	2	1	lives_nn

			contending_jj				deadly_jj
194	2	1	religions_nn	194	2	1	virus_nn
			criminal_jj				developing_jj
195	2	1	offence_nn	195	2	1	world_nn
196	2	1	current_jj system_nn	196	2	1	different_jj ways_nn
							different_jj
197	2	1	daily_jj mail_nn	197	2	1	countries_nn
198	2	1	daily_jj telegraph_nn	198	2	1	different_jj faiths_nn
199	2	1	day_jj life_nn	199	2	1	different_jj person_nn
							digital_jj
200	2	1	day_jj spending_nn	200	2	1	space_nn
							domestic_jj
201	2	1	decisive_jj action_nn	201	2	1	consumption_nn
			decisive_jj				domestic_jj
202	2	1	charge_nn	202	2	1	politics_nn
							domestic_jj
203	2	1	deep_jj cuts_nn	203	2	1	tourists_nn
							domestic_jj
204	2	1	deep_jj sea_nn	204	2	1	violence_nn
			democracy_jj				
205	2	1	group_nn	205	2	1	eastern_jj country_nn
			democratic_jj				economic_jj ,_
206	2	1	forces_nn	206	2	1	trade_nn
							economic_jj
207	2	1	desperate_jj need_nn	207	2	1	affairs_nn

			developing_jj				economic_jj
208	2	1	countries_nn	208	2	1	burden_nn
							economic_jj
209	2	1	different_jj parts_nn	209	2	1	challenges_nn
210	2	1	different_jj ways_nn	210	2	1	economic_jj field_nn
			difficult_jj				economic_jj
211	2	1	question_nn	211	2	1	indicators_nn
							economic_jj
212	2	1	digital_jj wars_nn	212	2	1	meltdown_nn
			economic_jj				economic_jj
213	2	1	collapse_nn	213	2	1	opportunities_nn
			economic_jj				
214	2	1	rationale_nn	214	2	1	economic_jj terms_nn
			educational_jj				economic_jj
215	2	1	system_nn	215	2	1	wheel_nn
							effective_jj
216	2	1	effective_jj way_nn	216	2	1	response_nn
217	2	1	english_jj laws_nn	217	2	1	elected_jj leader_nn
							elected_jj
218	2	1	english_jj votes_nn	218	2	1	representatives_nn
			equal_jj				electoral_jj
219	2	1	protection_nn	219	2	1	exercise_nn
			european_jj				electoral_jj
220	2	1	elections_nn	220	2	1	process_nn

			european_jj				electronic_jj
221	2	1	cause_nn	221	2	1	crimes_nn
			european_jj				empowering_jj
222	2	1	community_nn	222	2	1	women_nn
			european_jj				
223	2	1	countries_nn	223	2	1	english_jj papers_nn
			european_jj				environmental_jj
224	2	1	currency_nn	224	2	1	issues_nn
			european_jj				
225	2	1	interests_nn	225	2	1	erstwhile_jj fata_nn
			european_jj				
226	2	1	leadership_nn	226	2	1	essential_jj items_nn
			european_jj				
227	2	1	partners_nn	227	2	1	european_jj union_nn
			european_jj				
228	2	1	states_nn	228	2	1	excellent_jj men_nn
			european_jj				existential_jj
229	2	1	unions_nn	229	2	1	crisis_nn
							expired_jj
230	2	1	everyday_jj life_nn	230	2	1	medicines_nn
			exasperated_jj				
231	2	1	creditors_nn	231	2	1	external_jj affairs_nn
232	2	1	extra_jj money_nn	232	2	1	favourite_jj food_nn
							federal_jj
233	2	1	fair_jj bet_nn	233	2	1	government_nn

			federal_jj				financial_jj
234	2	1	government_nn	234	2	1	institutions_nn
			female_jj				financial_jj
235	2	1	homicide_nn	235	2	1	services_nn
			feminist_jj				fiscal_jj
236	2	1	activism_nn	236	2	1	deficits_nn
							forced_jj
237	2	1	financial_jj news_nn	237	2	1	displacement_nn
238	2	1	financial_jj year_nnt	238	2	1	foreign_jj affairs_nn
			fiscal_jj				foreign_jj
239	2	1	approach_nn	239	2	1	investors_nn
240	2	1	fiscal_jj rules_nn	240	2	1	foreign_jj tourists_nn
			fleeing_jj				
241	2	1	persecution_nn	241	2	1	foreign_jj troops_nn
							foreseeable_jj
242	2	1	focal_jj point_nn	242	2	1	future_nn
			footballing_jj				friendly_jj
243	2	1	nations_nn	243	2	1	relations_nn
244	2	1	forced_jj labour_nn	244	2	1	full_jj advantage_nn
							full_jj
245	2	1	foreign_jj affairs_nn	245	2	1	circle_nn
			foreign_jj				fundamental_jj
246	2	1	minister_nn	246	2	1	rights_nn
							gainful_jj
247	2	1	free_jj childcare_nn	247	2	1	employment_nn

			free_jj				general_jj
248	2	1	entitlement_nn	248	2	1	assembly_nn
							global_jj
249	2	1	free_jj tv_nn	249	2	1	media_nn
			future_jj				global_jj
250	2	1	governments_nn	250	2	1	issues_nn
							global_jj
251	2	1	future_jj king_nn	251	2	1	scene_nn
			german_jj				global_jj
252	2	1	chancellor_nn	252	2	1	war_nn
							good_jj
253	2	1	german_jj people_nn	253	2	1	faith_nn
			global_jj				good_jj
254	2	1	warming_nn	254	2	1	governance_nn
			good_jj				good_jj
255	2	1	behaviour_nn	255	2	1	quality_nn
			governing_jj				growing_jj
256	2	1	body_nn	256	2	1	number_nn
							guiding_jj
257	2	1	grand_jj project_nn	257	2	1	principles_nn
			grand_jj				harsh_jj
258	2	1	statements_nn	258	2	1	reality_nn
			greek_jj				hegemonic_jj
259	2	1	economy_nn	259	2	1	ambitions_nn

			greek_jj				high_jj
260	2	1	counterparts_nn	260	2	1	profile_nn
							higher_jjr
261	2	1	greek_jj debt_nn	261	2	1	education_nn
							highest_jjt
262	2	1	greek_jj debts_nn	262	2	1	number_nn
							hindu_jj
263	2	1	greek_jj default_nn	263	2	1	bigotry_nn
			green_jj				huge_jj
264	2	1	movement_nn	264	2	1	setback_nn
							human_jj
265	2	1	green_jj party_nn	265	2	1	dilemma_nn
266	2	1	hard_jj line_nn	266	2	1	human_jj tragedy_nn
							humanitarian_jj
267	2	1	headlong_jj rush_nn	267	2	1	crisis_nn
							hybrid_jj
268	2	1	heavy_jj price_nn	268	2	1	war_nn
			higher_jjr				hypersonic_jj
269	2	1	productivity_nn	269	2	1	missiles_nn
							ideological_jj
270	2	1	highest_jjt level_nn	270	2	1	distress_nn
			historic_jj				
271	2	1	proportions_nn	271	2	1	important_jj issue_nn
							income_jj
272	2	1	horrific_jj crimes_nn	272	2	1	countries_nn

273	2	1	huge_jj amounts_nn	273	2	1	increasing_jj day_nnt increasing_jj
274	2	1	human_jj need_nn	274	2	1	exports_nn indian_jj
275	2	1	human_jj needs_nn humanitarian_jj	275	2	1	occupation_nn indian_jj
276	2	1	crisis_nn hung_jj	276	2	1	constitution_nn
277	2	1	parliament_nn important_jj	277	2	1	indian_jj designs_nn indian_jj
278	2	1	question_nn	278	2	1	moves_nn indian_jj
279	2	1	important_jj role_nn inclined_jj	279	2	1	occupation_nn
280	2	1	segments_nn increased_jj	280	2	1	indian_jj policies_nn
281	2	1	capacity_nn independent_jj	281	2	1	indian_jj security_nn indian_jj
282	2	1	schools_nn indeterminate_jj	282	2	1	state_nn indirect_jj
283	2	1	sentences_nn	283	2	1	taxation_nn individual_jj
284	2	1	internal_jj party_nn internal_jj	284	2	1	capacities_nn indonesian_jj
285	2	1	politics_nn	285	2	1	government_nn

			international_jj				industrial_jj
286	2	1	community_nn	286	2	1	complex_nn
			international_jj				informal_jj
287	2	1	effort_nn	287	2	1	summit_nn
							interior_jj
288	2	1	iraqi_jj people_nn	288	2	1	ministry_nn
			islamist_jj				interior_jj
289	2	1	terrorism_nn	289	2	1	sindh_nn
							international_jj
290	2	1	jihadi_jj ideology_nn	290	2	1	security_nn
							international_jj
291	2	1	judicial_jj review_nn	291	2	1	relations_nn
							islamic_jj
292	2	1	labour_jj mps_nn	292	2	1	state_nn
							islamic_jj
293	2	1	labour_jj needs_nn	293	2	1	terrorists_nn
			labour_jj				israeli_jj
294	2	1	supporters_nn	294	2	1	forces_nn
							israeli_jj
295	2	1	labour_jj voters_nn	295	2	1	model_nn
			large_jj				
296	2	1	investments_nn	296	2	1	judicial_jj system_nn
							kashmiri_jj
297	2	1	large_jj number_nn	297	2	1	muslims_nn

							kashmiri_jj
298	2	1	leaky_jj pipeline_nn	298	2	1	women_nn
			leftwing_jj				
299	2	1	leadership_nn	299	2	1	kinetic_jj means_nn
							leading_jj
300	2	1	legal_jj highs_nn	300	2	1	cause_nn
301	2	1	legal_jj provision_nn	301	2	1	led_jj government_nn
							legal_jj
302	2	1	lgbti_jj rights_nn	302	2	1	cover_nn
			local_jj				legal_jj
303	2	1	authorities_nn	303	2	1	experts_nn
							little_jj
304	2	1	local_jj media_nn	304	2	1	world_nn
							live_jj
305	2	1	long_jj list_nn	305	2	1	stream_nn
							live_jj
306	2	1	longer_jjr term_nn	306	2	1	births_nn
307	2	1	major_jj parties_nn	307	2	1	living_jj organism_nn
							long_jj
308	2	1	major_jj speech_nn	308	2	1	run_nn
							long_jj
309	2	1	male_jj mps_nn	309	2	1	time_nnt
			mammoth_jj				long_jj
310	2	1	demonstration_nn	310	2	1	way_nn

			mass_jj				loved_jj
311	2	1	shootings_nn	311	2	1	ones_nn
							lower_jjr
312	2	1	median_jj pay_nn	312	2	1	rates_nn
							main_jj
313	2	1	mental_jj health_nn	313	2	1	reason_nn
			metropolitan_jj				major_jj
314	2	1	police_nn	314	2	1	part_nn
			military_jj				martial_jj
315	2	1	personnel_nn	315	2	1	law_nn
							maternal_jj
316	2	1	modern_jj world_nn	316	2	1	complications_nn
317	2	1	moral_jj crusade_nn	317	2	1	maternal_jj death_nn
			moral_jj				meanest_jjt
318	2	1	obligation_nn	318	2	1	doberman_nn
			multinational_jj				mega_jj
319	2	1	corporations_nn	319	2	1	corruption_nn
			multinational_jj				merged_jj
320	2	1	sponsors_nn	320	2	1	districts_nn
			multiparty_jj				middle_jj
321	2	1	politics_nn	321	2	1	layer_nn
322	2	1	mutilated_jj body_nn	322	2	1	military_jj matters_nn
							military_jj
323	2	1	national_jj audit_nn	323	2	1	ruler_nn

							modern_jj
324	2	1	national_jj debt_nn	324	2	1	apartheid_nn
			national_jj				month_jj
325	2	1	football_nn	325	2	1	record_nn
			national_jj				multiple_jj
326	2	1	governments_nn	326	2	1	challenges_nn
			national_jj				multiple_jj
327	2	1	interest_nn	327	2	1	patients_nn
			national_jj				muslim_jj
328	2	1	politics_nn	328	2	1	poets_nn
			national_jj				
329	2	1	security_nn	329	2	1	mysterious_jj gas_nn
			necessary_jj				national_jj
330	2	1	precondition_nn	330	2	1	assembly_nn
			neoliberal_jj				national_jj
331	2	1	policies_nn	331	2	1	interests_nn
			new_jj				national_jj
332	2	1	administration_nn	332	2	1	register_nn
			new_jj				natural_jj
333	2	1	committee_nn	333	2	1	agriculture_nn
334	2	1	new_jj justice_nn	334	2	1	natural_jj process_nn
			new_jj				nazi_jj
335	2	1	parliament_nn	335	2	1	agenda_nn
							nazi_jj
336	2	1	new_jj powers_nn	336	2	1	leader_nn

							near_jj
337	2	1	new_jj study_nn	337	2	1	future_nn
							net_jj
338	2	1	new_jj world_nn	338	2	1	result_nn
			noisy_jj				new_jj
339	2	1	confrontation_nn	339	2	1	year_nnt
			northern_jj				new_jj
340	2	1	academics_nn	340	2	1	era_nn
			nuclear_jj				only_jj
341	2	1	facilities_nn	341	2	1	country_nn
			nuclear_jj				only_jj
342	2	1	programme_nn	342	2	1	party_nn
			nuclear_jj				operation_jj
343	2	1	weapon_nn	343	2	1	reports_nn
			nuclear_jj				
344	2	1	weapons_nn	344	2	1	ordinary_jj people_nn
							other_jj
345	2	1	old_jj girl_nn	345	2	1	areas_nn
							other_jj
346	2	1	only_jj way_nn	346	2	1	cities_nn
			original_jj				other_jj
347	2	1	decision_nn	347	2	1	parties_nn
			other_jj				other_jj
348	2	1	payments_nn	348	2	1	person_nn

other_jj				other_jj			
349	2	1	candidates_nn	349	2	1	side_nn
							outstanding_jj
350	2	1	other_jj day_nnt	350	2	1	disputes_nn
							painful_jj
351	2	1	other_jj family_nn	351	2	1	thorn_nn
							pakistani_jj
352	2	1	other_jj half_nn	352	2	1	expatriates_nn
							palestinian_jj
353	2	1	other_jj places_nn	353	2	1	economy_nn
							palestinian_jj
354	2	1	other_jj side_nn	354	2	1	population_nn
355	2	1	other_jj way_nn	355	2	1	particular_jj area_nn
							passing_jj
356	2	1	paid_jj workers_nn	356	2	1	day_nnt
			parliamentary_jj				past_jj
357	2	1	committees_nn	357	2	1	year_nnt
							philosophical_jj
358	2	1	party_jj rule_nn	358	2	1	observation_nn
							physical_jj
359	2	1	penal_jj policy_nn	359	2	1	activity_nn
360	2	1	penal_jj reform_nn	360	2	1	physical_jj health_nn
			pharmaceutical_jj				political_jj
361	2	1	companies_nn	361	2	1	elite_nn

			pharmaceutical_jj				political_jj
362	2	1	industry_nn	362	2	1	stability_nn
			political_jj				political_jj
363	2	1	agenda_nn	363	2	1	victimisation_nn
364	2	1	political_jj class_nn	364	2	1	political_jj career_nn
			political_jj				
365	2	1	climate_nn	365	2	1	political_jj forces_nn
			political_jj				political_jj
366	2	1	culture_nn	366	2	1	instability_nn
			political_jj				
367	2	1	economist_nn	367	2	1	political_jj leader_nn
							political_jj
368	2	1	political_jj future_nn	368	2	1	mileage_nn
							political_jj
369	2	1	political_jj mood_nn	369	2	1	pundits_nn
							political_jj
370	2	1	political_jj ones_nn	370	2	1	solution_nn
			political_jj				political_jj
371	2	1	opinion_nn	371	2	1	stability_nn
							political_jj
372	2	1	political_jj party_nn	372	2	1	vendetta_nn
							political_jj
373	2	1	political_jj risk_nn	373	2	1	wing_nn
			political_jj				
374	2	1	strategy_nn	374	2	1	poor_jj conditions_nn

			political_jj				
375	2	1	system_nn	375	2	1	popular_jj leader_nn popular_jj
376	2	1	poor_jj countries_nn	376	2	1	sovereignty_nn
377	2	1	poor_jj people_nn	377	2	1	popular_jj support_nn positive_jj
378	2	1	poorer_jjr pupils_nn	378	2	1	side_nn powerful_jj
379	2	1	popular_jj vote_nn practical_jj	379	2	1	country_nn preventive_jj
380	2	1	problems_nn presidential_jj	380	2	1	measures_nn
381	2	1	palace_nn presidential_jj	381	2	1	private_jj partners_nn proper_jj
382	2	1	plane_nn prime_jj	382	2	1	knowledge_nn
383	2	1	ministers_nn	383	2	1	protesting_jj mob_nn provincial_jj
384	2	1	private_jj schools_nn proportional_jj	384	2	1	government_nn provincial_jj
385	2	1	representation_nn	385	2	1	assembly_nn psychological_jj
386	2	1	public_jj service_nn	386	2	1	cost_nn public_jj
387	2	1	public_jj anger_nn	387	2	1	assets_nn

							public_jj
388	2	1	public_jj bodies_nn	388	2	1	spaces_nn
			public_jj				raised_jj
389	2	1	confidence_nn	389	2	1	beds_nn
			public_jj				recent_jj
390	2	1	expenditure_nn	390	2	1	days_nnt
			public_jj				recent_jj
391	2	1	finances_nn	391	2	1	events_nn
			public_jj				recent_jj
392	2	1	institution_nn	392	2	1	price_nn
393	2	1	public_jj interest_nn	393	2	1	regional_jj peace_nn
394	2	1	public_jj policy_nn	394	2	1	right_jj direction_nn
							rough_jj
395	2	1	public_jj realm_nn	395	2	1	music_nn
							rudderless_jj
396	2	1	public_jj service_nn	396	2	1	journey_nn
							rural_jj
397	2	1	quality_jj jobs_nn	397	2	1	areas_nn
398	2	1	racial_jj hate_nn	398	2	1	russian_jj military_nn
							sacred_jj
399	2	1	real_jj issue_nn	399	2	1	thread_nn
							saudi_jj
400	2	1	real_jj issues_nn	400	2	1	crown_nn
							scheduled_jj
401	2	1	real_jj politics_nn	401	2	1	tribes_nn

							senior_jj
402	2	1	real_jj sense_nn	402	2	1	police_nn
403	2	1	real_jj world_nn	403	2	1	serious_jj threat_nn
			reasonable_jj				shorter_jjr
404	2	1	chastisement_nn	404	2	1	man_nn
							shrewd_jj
405	2	1	recent_jj history_nn	405	2	1	politician_nn
			refreshing_jj				sicilian_jj
406	2	1	change_nn	406	2	1	mafia_nn
							slanted_jj
407	2	1	regional_jj cities_nn	407	2	1	coverage_nn
							social_jj
408	2	1	regressive_jj tax_nn	408	2	1	media_nn
							social_jj
409	2	1	right_jj favours_nn	409	2	1	institutions_nn
							social_jj
410	2	1	right_jj thing_nn	410	2	1	values_nn
411	2	1	rightwing_jj plot_nn	411	2	1	social_jj welfare_nn
							solemn_jj
412	2	1	sajid_jj javid_nn	412	2	1	faces_nn
413	2	1	scottish_jj mps_nn	413	2	1	speed_jj internet_nn
							spiritual_jj
414	2	1	scottish_jj voters_nn	414	2	1	pantheon_nn
			secondary_jj				strategic_jj
415	2	1	school_nn	415	2	1	ambitions_nn

			select_jj				strategic_jj
416	2	1	committee_nn	416	2	1	goal_nn
			serious_jj				strategic_jj
417	2	1	assaults_nn	417	2	1	implications_nn
			serious_jj				strategic_jj
418	2	1	concern_nn	418	2	1	partnership_nn
							strict_jj
419	2	1	serious_jj threat_nn	419	2	1	laws_nn
							strong_jj
420	2	1	sex_jj marriage_nn	420	2	1	economy_nn
421	2	1	short_jj supply_nn	421	2	1	supreme_jj courts_nn
							sustainable_jj
422	2	1	single_jj market_nn	422	2	1	development_nn
423	2	1	small_jj beer_nn	423	2	1	syrian_jj refugees_nn
							taliban_jj
424	2	1	social_jj housing_nn	424	2	1	peace_nn
			social_jj				terrorist_jj
425	2	1	problem_nn	425	2	1	organization_nn
							terrorist_jj
426	2	1	social_jj attitudes_nn	426	2	1	acts_nn
							terrorist_jj
427	2	1	social_jj care_nn	427	2	1	organization_nn
			social_jj				terrorist_jj
428	2	1	democrats_nn	428	2	1	organizations_nn
429	2	1	social_jj mobility_nn	429	2	1	testing_jj times_nnt

			social_jj				torn_jj
430	2	1	response_nn	430	2	1	country_nn
							tough_jj
431	2	1	sorry_jj history_nn	431	2	1	times_nnt
			strategic_jj				
432	2	1	choices_nn	432	2	1	traumatic_jj stress_nn
			successive_jj				tribal_jj
433	2	1	governments_nn	433	2	1	elders_nn
			sustainable_jj				underage_jj
434	2	1	development_nn	434	2	1	marriage_nn
			temporary_jj				uphill_jj
435	2	1	imposition_nn	435	2	1	battle_nn
			temporary_jj				young_jj
436	2	1	price_nn	436	2	1	leader_nn
							young_jj
437	2	1	term_jj plan_nn	437	2	1	people_nn
			terrorist_jj				
438	2	1	attacks_nn	438	2	1	young_jj women_nn
439	2	1	tiny_jj fraction_nn	438	2	1	young_jj women_nn
440	2	1	top_jj priority_nn				
441	2	1	tory_jj health_nn				
442	2	1	tory_jj manifesto_nn				
443	2	1	tory_jj ministers_nn				
444	2	1	turkish_jj constitution_nn				
445	2	1	turkish_jj president_nn				

446	2	1	turkish_jj society_nn
447	2	1	unconscious_jj bias_nn
448	2	1	unequal_jj economy_nn
449	2	1	unintended_jj consequences_nn
450	2	1	vast_jj majority_nn
451	2	1	vicious_jj conflict_nn
452	2	1	vital_jj role_nn
453	2	1	vulnerable_jj groups_nn
454	2	1	vulnerable_jj people_nn
455	2	1	wage_jj economy_nn
456	2	1	western_jj powers_nn
457	2	1	white_jj men_nn
458	2	1	wide_jj parties_nn
459	2	1	work_jj benefits_nn
460	2	1	year_jj plan_nn
461	2	1	young_jj workers_nn
462	2	1	younger_jjr people_nn

APPENDIX-F

ADV+V

V

#Total No. of Cluster Types:

1107

ENG

#Total No. of Cluster Types:

839

PAK

#Total No. of Cluster Tokens:

1142

#Total No. of Cluster

Tokens: 896

			already_rr				online_rr
1	5	1	made_vvn	1	11	1	plot_vv
			never_rr				more_rrr : _:
2	4	1	influenced_vvn	2	8	1	nab_vv
3	3	1	now_rt clear_vv	3	7	1	more_rrr : _: blast_vv
							no_rr
4	3	1	now_rt support_vv	4	5	1	denying_vvg
			ago_ra , _:				also_rr
5	2	1	needs_vvz	5	3	1	chose_vvd
							also_rr
6	2	1	also_rr points_vvz	6	3	1	like_vvi
			also_rr				now_rt
7	2	1	suggests_vvz	7	3	1	seems_vvz
			always_rr				
8	2	1	going_vvgk	8	2	1	actually_rr made_vvn
			always_rr				already_rr
9	2	1	regretted_vvd	9	2	1	announced_vvn

			closely_rr				already_rr
10	2	1	watched_vvn	10	2	1	working_vvg
			democratically_rr				also_rr
11	2	1	elected_vvn	11	2	1	acknowledge_vvi
			desperately_rr				also_rr
12	2	1	needs_vvz	12	2	1	help_vvi
							also_rr
13	2	1	far_rr suggests_vvz	13	2	1	joined_vvn
			jointly_rr				also_rr
14	2	1	define_vvi	14	2	1	made_vvn
			now_rt				also_rr
15	2	1	represented_vvn	15	2	1	makes_vvz
							also_rr
16	2	1	only_rr came_vvd	16	2	1	need_vv
17	2	1	only_rr take_vv	17	2	1	also_rr provides_vvz
			over_rp , ,				
18	2	1	takes_vvz	18	2	1	always_rr called_vvn
			perhaps_rr				falsely_rr
19	2	1	emerging_vvg	19	2	1	accused_vvn
			possibly_rr				further_rrr
20	2	1	disagree_vvi	20	2	1	complicated_vvn
			publicly_rr				how_rrq
21	2	1	owned_vvn	21	2	1	come_vvn
			repeatedly_rr				little_rr
22	2	1	targeted_vvn	22	2	1	doubt_vv

							miserably_rr
23	2	1	rightly_rr gave_vvd	23	2	1	failed_vvn
							necessarily_rr
24	2	1	still_rr taking_vvg	24	2	1	match_vvi
			subsequently_rr				
25	2	1	said_vvn	25	2	1	newly_rr formed_vvn
			today_rt : _:				only_rr
26	2	1	sign_vv	26	2	1	fear_vvi
			urgently_rr				only_rr
27	2	1	required_vvn	27	2	1	hope_vvi
			widely_rr				only_rr
28	2	1	attributed_vvn	28	2	1	look_vvi
							reluctantly_rr
				29	2	1	took_vvd
				30	2	1	so_rr let_vv
				31	2	1	up_rp ,_, meant_vvd

APPENDIX-G

NN+NN

#Total No. of Cluster Types:

2475

ENG

#Total No. of Cluster Types:

2366

PAK

#Total No. of Cluster Tokens:

2890

#Total No. of Cluster

Tokens: 2795

1	24	1	licence_nn1 fee_nn	1	16	1	growth_nn1 rate_nn account_nn1
2	13	1	world_nn1 cup_nn	2	12	1	deficit_nn gdp_nn1
3	10	1	lib_nn1 dems_nn	3	12	1	growth_nn finder_nn1
4	9	1	tax_nn1 credits_nn justice_nn1	4	11	1	tool_nn plot_nn1
5	7	1	secretary_nn justice_nn1	5	11	1	finder_nn health_nn1
6	7	1	system_nn public_nn1	6	8	1	care_nn army_nn1
7	7	1	opinion_nn culture_nn1	7	7	1	chief_nn investment_nn1
8	6	1	secretary_nn education_nn1	8	7	1	companies_nn money_nn1
9	6	1	secretary_nn	9	7	1	laundering_nn water_nn1
10	6	1	income_nn1 tax_nn	10	7	1	crisis_nn

			million_nno				world_nn1
11	6	1	people_nn	11	7	1	war_nn
			prison_nn1				
12	6	1	population_nn	12	6	1	gdp_nn1 growth_nn
			resolution_nn1				
13	6	1	foundation_nn	13	6	1	hate_nn1 speech_nn
			election_nn1				
14	5	1	campaign_nn	14	6	1	million_nno people_nn
			election_nn1				war_nn1
15	5	1	result_nn	15	6	1	crimes_nn
			finance_nn1				amendment_nn1
16	5	1	minister_nn	16	5	1	bill_nn
			health_nn1				apex_nn1
17	5	1	service_nn	17	5	1	court_nn
			magna_nn121				
18	5	1	carta_nn	18	5	1	child_nn1 marriage_nn
			mansion_nn1				law_nn1
19	5	1	house_nn	19	5	1	enforcement_nn
			stock_nn1				opposition_nn1
20	5	1	market_nn	20	5	1	parties_nn
			trade_nn1				
21	5	1	unions_nn	21	5	1	steam_nn1 engine_nn
			world_nn1				account_nn1
22	5	1	service_nn	22	4	1	deficits_nn

			billionaire_nn1				cabinet_nn1
23	4	1	owners_nn	23	4	1	committee_nn
			child_nn1				child_nn1
24	4	1	poverty_nn	24	4	1	marriage_nn
			climate_nn1				climate_nn1
25	4	1	change_nn	25	4	1	change_nn
							climate_nn1
26	4	1	death_nn1 toll_nn	26	4	1	change_nn
			default_nn1				education_nn1
27	4	1	tariffs_nn	27	4	1	sector_nn
			development_nn1				enforcement_nn1
28	4	1	goals_nn	28	4	1	agencies_nn
			guardians_nn2				greenhouse_nn1
29	4	1	independence_nn	29	4	1	gases_nn
			intelligence_nn1				indemnity_nn1
30	4	1	agencies_nn	30	4	1	bond_nn
			leadership_nn1				news_nn1
31	4	1	election_nn	31	4	1	desk_nn
			quality_nn1				peace_nn1
32	4	1	journalism_nn	32	4	1	deal_nn
			reader_nn1				
33	4	1	contribution_nn	33	4	1	peoples_nn2 party_nn
			welfare_nn1				
34	4	1	state_nn	34	4	1	per_nnu21 cent_nnu

			womens_nn2				price_nn1
35	4	1	football_nn	35	4	1	hike_nn
							rights_nn2
36	3	1	aid_nn1 workers_nn	36	4	1	violations_nn
			bailout_nn1				saving_nn1
37	3	1	terms_nn	37	4	1	rate_nn
							security_nn1
38	3	1	bank_nn1 levy_nn	38	4	1	apparatus_nn
			banking_nn1				security_nn1
39	3	1	sector_nn	39	4	1	council_nn
							service_nn1
40	3	1	benefit_nn1 cap_nn	40	4	1	delivery_nn
			business_nn1				unit_nn1
41	3	1	secretary_nn	41	4	1	offices_nn
			climate_nn1				world_nn1
42	3	1	emergency_nn	42	4	1	economy_nn
							accountability_nn1
43	3	1	coast_nn1 line_nn	43	3	1	bureau_nn
44	3	1	council_nn1 tax_nn	44	3	1	agro_nn1 chemical_nn
							arms_nn2
45	3	1	debt_nn1 relief_nn	45	3	1	race_nn
			defence_nn1				belly_nn1
46	3	1	secretary_nn	46	3	1	piece_nn
			deputy_nn1				
47	3	1	leadership_nn	47	3	1	billion_nno people_nn

			election_nn1				business_nn1
48	3	1	victory_nn	48	3	1	captains_nn
			energy_nn1				capital_nn1
49	3	1	market_nn	49	3	1	city_nn
							carbon_nn1
50	3	1	exit_nn1 polls_nn	50	3	1	dioxide_nn
			gender_nn1				
51	3	1	diversity_nn	51	3	1	care_nn1 system_nn
			government_nn1				
52	3	1	policy_nn	52	3	1	case_nn1 response_nn
			guardian_nn1				citizen_nn1
53	3	1	journalism_nn	53	3	1	amendment_nn
			house_nn1				
54	3	1	prices_nn	54	3	1	congress_nn1 party_nn
			house_nn1				
55	3	1	prices_nn	55	3	1	corona_nn1 virus_nn
			labour_nn1				development_nn1
56	3	1	market_nn	56	3	1	goals_nn
			leadership_nn1				drug_nn1
57	3	1	contest_nn	57	3	1	abuse_nn
			lives_nn2 ,_ ,				election_nn1
58	3	1	health_nn	58	3	1	commission_nn
59	3	1	m_nnu votes_nn	59	3	1	faith_nn1 ,_ , unity_nn
			ministers_nn2				finance_nn1
60	3	1	questions_nn	60	3	1	minister_nn

			network_nn1			
61	3	1	rail_nn	61	3	1 hate_nn1 speech_nn
			network_nn1			
62	3	1	rails_nn	62	3	1 health_nn1 sector_nn
			opinion_nn1			
63	3	1	polls_nn	63	3	1 systems_nn
64	3	1	party_nn1 leader_nn	64	3	1 home_nn1 minister_nn
65	3	1	per_nnu21 cent_nnu	65	3	1 million_nno
			press_nn1			
66	3	1	conference_nn	66	3	1 inflation_nn1 rate_nn
67	3	1	price_nn1 cap_nn	67	3	1 style_nn
			price_nn1			
68	3	1	regulation_nn	68	3	1 restraint_nn
			quality_nn1			
69	3	1	journalism_nn	69	3	1 mortality_nn1 ratio_nn
			referendum_nn1			
70	3	1	bill_nn	70	3	1 verdict_nn
			republican_nn1			
71	3	1	party_nn	71	3	1 leadership_nn
72	3	1	root_nn1 causes_nn	72	3	1 negotiations_nn
			status_nn121			
73	3	1	quo_nn	73	3	1 police_nn2 officer_nn

74	3	1	tax_nn1 credits_nn	74	3	1	policy_nn1 makers_nn
75	3	1	tax_nn1 rates_nn union_nn1	75	3	1	poverty_nn1 line_nn security_nn1
76	3	1	address_nn union_nn1	76	3	1	deterrent_nn security_nn1
77	3	1	members_nn union_nn1	77	3	1	establishment_nn security_nn1
78	3	1	movement_nn	78	3	1	forces_nn spike_nn1
79	3	1	welfare_nn1 bill_nn abduction_nn1 ,_	79	3	1	year_nnt stock_nn1
80	2	1	torture_nn accounts_nn2	80	3	1	exchange_nn
81	2	1	committee_nn	81	3	1	stock_nn1 market_nn term_nn1
82	2	1	air_nn1 strikes_nn archives_nn2	82	3	1	orientation_nn
83	2	1	centre_nn arms_nn2	83	3	1	tourism_nn1 sector_nn tv_nn1
84	2	1	embargo_nn asylum_nn1	84	3	1	channels_nn
85	2	1	seekers_nn	85	3	1	value_nn1 addition_nn
86	2	1	audit_nn1 office_nn austerity_nn1	86	3	1	war_nn1 strategy_nn world_nn1
87	2	1	policies_nn	87	3	1	bank_nn

88	2	1	back_nn1 foot_nn banking_nn1	88	3	1	world_nn1 health_nn agriculture_nn1
89	2	1	crisis_nn	89	2	1	practices_nn agriculture_nn1
90	2	1	bias_nn1 tests_nn britains_nn2	90	2	1	process_nn amendment_nn1)_)
91	2	1	recovery_nn brothers_nn2 ,_	91	2	1	act_nn amendment_nn1)_)
92	2	1	mothers_nn budget_nn1 . _.	92	2	1	bill_nn
93	2	1	mr_nnb budget_nn1	93	2	1	army_nn1 chiefs_nn bargaining_nn1
94	2	1	deficit_nn budget_nn1	94	2	1	chip_nn base_nn1
95	2	1	speech_nn	95	2	1	units_nn
96	2	1	bulk_nn1 data_nn carbon_nn1	96	2	1	begging_nn1 bowl_nn
97	2	1	emissions_nn	97	2	1	blame_nn1 game_nn brain_nn1
98	2	1	care_nn1 quality_nn cash_nn1	98	2	1	cells_nn building_nn1
99	2	1	injection_nn	99	2	1	blocks_nn cabinet_nn1
100	2	1	cash_nn1 terms_nn	100	2	1	members_nn

							capacity_nn1
101	2	1	child_nn1 sex_nn	101	2	1	building_nn
							chief_nn1
102	2	1	child_nn1 tax_nn	102	2	1	gen_nnb
			climate_nn1				child_nn1
103	2	1	anxiety_nn	103	2	1	birth_nn
			commission_nn1				child_nn1
104	2	1	efforts_nn	104	2	1	marriages_nn
			confederate_nn1				citizenship_nn1
105	2	1	flag_nn	105	2	1	amendment_nn
			conference_nn1				claim_nn1
106	2	1	speech_nn	106	2	1	'fake_nn
			consumer_nn1				congress_nn1
107	2	1	bills_nn	107	2	1	government_nn
			consumer_nn1				
108	2	1	detriment_nn	108	2	1	corner_nn1 bursts_nn
			control_nn1				council_nn1
109	2	1	legislation_nn	109	2	1	resolutions_nn
			coronavirus_nn1				
110	2	1	outbreak_nn	110	2	1	cream_nn1 soldiers_nn
			council_nn1				death_nn1
111	2	1	seats_nn	111	2	1	anniversary_nn
			country_nn1				death_nn1
112	2	1	back_nn	112	2	1	toll_nn

creditors_nn2						
113	2	1	terms_nn	113	2	1 dengue_nn1 virus_nn development_nn1
114	2	1	culture_nn1 war_nn	114	2	1 projects_nn dimension_nn1
115	2	1	cup_nn1 final_nn	115	2	1 survey_nn dog_nn1
116	2	1	debt_nn1 relief_nn deficit_nn1	116	2	1 fight_nn
117	2	1	reduction_nn deputy_nn1	117	2	1 dollar_nn1 rate_nn
118	2	1	leader_nn deputy_nn1	118	2	1 dollar_nn1 terms_nn education_nn1 ,_
119	2	1	leader_nn development_nn1	119	2	1 health_nn emergency_nn1
120	2	1	party_nn dispatch_nn1	120	2	1 response_nn employment_nn1
121	2	1	box_nn drug_nn1	121	2	1 opportunities_nn fake_nn1
122	2	1	companies_nn economy_nn1 ._.	122	2	1 news_nn flag_nn1
123	2	1	mr_nnb election_nn1	123	2	1 carrier_nn fossil_nn1
124	2	1	defeat_nn	124	2	1 fuels_nn
125	2	1	elite_nn1 jobs_nn	125	2	1 freedom_nn1 party_nn

			energy_nn1				gas_nn1
126	2	1	companies_nn	126	2	1	leak_nn
			eurozone_nn1				gender_nn1
127	2	1	finance_nn	127	2	1	equality_nn
			executive_nn1				gender_nn1
128	2	1	presidency_nn	128	2	1	inequality_nn
							governments_nn2
129	2	1	family_nn1 life_nn	129	2	1	failure_nn
			family_nn1				greenhouse_nn1
130	2	1	members_nn	130	2	1	effect_nn
			fathers_nn2 , ,				
131	2	1	brothers_nn	131	2	1	growth_nn1 rates_nn
			finance_nn1				headquarters_nn
132	2	1	ministers_nn	132	2	1	hospital_nn
			football_nn1				
133	2	1	association_nn	133	2	1	health_nn1 facility_nn
			football_nn1				health_nn1
134	2	1	fans_nn	134	2	1	facilities_nn
135	2	1	gdp_nn1 limit_nn	135	2	1	health_nn1 issues_nn
			government_nn1				
136	2	1	spending_nn	136	2	1	health_nn1 rights_nn
			grand_nn121				
137	2	1	prix_nn	137	2	1	health_nn1 service_nn
			gun_nn1				
138	2	1	violence_nn	138	2	1	health_nn1 system_nn

139	2	1	hate_nn1 crime_nn health_nn1	139	2	1	health_nn1 units_nn healthcare_nn1
140	2	1	secretary_nn herd_nn1	140	2	1	sector_nn
141	2	1	immunity_nn history_nn1	141	2	1	helping_nn1 hand_nn holi_nn1
142	2	1	museum_nn home_nn1	142	2	1	hai_nn
143	2	1	office_nn home_nn1	143	2	1	house_nn1 arrest_nn impact_nn1
144	2	1	secretary_nn homicide_nn1	144	2	1	assessment_nn implementation_nn1
145	2	1	victims_nn house_nn1	145	2	1	bench_nn intelligence_nn1
146	2	1	speech_nn impeachment_nn1	146	2	1	agencies_nn interest_nn1
147	2	1	trial_nn independence_nn1	147	2	1	groups_nn investment_nn1
148	2	1	party_nn infrastructure_nn1	148	2	1	rate_nn job_nn1
149	2	1	spending_nn interest_nn1	149	2	1	market_nn job_nn1
150	2	1	rates_nn leadership_nn1	150	2	1	opportunities_nn
151	2	1	candidates_nn	151	2	1	labour_nn1 force_nn

			leadership_nn1				
152	2	1	contenders_nn	152	2	1	lagoon_nn1 side_nn
			leadership_nn1				law_nn1
153	2	1	race_nn	153	2	1	enforcement_nn
							law_nn1
154	2	1	lib_nn1 dems_nn	154	2	1	administrator_nn
							lawyers_nn2
155	2	1	lieu_nn1 scheme_nn	155	2	1	movement_nn
			life_nn1				looks_nn2 ,_,
156	2	1	chances_nn	156	2	1	people_nn
							manufacturing_nn1
157	2	1	lions_nn2 share_nn	157	2	1	facilities_nn
			living_nn1				media_nn ,_,
158	2	1	standards_nn	158	2	1	academics_nn
			man_nn1 ,_,				media_nn
159	2	1	woman_nn	159	2	1	giant_nn
			member_nn1				
160	2	1	states_nn	160	2	1	media_nn outlets_nn
			morning_nnt1				media_nn
161	2	1	shot_nn	161	2	1	platforms_nn
			mothers_nn2 ,_,				media_nn
162	2	1	sisters_nn	162	2	1	users_nn
			nations_nn2				
163	2	1	report_nn	163	2	1	migrant_nn1 crisis_nn
164	2	1	no_nn1 vote_nn	164	2	1	million_nno homes_nn

			opposition_nn1				million_nno
165	2	1	leader_nn	165	2	1	women_nn
			opposition_nn1				mineral_nn1
166	2	1	parties_nn	166	2	1	resources_nn
			parole_nn1				mortality_nn1
167	2	1	board_nn	167	2	1	figures_nn
			partners_nn2 ,_ ,				musharraf_nn1
168	2	1	fathers_nn	168	2	1	verdict_nn
			party_nn1				
169	2	1	membership_nn	169	2	1	nab_nn1 chairman_nn
			party_nn1				nnu
170	2	1	rivalries_nn	170	2	1	billion_nno in
							nnu
171	2	1	pay_nn1 packet_nn	171	2	1	billion_nno in
			pensions_nn2				nnu
172	2	1	secretary_nn	172	2	1	billion_nno to
			peoples_nn2				obl_nn1
173	2	1	history_nn	173	2	1	raid_nn
			peoples_nn2				occupation_nn1
174	2	1	priorities_nn	174	2	1	forces_nn
			police_nn2				ocean_nn1
175	2	1	officers_nn	175	2	1	side_nn
176	2	1	poverty_nn1 pay_nn	176	2	1	ocean_nn1 region_nn
			prison_nn1				opinion_nn1
177	2	1	officers_nn	177	2	1	makers_nn

			prison_nn1				opposition_nn1
178	2	1	service_nn	178	2	1	leaders_nn
			productivity_nn1				opposition_nn1
179	2	1	puzzle_nn	179	2	1	party_nn
			protest_nn1				
180	2	1	movement_nn	180	2	1	party_nn1 leaders_nn
			public_nn1				peace_nn1
181	2	1	opinion_nn	181	2	1	agreement_nn
			record_nn1				peace_nn1
182	2	1	books_nn	182	2	1	talks_nn
			referendum_nn1				
183	2	1	campaign_nn	183	2	1	planet_nn1 earth_nn
			republican_nn1				politician_nn1
184	2	1	senator_nn	184	2	1	whod_nn
			review_nn1				politicians_nn2 , ,
185	2	1	process_nn	185	2	1	educationists_nn
			rights_nn2				power_nn1
186	2	1	council_nn	186	2	1	structures_nn
187	2	1	rights_nn2 law_nn	187	2	1	print_nn1 media_nn
			rights_nn2				
188	2	1	violations_nn	188	2	1	public_nn1 opinion_nn
							response_nn1
189	2	1	safety_nn1 net_nn	189	2	1	committee_nn
			school_nn1				
190	2	1	pupils_nn	190	2	1	restraint_nn1 act_nn

			second_nnt1 ,_,				rights_nn2
191	2	1	days_nnt	191	2	1	violation_nn
			security_nn1				
192	2	1	system_nn	192	2	1	road_nn1 initiative_nn
			security_nn1				science_nn1
193	2	1	services_nn	193	2	1	minister_nn
			service_nn1				sector_nn1
194	2	1	sector_nn	194	2	1	hospitals_nn
			service_nn1				security_nn1
195	2	1	professionals_nn	195	2	1	discourse_nn
							services_nn2
196	2	1	sex_nn1 abuse_nn	196	2	1	hospital_nn
			society_nn1				state_nn1
197	2	1	weekly_nn	197	2	1	department_nn
			standards_nn2				
198	2	1	authority_nn	198	2	1	state_nn1 ideology_nn
			state_nn1				states_nn2
199	2	1	education_nn	199	2	1	part_nn
			strike_nn1				
200	2	1	action_nn	200	2	1	status_nn121 quo_nn
			support_nn1				
201	2	1	staff_nn	201	2	1	steam_nn1 engines_nn
202	2	1	tax_nn1 bill_nn	202	2	1	stock_nn1 markets_nn
			tax_nn1				stomach_nn1 ,_,
203	2	1	revenues_nn	203	2	1	eyes_nn

			taxation_nn1				support_nn1
204	2	1	policies_nn	204	2	1	programs_nn
			taxpayers_nn2				suspension_nn1
205	2	1	money_nn	205	2	1	claim_nn
			terror_nn1				teaching_nn1
206	2	1	offences_nn	206	2	1	institutions_nn
			terrorism_nn1				tear_nn1
207	2	1	bill_nn	207	2	1	gas_nn
							telenor_nn1
208	2	1	trade_nn1 talks_nn	208	2	1	suspension_nn
							thousand_nno
209	2	1	trade_nn1 deal_nn	209	2	1	years_nnt
							tourism_nn1
210	2	1	trade_nn1 talks_nn	210	2	1	industry_nn
			treaty_nn1				tourism_nn1
211	2	1	change_nn	211	2	1	ministry_nn
			troll_nn1				
212	2	1	factories_nn	212	2	1	traffic_nn1 issue_nn
							traffic_nn1
213	2	1	troll_nn1 factory_nn	213	2	1	jams_nn
							transport_nn1
214	2	1	tv_nn1 licences_nn	214	2	1	system_nn
			utility_nn1				treasury_nn1
215	2	1	companies_nn	215	2	1	benches_nn

							trial_nn1
216	2	1	war_nn1 crimes_nn	216	2	1	court_nn
			web_nn1				voucher_nn1
217	2	1	presence_nn	217	2	1	scheme_nn
			weeks_nnt2				water_nn1
218	2	1	events_nn	218	2	1	shortage_nn
			weeks_nnt2				weapon_nn1
219	2	1	summit_nn	219	2	1	systems_nn
			welfare_nn1				
220	2	1	handouts_nn	220	2	1	wheat_nn1 crisis_nn
			welfare_nn1				words_nn2 ,_ ,
221	2	1	reform_nn	221	2	1	places_nn
			welfare_nn1				
222	2	1	spending_nn	222	2	1	workers_nn2 party_nn
			women_nn2				
223	2	1	enclave_nn	223	2	1	world_nn1 country_nn

APPENDIX-H

Noun +			
Verb			
#Total No. of Cluster Types:		#Total No. of Cluster Types:	
1944	ENG	1745	PAK

#Total No. of Cluster			#Total No. of Cluster		
Tokens: 2012			Tokens: 1805		
		independence_nn1			companies_nn2
1	4	1 means_vvz	1	7	1 agree_vv
		journalism_nn1			government_nn1
2	4	1 thats_vvz	2	6	1 needs_vvz
					trees_nn2
3	4	1 labour_nn1 needs_vvz	3	6	1 produce_vv
		platforms_nn2			president_nn1
4	4	1 giving_vvg	4	5	1 oks_vvz
		rights_nn2			words_nn2
5	4	1 act_vv	5	4	1 kill_vv
		tax_nn1			bin_nn1
6	4	1 credits_vvz	6	3	1 laden_vvn
		ends_nn2			binge_nn1
7	3	1 meet_vvi	7	3	1 watching_vvg
		guardians_nn2			verdict_nn1
8	3	1 measured_vvd	8	3	1 exposed_vvd
		lib_nn1			affair_nn1 ,_
9	3	1 dems_vvz	9	2	1 followed_vvn
		security_nn1			beacon_nn1 !_!
10	3	1 based_vvn	10	2	1 sighed_vvd
		uncertainty_nn1			blame_nn1
11	3	1 abound_vv	11	2	1 lies_vvz

			activists_nn2 ,_,				chairs_nn2
12	2	1	championing_vvg	12	2	1	meeting_vvg
							deadline_nn1
13	2	1	agenda_nn1 led_vvn	13	2	1	set_vvn
			bill_nn1				decision_nn1
14	2	1	gives_vvz	14	2	1	taken_vvn
							fact_nn1
15	2	1	britains_nn2 place_vv	15	2	1	remains_vvz
			budget_nn1				field_nn1
16	2	1	marks_vvz	16	2	1	generated_vvn
			campaigners_nn2				forebears_nn2
17	2	1	want_vv	17	2	1	crushed_vvd
			challenges_nn2				fraternity_nn1
18	2	1	facing_vvg	18	2	1	include_vv
			changes_nn2				government_nn1
19	2	1	spearheaded_vvn	19	2	1	tries_vvz
			children_nn2				guru_nn1
20	2	1	born_vvn	20	2	1	nanaks_vvz
			consumers_nn2				jp_nn1
21	2	1	continue_vv	21	2	1	called_vvn
							judge_nn1
22	2	1	court_nn1 ruled_vvd	22	2	1	wrote_vvd
			flag_nn1				people_nn
23	2	1	flying_vvg	23	2	1	look_vv

			governments_nn2				people_nn
24	2	1	need_vv	24	2	1	looking_vvg
							people_nn
25	2	1	issue_nn1 facing_vvg	25	2	1	update_vv
							people_nn
26	2	1	law_nn1 banning_vvg	26	2	1	want_vv
			leader_nn1				
27	2	1	look_vv	27	2	1	people_nn watch_vv
			level_nn1				people_nn
28	2	1	playing_vvg	28	2	1	working_vvg
			levels_nn2 ,_.				plea_nn1
29	2	1	borrowing_vvg	29	2	1	filed_vvn
			minister_nn1				police_nn2
30	2	1	sees_vvz	30	2	1	reform_vv
			ministers_nn2				situation_nn1
31	2	1	said_vvd	31	2	1	got_vvd
							threat_nn1
32	2	1	nnu cut_vvn in	32	2	1	posed_vvn
			party_nn1				war_nn1 ._.
33	2	1	led_vvd	33	2	1	read_vv
			people_nn				wont_nn1
34	2	1	working_vvg	34	2	1	give_vv
			people_nn				world_nn1 ._.
35	2	1	need_vv	35	2	1	read_vv

			plan_nn1				world_nn1
36	2	1	set_vvn	36	2	1	chooses_vvz
			price_nn1				world_nn1
37	2	1	rises_vvz	37	2	1	needs_vvz
			priorities_nn2 ,_,				
38	2	1	spending_vvg	38	2	1	world_nn1 took_vvd
							xi_nn1
39	2	1	report_nn1 found_vvn	39	2	1	j Jinping_vvg
40	2	1	report_nn1 points_vvz				
			reporter_nn1				
41	2	1	asked_vvd				
			research_nn1				
42	2	1	suggests_vvz				
			shopping_nn1				
43	2	1	reminds_vvz				
44	2	1	slogans_nn2 hit_vvd				
45	2	1	tax_nn1 credits_vvz				
			version_nn1				
46	2	1	referred_vvn				
47	2	1	violence_nn1 vary_vv				
48	2	1	voters_nn2 think_vv				
			woman_nn1				
49	2	1	sleeping_vvg				
50	2	1	women_nn2 make_vv				

yesterdays_nn2			
51	2	1	report_vv

S



APPENDIX-I

VV+N			
N			
#Total No. of Cluster Types:		#Total No. of Cluster Types:	
972	ENG	841	PAK

#Total No. of Cluster Tokens:				#Total No. of Cluster Tokens:			
1032				903			
			delivering_vvg				vvi
1	10	1	quality_nn	1	10	1	roads_nn2 in
							nab_vv0
2	5	1	giving_vvg rise_nn	2	8	1	files_nn
			eliminating_vvg				make_vvi
3	3	1	poverty_nn	3	6	1	base_nn
							oks_vvz
4	3	1	find_vvi ways_nn	4	5	1	bill_nn
			keep_vvi				produce_vv0
5	3	1	guardian_nn	5	5	1	oxygen_nn
							taken_vvn
6	3	1	make_vvi sense_nn	6	4	1	place_nn
			makes_vvz				
7	3	1	sense_nn	7	4	1	taking_vvg place_nn
			playing_vvg				teaching_vvg
8	3	1	field_nn	8	3	1	hospitals_nn
9	3	1	take_vvi years_nnt	9	2	1	celebrate_vvi holi_nn
			alarm_vv0				collect_vvi
10	2	1	bells_nn	10	2	1	revenue_nn
			become_vvi				confirms_vvz
11	2	1	academies_nn	11	2	1	telenor_nn
			beggars_vvz				dengue_vvi
12	2	1	belief_nn	12	2	1	control_nn

			blamed_vvn				find_vvi
13	2	1	femicides_nn	13	2	1	relief_nn
							forgetting_vvg
14	2	1	buy_vvi shares_nn	14	2	1	words_nn
							given_vvn
15	2	1	cut_vvi carbon_nn	15	2	1	rise_nn
			cutting_vvg				
16	2	1	taxes_nn	16	2	1	hate_vv0 speech_nn
			expected_vvd . _.				help_vvi
17	2	1	mr_nnb	17	2	1	people_nn
			find_vvi				kill_vv0 . _.
18	2	1	consensus_nn	18	2	1	words_nn
			galvanised_vvn				loved_vvn
19	2	1	action_nn	19	2	1	ones_nn
							lynch_vv0
20	2	1	get_vv0 society_nn	20	2	1	mobs_nn
			hit_vv0				
21	2	1	nations_nn	21	2	1	met_vvd bureau_nn
			housing_vvg				play_vvi
22	2	1	benefit_nn	22	2	1	saviour_nn
			laughing_vvg				
23	2	1	gas_nn	23	2	1	provide_vvi relief_nn
			made_vvn				render_vvi
24	2	1	things_nn	24	2	1	musharraf_nn

							stop_vv0
25	2	1	make_vvi ends_nn	25	2	1	hate_nn
			make_vvi				take_vv0
26	2	1	womens_nn	26	2	1	place_nn
							take_vvi
27	2	1	match_vvi tax_nn	27	2	1	action_nn
			needs_vvz				taken_vvn
28	2	1	leisure_nn	28	2	1	care_nn
			needs_vvz				taken_vvn
29	2	1	time_nnt	29	2	1	oath_nn
			preventing_vvg				
30	2	1	infections_nn	30	2	1	taking_vvg shape_nn
			receiving_vvg				took_vvd
31	2	1	tax_nn	31	2	1	oath_nn
			reform_vvi				took_vvd
32	2	1	britains_nn	32	2	1	place_nn
33	2	1	see_vvi mr_nnb	33	2	1	treat_vv0 women_nn
34	2	1	sees_vvz things_nn				
35	2	1	sinn_vv0 fins_nn				
36	2	1	spending_vvg levels_nn				
37	2	1	take_vvi holidays_nn				
38	2	1	take_vvi place_nn				
39	2	1	take_vvi time_nnt				
40	2	1	womens_vvz football_nn				
41	2	1	womens_vvz strike_nn				

			won_vvd
42	2	1	labour_nn



APPENDIX-J

			VV +
			ADV

#Total No. of Cluster Types:				#Total No. of Cluster Types:			
1037		ENG		724		PAK	
#Total No. of Cluster Tokens:				#Total No. of Cluster Tokens:			
1161				898			
				read_vv0			
1	7	1	points_vvz out_rp	1	98	1	more_rrr
				facilitate_vvi			
2	5	1	end_vvi up_rp	2	11	1	online_rr
3	4	1	cut_vvn off_rp	3	8	1	set_vvi up_rp
explains_vvz				set_vvn			
4	4	1	why_rrq	4	6	1	up_rp
				going_vvg			
5	4	1	go_vvi ahead_rl	5	4	1	on_rp
				make_vvi			
6	4	1	make_vv0 up_rp	6	4	1	up_rp
pointed_vvn				wake_vv0			
7	4	1	out_rp	7	4	1	up_rp
				building_vvg			
8	4	1	rule_vvi out_rp	8	3	1	up_rp
				carried_vvn			
9	4	1	set_vvd up_rp	9	3	1	out_rp
shows_vvz							
10	4	1	how_rrq	10	3	1	come_vvi together_rl
carried_vvn				come_vvi			
11	3	1	out_rp	11	3	1	up_rp

12	3	1	clean_vvi up_rp	12	3	1	handed_vvd over_rp pointed_vvn
13	3	1	come_vvi in_rp	13	3	1	out_rp
14	3	1	come_vvi up_rp	14	3	1	sit_vv0 in_rp take_vvi
15	3	1	get_vvi on_rp	15	3	1	up_rp
16	3	1	give_vvi up_rp	16	2	1	based_vvn purely_rr bring_vvi
17	3	1	goes_vvz well_rr levelling_vvg	17	2	1	about_rp
18	3	1	up_rp	18	2	1	brought_vvn back_rp build_vv0
19	3	1	locked_vvn up_rp opened_vvn	19	2	1	up_rp build_vvi
20	3	1	up_rp	20	2	1	up_rp celebrated_vvn
21	3	1	picked_vvn up_rp	21	2	1	enough_rr discussed_vvn
22	3	1	prop_vvi up_rp	22	2	1	intellectually_rr drying_vvg
23	3	1	remain_vvi so_rr ripped_vvn	23	2	1	up_rp focusing_vvg
24	3	1	off_rp risen_vvn	24	2	1	on_rp get_vvi
25	3	1	sharply_rr	25	2	1	away_rl

							gives_vvz
26	3	1	see_vvi how_rrq	26	2	1	away_rl
							go_vv0
27	3	1	set_vvd out_rp	27	2	1	ahead_rl
							go_vv0
28	3	1	set_vvn up_rp	28	2	1	back_rp
							goes_vvz
29	3	1	sign_vvi up_rp	29	2	1	on_rp
							gone_vvn
30	3	1	went_vvd on_rp	30	2	1	up_rp
			approaching_vvg				keep_vv0
31	2	1	more_rrr	31	2	1	on_rp
							lets_vvz
32	2	1	built_vvd up_rp	32	2	1	just_rr
							living_vvg
33	2	1	carry_vvi out_rp	33	2	1	in_rp
							looking_vvg
34	2	1	carve_vvi out_rp	34	2	1	forward_rl
			coming_vvg				point_vvi
35	2	1	out_rp	35	2	1	out_rp
			dropped_vvn				pulled_vvd
36	2	1	down_rp	36	2	1	out_rp
							read_vvd
37	2	1	ended_vvd up_rp	37	2	1	more_rrr

							ruled_vvn
38	2	1	find_vvi out_rp	38	2	1	out_rp
			getting_vvg				seek_vvi
39	2	1	almost_rr	39	2	1	out_rp
							shored_vvd
40	2	1	go_vv0 far_rr	40	2	1	up_rp
							spelt_vvn
41	2	1	go_vvi down_rp	41	2	1	out_rp
							step_vvi
42	2	1	go_vvi on_rp	42	2	1	up_rp
							turned_vvd
43	2	1	go_vvi up_rp	43	2	1	out_rp
							woken_vvn
44	2	1	goes_vvz away_rl	44	2	1	up_rp
45	2	1	goes_vvz on_rp	45	2	1	work_vvi together_rl
46	2	1	going_vvg on_rp				
47	2	1	grown_vvn up_rp				
48	2	1	hollowed_vvn out_rp				
49	2	1	keep_vvi up_rp				
50	2	1	live_vvi on_rp				
			locking_vvg				
51	2	1	up_rp				
52	2	1	look_vv0 on_rp				
53	2	1	move_vvi forward_rl				
54	2	1	move_vvi on_rp				

			moved_vvn
55	2	1	on_rp
			opening_vvg
56	2	1	up_rp
57	2	1	pay_vvi off_rp
			pointed_vvd
58	2	1	out_rp
			put_vvn
59	2	1	forward_rl
60	2	1	put_vvn off_rp
61	2	1	remains_vvz fundamentally_rr
62	2	1	resigned_vvd over_rp
			ripping_vvg
63	2	1	up_rp
64	2	1	send_vvi in_rp
65	2	1	set_vv0 aside_rl
66	2	1	set_vvn out_rp
67	2	1	shore_vvi up_rp
68	2	1	shown_vvn at_rr
69	2	1	shows_vvz how_rrq
70	2	1	sign_vv0 up_rp
			snuffed_vvd
71	2	1	out_rp
72	2	1	take_vvi on_rp
73	2	1	take_vvi up_rp

74	2	1	taking_vvg more_rgr talking_vvg
75	2	1	up_rp
76	2	1	top_vvi up_rp topping_vvg
77	2	1	up_rp
78	2	1	vary_vv0 immensely_rr
79	2	1	work_vv0 out_rp
80	2	1	work_vvi together_rl



APPENDIX-K

Students' Interview questions

This is a semi-structured interview purely designed to seek your opinion relevant to the research study. Your involvement will be kept secret. Please answer the following questions earnestly. Your response would be highly appreciated.

1. What is your opinion about English language teaching/learning in your country?

2. You have been learning English language since your childhood. How do you like to use English language in routine communication?
3. What is your opinion about the existing English language teaching/learning approaches?
4. According to your opinion, which approach is better in language learning/acquisition in the perspective of inductive and deductive approaches?
5. What do you mean by Corpus Driven Language Learning, how you come to know this term?
6. Have you processed concordance, while concordancing you focus on meaning or form?
7. What persuaded you to prefer CDLL approach in language learning?

APPENDIX-L

4/21/2020 7:40 PM

Coding Summary By Code

Qualitative Analysis NVivo

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Node

Nodes\\Corpus Driven Language Learning

Document

Files\\Ahmad

No		0.1287	1			
				1	SH	4/14/2020 11:55 AM

Corpus Driven Language Learning is an inductive approach learning language in which a large number of corpora is utilized and learners are asked to do pattern hunting that helps them in identifying linguistic patterns. I came to know this

Files\\Areej

No		0.0564	1			
				1	SH	4/14/2020 12:13 PM

Corpus is actually the data of authentic language collected electronically. The language learned from corpus by operating on it, is called corpus driven language learning.

Files\\Fazeelat

No		0.0859	1			
				1	SH	4/14/2020 12:21 PM

Corpus driven approach is inductive learning in that it chooses a targeted corpus and cyclical analysis and interpretation of corpus data

Files\\Isha

No		0.0632	1			
				1	SH	4/14/2020 12:26 PM

Corpus is a big collection of naturally occurring language produced by native speakers gathered information from both speaking and written language can find difference between formal and informal language with the use of corpus

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\rida

No		0.0529	1
----	--	--------	---

1	SH	4/14/2020 12:17 PM
---	----	--------------------

Corpus driven learning means that a person learns language by exploring different features of language in the collection of the authentic text, representative of the language.

Files\\usama shoukat

No		0.0860	1
----	--	--------	---

1	SH	4/14/2020 12:57 PM
---	----	--------------------

Learning any language with the help of prepared corpus (a collection of written texts, especially the entire works of a particular author or a body of writing on a particular subject) while understanding, analysing and then using the features

Files\\zainab

No		0.0693	1
----	--	--------	---

1	SH	4/14/2020 1:02 PM
---	----	-------------------

Corpus Driven Language Learning means to analyse and interpret a large collection of real-life language (i.e language used in routine communication

Files\\Zulykha

No		0.0567	1
----	--	--------	---

1	SH	4/14/2020 1:10 PM
---	----	-------------------

In CDLL, a larger amount of text is collected, stored in computers or in hard form and then various features are analyzed. Corpus is full of expressions of native speakers. Students recognize patterns of language, frequency,

Nodes\\Corpus Driven Language Learning\\Autonomous Learning

Document

Files\\rida

No	0.0745	1			
			1	SH	4/14/2020 12:48 PM

One of the major reason was autonomous learning. It gave me freedom to learn independently. It includes the use of advance apps and softwares. I love to use softwares and applications for learning. It is fun to learn and enjoy at the same

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\usama shoukat

No	0.0785	1			
			1	SH	4/14/2020 12:58 PM

I think it is better to explore the knowledge without the teacher's pressure. I think it can make me more active and I can perform better in that language by connecting different people and studying different materials of different authors on a

Files\\Zulykha

No	0.0292	1			
			1	SH	4/14/2020 1:11 PM

We somehow focus on the meaning but our main focus is on the form of words. Grammar, vocabulary etc. We check the sense of the words.

Nodes\\Corpus Driven Language Learning\\CDLL Preference

Document

Files\\Ahmad

No	0.0601	1			
----	--------	---	--	--	--

1 SH 4/14/2020
11:56 AM

Corpus Based approach helps me in constructing my own knowledge. That is one important factor that persuaded me to prefer it.

Files\\Isha

No 0.0576 1

1 SH 4/14/2020
12:31 PM

In CDLL approach corpus and collocations are finds and gives proper meanings of information both in speaking and written language, corpus plays a vital role in DDL to know which type of data is used in CALL.

Files\\khalil

No 0.0514 1

1 SH 4/14/2020
12:42 PM

Because in traditional way of learning student just cramm thing to close the mouth of teacher or his examiner and that is not not learning .Specially while learning language we should not learn things to close mouth of examiner



Reports\\Coding Summary By Code Report

Page 3 of 22

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\zainab

No 0.0969 1

1 SH 4/14/2020
1:03 PM

CDLL is a unique approach that enables the user of the English language to get access to authentic text every time including hundreds and thousands of words, phrases, and sentences with just a single click.

Files\\Zulykha

No 0.0221 1

1 SH 4/14/2020
1:12 PM

CDLL is an exciting approach and student become active member by being involved in exploring the text.

Nodes\\Corpus Driven Language Learning\\Concordance

Document

Files\\Ahmad

No	0.0482	1
----	--------	---

1	SH	4/14/2020 11:56 AM
---	----	-----------------------

Yes, I have processed concordance. While concordancing I focus more on the meaning of lexical items.

Files\\Fazeelat

No	0.0186	1
----	--------	---

1	SH	4/14/2020 12:21 PM
---	----	-----------------------

Concordance focus on meaning

Files\\Isha

No	0.0545	1
----	--------	---

1	SH	4/14/2020 12:27 PM
---	----	-----------------------

Concordance is an essential tool to gain access to corpus, is a computer program utilize for searching and analyzing a particular word, it helps learners to search target word by typing that word

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\Sara kalsoom

No	0.0636	1
----	--------	---

1	SH	4/14/2020 12:52 PM
---	----	-----------------------

while concordancing we usually use the activities that are mainly focused on form and we also use corpus based on forms .

Nodes\\Corpus Driven Language Learning\\Corpus Information

Document

Files\\Isha

No	0.0596	1			
			1	SH	4/14/2020 12:27 PM

in corpus we can find collocations with the help of software COCA and many QUIZLET, we come to know about Corpus through the research given by our professor in which we edit make editorials in the form of Corpus.

Files\\khalil

No	0.0320	1			
			1	SH	4/14/2020 12:40 PM

I firstly heard this term while studying DDL approaches and I think is the best method for the university learners for any language learning

Files\\rida

No	0.0314	1			
			1	SH	4/14/2020 12:47 PM

Corpus is the collection of text of the target language. I came to know about CDLL through my teacher.

Files\\Sara kalsoom

No	0.0318	1			
			1	SH	4/14/2020 12:51 PM

my teacher used this term and i first heard this from him .

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\usama shoukat

No		0.0430	1
----	--	--------	---

1	SH	4/14/2020 12:57 PM
---	----	-----------------------

I firstly heard this term while studying DDL approaches and I think is the best method for the university learners for any language learning

Files\Zulykha

No		0.0290	1
----	--	--------	---

1	SH	4/14/2020 1:10 PM
---	----	----------------------

I came to know this term by scholarly articles shared by my teacher. I have practiced this approach because it's a part of my studies.

Nodes\Corpus Driven Language Learning\Features of Language in Corpus**Document****Files\khalil**

No		0.0635	1
----	--	--------	---

1	SH	4/14/2020 12:40 PM
---	----	-----------------------

Learning any language with the help of prepared corpus (a collection of written texts, especially the entire works of a particular author or a body of writing on a particular subject) while understanding, analysing and then using the features

Files\Zulykha

No		0.0305	1
----	--	--------	---

1	SH	4/14/2020 1:00 PM
---	----	----------------------

DDL is interaction between corpus and student. This technique is helpful to learn language by exploring different features of authentic text.

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
Nodes\\Corpus Driven Language Learning\\Generalizing in CDDL						
Document						
Files\\zainab						
No		0.0693	1			
				1	SH	4/14/2020 1:07 PM
helps in generalising a set of rules or particular theory. I came to know this term in my university life as it is currently a part of my degree.						

Nodes\\Corpus Driven Language Learning\\Idea of CDL**Document****Files\\Areej**

No		0.0331	1			
				1	SH	4/14/2020 12:12 PM
I got to know about corpus driven language learning in 3rd semester of my BS program in university.						

Files\\Fazeelat

No	0.0236	1
----	--------	---

1	SH	4/14/2020 12:21 PM
---	----	-----------------------

I got this information from my teacher

Reports\\Coding Summary By Code Report

Page 7 of 22

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------------	---------------------	----------------------	----------------

Nodes\\Corpus Driven Language Learning\\Line in concordance

Document

Files\\Isha

No	0.0465	1
----	--------	---

1	SH	4/14/2020 12:28 PM
---	----	-----------------------

in concordance we focus on line and on that line we find the simple meanings of words, by finding the sense of a line we focus on words which gives proper meanings.

Nodes\\Corpus Driven Language Learning\\Meaning in Concordance

Document

Files\\zainab

No	0.0688	1			
			1	SH	4/14/2020 1:03 PM

While concordancing, I focus on meaning because even a single word expresses various meanings depending on the context, user, or situation, etc.

Nodes\\Corpus Driven Language Learning\\Multiple exposure of words

Document

Files\\Areej

No	0.0324	1			
			1	SH	4/14/2020 12:15 PM

By operating on corpus, you get the multiple exposures of a word you search in target language. Then

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------------	---------------------	----------------------	----------------

Nodes\\Corpus Driven Language Learning\\Process of concordance

Document

Files\\Areej

No	0.0396	1			
----	--------	---	--	--	--

1 SH 4/14/2020
12:14 PM

I did process concordance using ant conc, COCA and BNC. I usually focus on form; the way a word is used in the context.

Files\\khalil

No 0.0356 1

1 SH 4/14/2020
12:41 PM

I have processed concordance while using COCA .and i mostly focused on the meaning of words because i found much more words with which i was quite unfamiliar

Files\\usama shoukat

No 0.0703 1

1 SH 4/14/2020
12:57 PM

I have processed concordance while using COCA. While concordancing my main focus was on form but sometimes I also focus on meaning because there were some new words for me and I couldn't understand the actual forms and their

Files\\Zulykha

No 0.0305 1

1 SH 4/14/2020
1:10 PM

we have processed concordance by using several softwares. We have complied corpus of editorials and then processed concordance by using COCA.

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Nodes\\ELT Approaches

Document

Files\\Ahmad

No	0.0629	1				
			1	SH	4/14/2020	11:53 AM

I think the existing English language teaching/learning approaches are dated & needs drastic updates so that the output is better.

Files\\Sara kalsoom

No	0.0870	1				
			1	SH	4/14/2020	12:51 PM

These approaches teaches us vocabulary, grammar and improve our speaking and skills , in my opinion these approaches are basic need of a person who wants to learn .

Files\\zainab

No	0.0978	1				
			1	SH	4/14/2020	1:01 PM

the existing approaches like DDL, CALL, CDLL which have enhanced the acceptance, usage, and productivity of English language for all learners, regardless of their economic, social, or educational background.

Nodes\\ELT Approaches\\ALM Approach

Document

Files\\khalil

No	0.0378	1				
			1	SH	4/14/2020	12:28 PM

ALM is being widely used in our country .I think this method is not good because when children cramm the things they make mind that learning is possible with cramming .

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\usama shoukat

No		0.0554	1	1	SH	4/14/2020 12:55 PM
----	--	--------	---	---	----	--------------------

ALM is widely used in my country. I think it is good at initial level but I am not satisfied with this method at matric or higher level. It doesn't improve language speaking ability.

Files\\Zulykha

No		0.0382	1	1	SH	4/14/2020 1:08 PM
----	--	--------	---	---	----	-------------------

Audio-lingual Method and its still popular today, and whose influence can be seen in a variety of drill based techniques and exercises. It is based on theory of behaviourism.

Nodes\\ELT Approaches\\Better Approach of ELT**Document****Files\\Ahmad**

No		0.0899	1	1	SH	4/14/2020 11:54 AM
----	--	--------	---	---	----	--------------------

In my opinion, both the approaches have their pros and cons and used accordingly. Inductive approach takes a lot of practice & skill that new comer would lake but it is also more fruitful.

Nodes\\ELT Approaches\\DDL Approach

Document

Files\\Areej

No	0.0457	1			
			1	SH	4/14/2020 12:12 PM

you can integrate DDL in your classroom if your students are willing to learn more and you have enough equipments for a language classroom.

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------------	---------------------	----------------------	----------------

Nodes\\ELT Approaches\\Deductive Approach in ELT

Document

Files\\khalil

No	0.0223	1			
			1	SH	4/14/2020 12:38 PM

Deductive approach is better for language beginners as they have not much knowledge about language.

Files\\Sara kalsoom

No	0.0844	1			
			1	SH	4/14/2020 12:51 PM

IN MY WAY/ OPINION deductive reasoning/approaches is better than inductive because it starts with a statement or hypothesis and then tests to see if it's true .

Files\\usama shoukat

No 0.0454 1

1 SH 4/14/2020
12:56 PM

Deductive approach is better for language beginners as they have not much knowledge about language. But when they have learnt language to some extent,

Nodes\\ELT Approaches\\GTM Approach

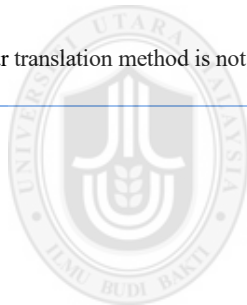
Document

Files\\Areej

No 0.0370 1

1 SH 4/14/2020
12:11 PM

Grammar translation method is not going to help us in our practical lives. We need to change our teaching system



UUM
Universiti Utara Malaysia

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\Isha

No 0.1036 1

1 SH 4/14/2020
12:25 PM

The approach which is still used in Pakistan is GTM(Grammar Translation Method) in which focus is on vocabulary and pronunciation because it gives fluency, accuracy and GTM translates native language into target language, by using this method teachers explain target language into native language and gives information about replacement of words and

Files\\Zulykha

No	0.0479	1			
			1	SH	4/14/2020 1:07 PM

Grammar Translation Method is a traditional method of language teaching still used in our country. According to this method language is learned by translating, focus is on accuracy, grammar rules. Students role is passive.

Nodes\\ELT Approaches\\Inductive Approach

Document

Files\\Fazeelat

No	0.0492	1			
			1	SH	4/14/2020 12:20 PM

According to my experience inductive learning is better for language learning

Files\\Isha

No	0.0885	1			
			1	SH	4/14/2020 12:26 PM

Inductive approach is better as in this approach we can make one's own rule and constructs our knowledge through different sources, data is collected on the possible reasons why it is occurs and the experiments are done on that hypothesis, as in deductive approach theory rejected or accepted which is not confirmed.

Files\\khalil

No	0.0547	1			
			1	SH	4/14/2020 12:30 PM

when they have learnt language to some extent, then they should use inductive method. I think it is better because it tends learners to formalize the rules by analysing different examples. It improves their analytical and researching skills.

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\usama shoukat

No		0.0578	1			
				1	SH	4/14/2020 12:56 PM

they should use inductive method. I think it is better because it tends learners to formalize the rules by analysing different examples. It improves their analytical and researching skills.

Files\\Zulykha

No		0.0901	1			
				1	SH	4/14/2020 1:00 PM

Inductive approach is much better than deductive for language acquisition. It is base of DDL. It entirely depends on noticing by students or teachers. Student becomes active and acquire knowledge by interacting with the authentic text. Students learn new structures, laws are not defined. Language is explored. Hypothesis is made through observation and

Nodes\\ELT Approaches\\option about CLT approach

Document

Files\\Areej

No		0.0233	1			
				1	SH	4/14/2020 12:12 PM

I think communicative approach is best to teach students at school level

Files\\rida

No		0.0754	1			
				1	SH	4/14/2020 12:46 PM

teachers now communicate with students in English, but it is still not enough to meet the needs of actual and more practical learning. We need to use more advance modes of learning language such as, use of advance applications to

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\Zulykha

No		0.0649	1
----	--	--------	---

1	SH	4/14/2020
---	----	-----------

1:08 PM

CLT(Communicative language teaching) method is used. It's an approach that emphasizes authenticity, interaction, student-centered learning and task based activities for real world. So CLT is much better for language learning. Students learn by role play activities, dialogues, reading magazines etc.

Nodes\\English Language in Pakistan**Document****Files\\Ahmad**

No		0.0440	1
----	--	--------	---

1	SH	4/14/2020
---	----	-----------

11:51 AM

it should be taught and learned in Pakistan and it is being taught and leaned in Pakistan.

Files\\zainab

No		0.0997	1
----	--	--------	---

1	SH	4/14/2020
---	----	-----------

12:50 PM

an exceptional way that serves as a platform to improve the economic condition of the country, embracing the career, study, and travel opportunities that come with teaching and learning of this foreign language.

Nodes\\English Language in Pakistan\\Encouragement of Peers in Speaking Eng

Document

Files\\khalil

No	0.0620	1				
			1	SH	4/14/2020	12:37 PM

In university i get the chance to speak english and with my best friends only .Because they are not counting my mistakes ,they just want to improve my language learning ability .They encourage me and ofcourse i speak english with them and

Reports\\Coding Summary By Code Report

Page 15 of 22

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\usama shoukat

No	0.0578	1				
			1	SH	4/14/2020	12:55 PM

yes, When I am talking with my fellows specially Khalil, Shahzaib we use English language as we are student of this language. But in short, we do only 5% communication in English language.

Nodes\\English Language in Pakistan\\Learning English

Document

Files\\Fazeelat

No	0.0760	1				
			1	SH	4/14/2020	12:16 PM

Learing english us compulsory for doing a good job our country has a better learning system to now instead of old system

Files\\Sara kalsoom

No 0.0354 1

1 SH 4/14/2020 12:50 PM

In my opinion ,it has many benefits and uses for student and teacher

Files\\zainab

No 0.1348 1

1 SH 4/14/2020 1:01 PM

the inductive approach is better because it is task-based knowledge and involves experiment, reasoning, and applicational methods. It is based on explicit knowledge and is full of activity for the language learner, training his/her



Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Nodes\\English Language in Pakistan\\Listening & Speaking of University students

Document

Files\\Isha

No 0.0540 1

1 SH 4/14/2020 12:22 PM

speaking and listening capability in students becomes week due to this at university level students hesitate to speak English and give most of their time to correct their Grammar pronunciation,

Nodes\\English Language in Pakistan\\Personal Stance about ELT in Pak

Document

Files\\Areej

No	0.0548	1			
			1	SH	4/14/2020 11:59 AM

sp speaking and listening capability in students becomes weak due to this at university level students hesitate to speak English and give most of their time to correct their Grammar pronunciation

Files\\Isha

No	0.0317	1			
			1	SH	4/14/2020 12:23 PM

teachers should give their most of their time to students on speaking and listening at initial stage of learning.

Files\\khalil

No	0.0340	1			
			1	SH	4/14/2020 12:22 PM

I think in Pakistan most of the students are not willing to learn English. They think it boring and difficult subject during their matric or fsc class .

Files\\rida

No	0.1814	2			
			1	SH	4/14/2020 12:43 PM

I have observed, English language is learnt by students just to pass the exams till matric level. There is no other aim in the minds of students. Only a few people consider learning the language for the sake of communication and connectivity

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

2 SH 4/14/2020
12:44 PM

I personally faced many issues while learning English in school. The most basic problem was that teachers did not teach by following an appropriate method. I used to cram the meanings of words and how these words form meaning, I had no practical knowledge. Teachers had really wrong pronunciation of the words with no accent.

Files\\usama shoukat

No 0.0215 1

1 SH 4/14/2020
12:54 PM

I think it is not meeting with the world demands of English language.

Files\\Zulykha

No 0.1114 2

1 SH 4/14/2020
1:06 PM

People having good skills in English are admired. Language learning courses like TOFEL, ILETS are delivered and people pay much cost just to have Proficiency in English skills. People are good in Reading and Writing but not much

2 SH 4/14/2020
1:07 PM

Much stress was given to reading and writing. Oral language development was neglected in classrooms. Now being a university student we have language competence but we lack performance. We feel hesitation while producing our own

Nodes\\English Language in Pakistan\\Poor Teaching System

Document

Files\\Areej

No 0.0327 1

1 SH 4/14/2020
11:58 AM

There is a poor teaching system in our country especially for English language teaching or learning.

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\Isha

No		0.0501	1
----	--	--------	---

1	SH	4/14/2020 12:21 PM
---	----	-----------------------

since childhood we learn English language but our most focus is on reading and writing which weakens our speaking capability but through the best focus on Grammar we speak freely,

Nodes\English Language in Pakistan\Primary Level Teachers in Pakistan**Document****Files\khalil**

No		0.0441	1
----	--	--------	---

1	SH	4/14/2020 12:33 PM
---	----	-----------------------

In schools students are not properly guided specially in villages teachers are matric pass and they also teach English in cramming way and they don't create interest of students in This language .

Files\Zulykha

No		0.0451	1
----	--	--------	---

1	SH	4/14/2020 1:06 PM
---	----	----------------------

English language teaching and learning is given much importance in Private sectors mostly but now it is being taught from primary level in public sector educational institutes too. English is a global language.

Nodes\\English Language in Pakistan\\Source of English Learning

Document

Files\\usama shoukat

No	0.0724	1				
			1	SH	4/14/2020 12:53 PM	

The major source of English language learning is classrooms where English is not used as target language, rather we perform boring English spelling drills, some formal grammatical rules and practise and precise definitions of English

Reports\\Coding Summary By Code Report

Page 19 of 22

4/21/2020 7:40 PM

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------------	---------------------	----------------------	----------------

Nodes\\English Language in Pakistan\\Students of Underprivileged areas

Document

Files\\khalil

No	0.0306	1				
			1	SH	4/14/2020 12:24 PM	

students who are from lower areas dont speak english in class due to having low confidence .They can not properly learn it as language

Nodes\\English Language in Pakistan\\Usage of English in Routine

Document

Files\\Ahmad

No	0.0743	1			
			1	SH	4/14/2020 11:52 AM

Although I don't use it very often in daily communication, I use it a lot especially in more formal situations but in informal situation that use is limited.

Files\\Fazeelat

No	0.0598	1			
			1	SH	4/14/2020 12:17 PM

I like to use english during conversation with my parents and also with my all family members

Files\\Isha

No	0.0657	1			
			1	SH	4/14/2020 12:25 PM

we can use English in routine communication by speaking and learning new words of vocabulary through it our vocabulary would enhance and teacher should also appreciate and cooperate students on their knowledge and ability of

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
-----------	----------------	----------	-----------------------------	------------------	-------------------	-------------

Files\\khalil

No	0.0270	1			
			1	SH	4/14/2020 12:36 PM

in my village speaking English is really difficult .Because people make fun and it is useless to speak it in my village

Files\\rida

No	0.0529	1			
			1	SH	4/14/2020 12:45 PM

I did not found any favorable environment to communicate in English here. Most of the people do not know this language. People usually consider such person proudy and rude.

Files\\Sara kalsoom

No	0.0849	1			
			1	SH	4/14/2020 12:50 PM

I use English with some of my elders those who understand it , and practice it with children by teaching them english and it also makes my pronunciation better .

Files\\usama shoukat

No	0.0463	1			
			1	SH	4/14/2020 12:54 PM

I'm living in a small town of Jhang, It is difficult and useless to use English language here because many will not understand and some will make fun too

Files\\zainab

No	0.1213	1			
			1	SH	4/14/2020 1:00 PM

I prefer the use of the English Language in the form of speaking because the way you speak directly reflects your personality and your academic life. My fluent speaking skill always boosts my confidence and enable me to convey my

Aggregate	Classification	Coverage	Number Of Coding References	Reference Number	Coded By Initials	Modified On
Files\Zulykha						
No		0.0303	1			
				1	SH	4/14/2020
Speaking English in daily life is not a part of our habit but we should use English language in routine communication. We run after accuracy.						



APPENDIX-M



Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

Teachers' View about Corpus Driven Language Learning

Dear Participants

I would like to seek your views regarding corpus Driven Language Learning for research purpose. I am doing doctoral research into language learning of Pakistani undergraduates particularly, the effect of CDLL on learning foreign language (ESL/EFL). I would like to have your views for my data analysis.

Please read the following questions and answer them. Your contribution would matter positively in my research thesis. It is purely research oriented, your responses will be mentioned anonymously in the document.

Interview Questions

1. What is your general views regarding the English language learning status in Pakistan?
2. There are certain approaches of language teaching/learning, which approach/method you have preferred while teaching/learning, give reason?
3. Do you think development of procedural knowledge is better than that of declarative in case of ESL/EFL?
4. You are well aware of Corpus Linguistics, Have you ever compiled corpus for a specific purpose while teaching/learning?
5. What is your idea about Data Driven Learning method in general?
6. Corpus driven language Learning (CDLL) is actually derived from Data Driven Learning; have you ever taught or discussed this topic at any stage of teaching/learning during your academic career?
7. Do you think the method is worthy enough to be included in the course syllabus of ELT Method & Approaches of the degree program of the university?

Appendix-N

Code Summary

Faculty Interview

4/21/2020 12:26 PM

File Type	Number of Files	Number of Coding References	Number of Words Coded	Number of Paragraphs Coded
-----------	-----------------	-----------------------------	-----------------------	----------------------------

Node

Nickname: Nodes\\Effectiveness of CDLL

Classification:

Aggregated: No

Document	2	2	49	2
----------	---	---	----	---

Nickname: Nodes\\Effectiveness of CDLL\\Awareness of Corpus Linguistics

Classification:

Aggregated: No

Document	4	4	110	4
----------	---	---	-----	---

Nickname: Nodes\\Effectiveness of CDLL\\CDLL meets the requirements of International level

Classification:

Aggregated: No

Document	5	5	83	5
----------	---	---	----	---

Nickname: Nodes\\Effectiveness of CDLL\\Corpus tools for Pedagogical purposes

Classification:

Aggregated: No

Document	5	5	129	5
----------	---	---	-----	---

File Type	Number of Files	Number of Coding References	Number of Words Coded	Number of Paragraphs Coded
-----------	-----------------	-----------------------------	-----------------------	----------------------------

Nickname: Nodes\\Effectiveness of CDLL\\Corpus tools for Pedagogical purposes\\Awareness of Corpus Linguistics
Classification:

Aggregated: No

Document	3	4	114	4
----------	---	---	-----	---

Nickname: Nodes\\Effectiveness of CDLL\\Corpus tools for Pedagogical purposes\\syllabus contents & DDL
Classification:

Aggregated: No

Document	2	2	82	2
----------	---	---	----	---

Nickname: Nodes\\Effectiveness of CDLL\\Corpus tools for Pedagogical purposes\\Teaching Grammar & Data Driven Learning
Classification:

Aggregated: No

Document	1	1	20	1
----------	---	---	----	---

Nickname: Nodes\\Effectiveness of CDLL\\Idea

Classification:

Aggregated: No

Document	3	3	106	3
----------	---	---	-----	---

Nickname: Nodes\\Effectiveness of CDLL\\Significance of DDL

Classification:

Aggregated: No

Document	4	4	109	4
----------	---	---	-----	---

Reports\\Code Summary Report

Page 2 of 7

4/21/2020 12:26 PM

File Type	Number of Files	Number of Coding References	Number of Words Coded	Number of Paragraphs Coded
-----------	-----------------	-----------------------------	-----------------------	----------------------------

Nickname: Nodes\\Effectiveness of CDLL\\Teachers' Training and CDLL

Classification:

Aggregated: No

Document	1	1	22	1
----------	---	---	----	---

Nickname: Nodes\\English language in Pakistan

Classification:

Aggregated: No

Document	7	7	176	7
----------	---	---	-----	---

Nickname: Nodes\\English language in Pakistan\\Desire to learn English

Classification:

Aggregated: No

Document	2	2	97	2
----------	---	---	----	---

Nickname: Nodes\\English language in Pakistan\\English as Second language

Classification:

Aggregated: No

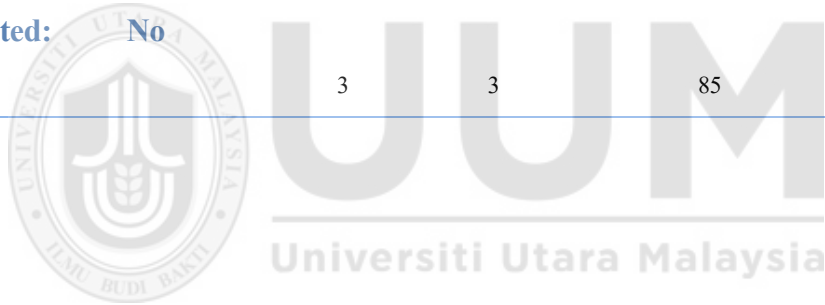
Document	1	1	21	1
----------	---	---	----	---

Nickname: Nodes\\English language in Pakistan\\English Teachers' incompetency

Classification:

Aggregated: No

Document	3	3	85	3
----------	---	---	----	---



File Type	Number of Files	Number of Coding References	Number of Words Coded	Number of Paragraphs Coded
-----------	-----------------	-----------------------------	-----------------------	----------------------------

Nickname: Nodes\\English language in Pakistan\\Factors affecting Language Learning

Classification:

Aggregated: No

Document	2	2	68	2
----------	---	---	----	---

Nickname: Nodes\\English language in Pakistan\\Functional approach of English in Socio-political scenario

Classification:

Aggregated: No

Document	1	1	43	1
----------	---	---	----	---

Nickname: Nodes\\English language in Pakistan\\learners' inconvenience

Classification:

Aggregated: No

Document	2	2	140	2
----------	---	---	-----	---

Nickname: Nodes\\English language in Pakistan\\Medium of teaching

Classification:

Aggregated: No

Document	1	1	60	1
----------	---	---	----	---

Nickname: Nodes\\English language in Pakistan\\requirements of EFL & ESL in Pakistan

Classification:

Aggregated: No

Document	1	1	28	1
----------	---	---	----	---

File Type	Number of Files	Number of Coding References	Number of Words Coded	Number of Paragraphs Coded
-----------	-----------------	-----------------------------	-----------------------	----------------------------

Nickname: Nodes\\English language in Pakistan\\Social disparities in English Language Learning

Classification:

Aggregated: No

Document	2	2	73	2
----------	---	---	----	---

Nickname: Nodes\\English language in Pakistan\\Status of English as an Official Language

Classification:

Aggregated: No

Document	1	1	23	2
----------	---	---	----	---

Nickname: Nodes\\English language in Pakistan\\Student background

Classification:

Aggregated: No

Document	2	2	70	2
----------	---	---	----	---

Nickname: Nodes\\Language Learning Approaches

Classification:

Aggregated: No

Document	1	1	8	1
----------	---	---	---	---

Nickname: Nodes\\Language Learning Approaches\\GTM with CLT approach

Classification:

Aggregated: No

Document	4	4	156	4
----------	---	---	-----	---

4/21/2020 12:26 PM

File Type	Number of Files	Number of Coding References	Number of Words Coded	Number of Paragraphs Coded
-----------	-----------------	-----------------------------	-----------------------	----------------------------

Nickname: Nodes\\Language Learning Approaches\\Idea of Corpus Driven Language Learning

Classification:

Aggregated: No

Document	3	3	90	3
----------	---	---	----	---

Nickname: Nodes\\Language Learning Approaches\\Mixed Method Approach

Classification:

Aggregated: No

Document	2	2	102	2
----------	---	---	-----	---

Nickname: Nodes\\Language Learning Approaches\\Personal stance

Classification:

Aggregated: No

Document	4	4	129	4
----------	---	---	-----	---

Nickname: Nodes\\Language Learning Approaches\\Prefernece of Procedural Knowledge

Classification:

Aggregated: No

File Type	Number of Files	Number of Coding References	Number of Words Coded	Number of Paragraphs Coded
-----------	-----------------	-----------------------------	-----------------------	----------------------------

Relationship

Nickname: Relationships\\Amna Arshad (Farukh) nosheen

Classification:

Aggregated: No

00
