



## Core Urdu Vocabulary for Chines Business Community in Pakistan, A Corpus-based Perspective

<sup>1</sup> Abrar Hussain Qureshi, <sup>2</sup> Shamim Akhter, <sup>3</sup> Musarat Shaheen

<sup>1</sup> Institute of Southern Punjab Multan, Pakistan: [abrarqureshi74f@gmail.com](mailto:abrarqureshi74f@gmail.com)

<sup>2</sup> Institute of Southern Punjab Multan, Pakistan: [misschudry96@gmail.com](mailto:misschudry96@gmail.com)

<sup>3</sup> Bahauddin Zakaria University Multan, Pakistan: [zainach530@gmail.com](mailto:zainach530@gmail.com)

ARTICLE DETAILS	ABSTRACT
<p><b>History</b>  <i>Revised format: 30 Nov 2019</i>  <i>Available Online: 31 Dec 2019</i></p> <hr/> <p><b>Keywords</b>  <i>Urdu, Language Learning, Vocabulary Building, Corpus, Word List, Pakistan</i></p> <hr/> <p><b>JEL Classification:</b>  <i>N10, N15</i></p>	<p>With the dawn of 21st century, the world has grown into a global village and the need for inter-communal interactions has also increased many times. Urdu language is said to be one the third biggest language of the world along with Chines and English and its speakers are constantly on the rise. With the emergence of the CPEC (China Pakistan Economic Corridor), Urdu has assumed ever increasing importance due to the geo-political and geo-economic condition of the south Asian region. The undertaken study is a systematic attempt in this regard to work out a list of most frequent words of the Urdu language for the Chinese business community in Pakistan. Schmitt (2000) asserts that that learning a non-native vocabulary is a continual process as the core vocabulary should encompass the ever changing linguistic needs of the time. The name of the Urdu corpus for the undertaken research is urTenTen that has been compiled from internet data. The corpus belongs to TenTen corpus family that is corpora of the web with more than ten billion words. The corpus has been tagged according to Unified Parts of Speech (POS) Standard in Indian Languages. In order to process data, “sketch Engine” has been used. List of frequent words for the Chines Business Community has been retrieved from urTenTen corpus with the help of sketch engine. The retrieved list of core Urdu vocabulary is supposed to be useful for the Chines business who is supposed to interact with the Urdu speakers of the region.</p>
<p>© 2019 The authors, under a Creative Commons Attribution-Non Commercial 4.0</p>	

Corresponding author's email address: [misschudry96@gmail.com](mailto:misschudry96@gmail.com)

**Recommended citation:** Qureshi, A. Q., Akhter, S. and Shaheen, M., (2019). Core Urdu Vocabulary for Chines Business Community in Pakistan, A Corpus-based Perspective. *Review of Economics and Development Studies*, 5 (4), 637-646

**DOI:** 10.26710/reads.v5i4.893

### 1. Introduction

Pakistan is a developing country of South Asia. The region has been under colonial rule for many decades. The result is disintegration and the deteriorating economy. Pakistan is not an exception in this regard. In these pursy times, CPEC is a massive opportunity to boost Pakistan Economy. China-Pakistan Economic Corridor (CPEC) is the major part of the One Belt One Road (OBOR) that has roots in the traditional Silk Road. This new trade facility will not only provide China a cost effective chance of trade to global regions but it will also upgrade and expand the existing Pakistani economy and infrastructure.

With the emergence of China-Pakistan Economic Corridor, there have also been signed many agreements between China and Pakistan. For this purpose, visits of the officials of both the countries have increased many times. Apart from these visits, many Chinese officials are constantly present in Pakistan for technical assistance. There is a significant language barrier between the communities of Pakistan and China. Moreover, Chinese and Urdu are absolutely different languages that have opposite linguistic origins. As a result, the officials of both the countries have to face a lot of linguistic obstacles that may hamper the true spirit of China-Pakistan Economic Corridor. Consequently, a wordlist of Urdu language that is quite frequent, is the need of the hour to help the Chinese officials in Pakistan for successful communication.

Significance of vocabulary especially in foreign language learning and teaching is an established phenomenon (Biemiller, 2004). It determines the procedure of foreign language learning. The traditional methods to handle foreign vocabulary have hampered linguistic creativity of the language learners.

One of the reasons of this obstructed linguistic creativity in the foreign language learners is the teachers centered approach of the learners of the foreign language who are at the lowest paradigm of autonomy (Neuman & Dwyer, 2009). In traditional foreign language class rooms, the teacher gets his own practice of vocabulary items and the learners' role is very passive.

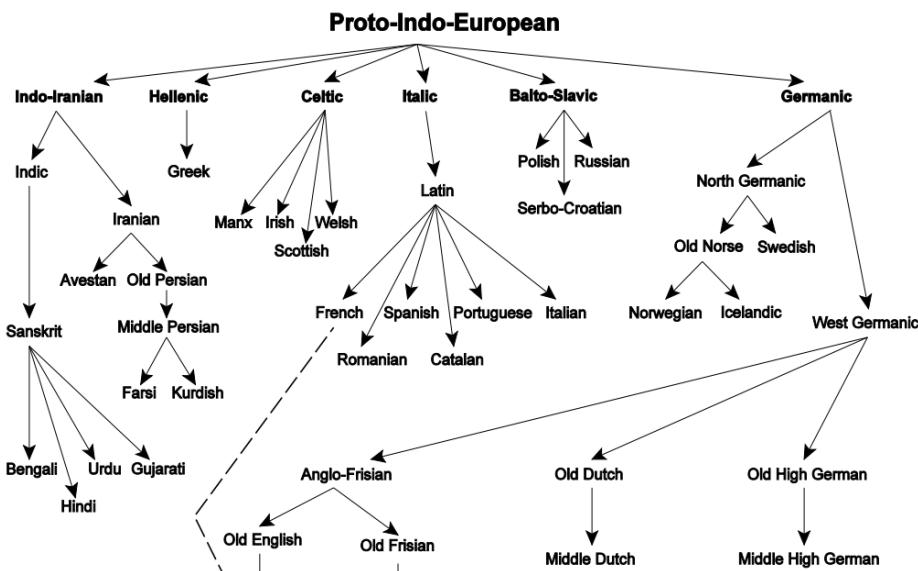
Secondly, with the advent of new approach of lexical item to vocabulary, the focus has been on a long range of vocabulary items from single words to multi-words items or phraseology in the words of Moon (2006). The various forms of Urdu words are کتاب (single word lexical item), بت پرستی (compound word), بولنا بند کرنا (idiom), دور کے ڈھول سہانے (proverb), ملکی خزانہ (collocation), etc. The handling of this type of complex vocabulary has never been an easy task.

In foreign language context, there has been felt a great need to improve this dismal situation, because learning a long range vocabulary of a foreign language was putting pressure on the overall comprehension of the learners. With the emergence of innovations in foreign language teaching, there have occurred many motivating techniques and there is a considerable shift to student centered approach and autonomy rather than following the orthodox approach of language teacher as the policeman of the classroom. The use of computers in language teaching in the form of great corpora has also improved the situation of foreign language learning.

According to Hughes and McCarthy (2001), the vocabulary of a language covers all aspects of social life. Qain (2002) describes vocabulary of a language in terms of depth, size, learner's autonomy and lexical organization. Presence of large multi-purpose corpora has made it possible to implement Qain (2002) learner friendly language theory with encouraging results. In this regard, retrieved frequency list with the help of large corpora has proved very useful in foreign language context.

#### Statement of the Problem:

Urdu is one of the top five most spoken languages of the world. Its ever increasing importance is due to socio-economic and geo-political position of the region where it is spoken, especially, the emergence of China-Pakistan Economic Corridor has increased the importance of Urdu language many times. In these conditions, it has become the need of the hour to work out for a specific vocabulary of Urdu language for Chinese business in Pakistan as the major portion of China-Pakistan Economic Corridor is being carried through in the Pakistani territories where Urdu is spoken as an official language. Consequently, a vocabulary list of Urdu language that is very frequent, can help the Chinese businessmen in Pakistan to communicate successfully over any speech event. As, Urdu is a pidgin language, its vocabulary is very diverse with lot of loan morphology.

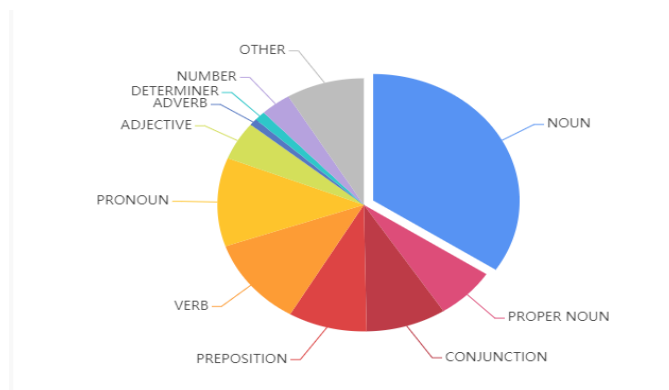


**Diagram 1. Urdu Geo-political origin**

As vocabulary plays a significant role in overall comprehension of the language, it is the need of the hour to work for the focused and core vocabulary of the Urdu language for Chinese so that that comprehension of the language may be improved in less time with maximum autonomy of the Urdu learners of the foreign language.

**2. Research Methodology**

Corpus linguistics has revolutionized the art and craft of foreign language teaching. Corpus is a collection of spoken or written text that is presented in electronic form. The corpus data is collected from real life situation to make the corpus more productive and helpful (Mukoroli, 2011). There are many corpus tools to work with the collection of electronic text that help in determining new emerging meanings, spellings, word sketch, n-grams, concordance, key words, frequency lists, etc. The name of the Urdu corpus for the undertaken research is urTenTen that has been compiled from internet data. The corpus belongs to TenTen corpus family that is corpora of the web with more than ten billion words. The corpus has been tagged according to Unified Parts of Speech (POS) Standard in Indian Languages. In order to process data, “sketch Engine” has been used. List of frequent words for the Urdu learners has been retrieved from urTenTen corpus with the help of sketch engine. The following diagram shows the weightage of the collection of corpus from various domains.



**Diagram.2 Courtesy to urTenTen corpus of the Urdu Web.**

**2.1 Place of Corpus in L2 Learning**

There have been brought many innovations with the help of corpus in nearly all field of linguistics. Especially, applied linguistics is the most influenced discipline in this regard. Linguistics in Corpus is a

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

Johns (1986) identifies the significance of concordance in L2 learning. He states that the issue of L2 vocabulary can be handled in a successful manner if the learners of the foreign language are exposed with the concordance of the lexical items. McCarthy (2001) establishes the significance of past perfect verb form in spoken and written discourse with the help of corpus. Corpus refers to the composition of the natural words, containing everything from a few sentences in huge bulk, written or oral for language learning. Recently, the word "Corpus" has been reserved for a collection of letters (or parts of the manuscript) that are stored and stored electronically. Because computer can pack and process many files information, electronic banking is generally larger than the academic literature before used in the study of words. The process is dead but the odds can play a part in the text collection, and it is designed for some purpose. Specific purpose of the decision model choice of texts, and the purpose is rather to store the books themselves because they are internal force. This separates the corpse from a library or electronic archive. This storage is kept there in a way that it can be learned in a linear way and effectively.

Cacoullos and Walker (2009) examines the multiple use of "will" and "going to" with the help of corpus analysis and conclude that their use is not determined by certainty or intention. The application of corpus linguistics in teaching can be divided into two aspects: the direct one: taking the relevant knowledge of corpus linguistics: means of developing a linguistic corpus, and applications of linguistic corpus as the teaching materials; and the indirect one: based on both corpus and computer technology, including compiling corpus-based dictionary, editing grammatical reference, textbook, developing multimedia courseware, language learning software, or evaluating or testing tools. Vocabulary is one of the three basic elements of language and is taken as the backbone of the whole language system as pointed out by Sinclair (1992) "If the language regarded as the bone of the tongue, the vocabulary offers vital organs, flesh and body. He further argued that without grammar, many can happen to express in languages, but nothing can be expressed without vocabulary.

Reppen (2009) asserts that corpus tools like MonoConc and Wordsmith can be motivating for the foreign language learners and teachers. He states that foreign language learners and teachers should develop their own corpora of multiple registers for better handling of vocabulary issues. Corpus linguistics has not same sense of potential as semantics, syntax, and pragmatics. Its real significance is in its empiricism as it takes into account the real patterns of language. Its analysis is objective and verifiability. It allows the collection of different genres and registers which makes it possible to show the wide repertoire of language. The corpus analysis is democratic in its approach as it provides equal chances to non-native speakers as to native speakers. Lastly, focused results can be obtained for pedagogical purposes.

According to Reppen (2009) corpus tools like MonoConc could help the teachers and the learners of the foreign language to develop their own material on vocabulary to determine the various shades of meanings and use. He also suggested using the reference corpora and multiple registers for the better handling of vocabulary issues.

The role of corpus in handling the challenging issues of foreign language vocabulary is an established phenomenon and it has changed the entire scenario across many branches of linguistics. Apart from this, it has led towards the autonomy of the L2 learners that was being hampered in the traditional class rooms of L2 learning (Hoey, 2000). Finally, corpus has helped in determining core vocabulary of any language that

was not possible in the past and the L2 learners had to cram long list of vocabulary items. As a result, the undertaken research has adopted a corpus perspective to retrieve a core list of frequent words for the Urdu learners of a foreign language. The core list of Urdu language may also be useful for Urdu lexicographers, lexicologists, grammarians, etc..

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

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

## **2.2 Core vocabulary List of the Urdu Language**

Urdu belongs to Indo-Aryan group within Indo-European family of languages. Its current place of use is South Asia. It is official language of Pakistan and one of the official languages of India. Urdu language is spoken by more than one hundred and twenty million people. Urdu is ranked as one of the top three languages of the world along with Chinese and English. Urdu has rich vocabulary and has the impact of many regional languages. Its vocabulary is also very diverse and not easy to handle for the foreign learners. With the emergence of corpus and its tools, there have occurred innovations in the field of foreign language learning (Granger, 1998). It has taken place the orthodox techniques being used in the traditional foreign language class rooms. Structurally, Urdu is a complex language. The reason behind the fact is that Urdu is a pidgin language and has the impact of many regional languages of the region. Consequently, learning Urdu as a foreign language is a challenging task. Core vocabulary, retrieved through corpus data, can be used to face the challenges of Urdu as a foreign language. The use of corpus and tools in the Urdu language are comparatively a new phenomenon but they have strong potential to motivate foreign learners of Urdu language. The retrieved Urdu list of frequent words can be reshaped according to the needs and levels of the learners. It will counter the traditional laborious wordlists of Urdu language that are not actively used in the real written and spoken discourse. The large number of Urdu files refers to the bulk of the corpus, resulting in a focused and core vocabulary list.

## **2.3 Word List**

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

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

Following is the list of frequent words of the Urdu language. The retrieved list is quite extensive but only the top two hundred frequent words have been indexed here. The list can be used as a point of departure and the list can be extended according to the level and the needs of learners of Urdu as a foreign language.

5/04/2019

Sketch Engine

word

(161,249 items | 52,181,119 total frequency)

Word	↓ Frequency	Per Million	Word	↓ Frequency	Per Million	Word	↓ Frequency	Per Million	Word	↓ Frequency	Per Million
1 کے	1,738,354	32,633	51 وقت	93,392	1,753	101 نام	50,181	942	151 اے	35,377	
2 کی	1,309,420	24,581	52 بات	89,983	1,689	102 امی	49,836	936	152 ہ	35,282	
3 میں	1,290,774	24,231	53 ہونے	89,638	1,683	103 بن	49,267	925	153 ک	35,267	
4 اور	1,114,561	20,923	54 اگر	88,894	1,669	104 اسلام	49,155	923	154 پاس	35,143	
5 سے	969,166	18,194	55 کرتے	88,233	1,656	105 مگر	49,105	922	155 ہوتی	34,980	
6 ہے	815,893	15,316	56 جب	87,423	1,641	106 حضرت	48,613	913	156 والی	34,731	
7 کا	761,732	14,300	57 والے	84,686	1,590	107 طور	47,618	894	157 افراد	34,507	
8 کو	671,929	12,614	58 لیکن	80,719	1,515	108 تمام	47,337	889	158 صلی	34,118	
9 اس	658,961	12,370	59 رہے	80,687	1,515	109 خود	46,270	869	159 اردو	34,002	
10 کہ	602,820	11,316	60 طرح	80,119	1,504	110 جاتا	45,221	849	160 عمل	33,879	
11 نے	600,351	11,270	61 کچھ	78,326	1,470	111 لے	44,966	844	161 یہاں	33,704	
12 پر	492,762	9,250	62 دیا	78,288	1,470	112 تھا۔	44,958	844	162 بی	33,484	
13 بھی	359,025	6,740	63 سب	75,902	1,425	113 دنیا	44,490	835	163 جانے	33,304	
14 یہ	331,925	6,231	64 بہت	72,443	1,360	114 گے	44,131	828	164 دور	33,299	
15 تو	331,622	6,225	65 تھے	72,265	1,357	115 سی	44,104	828	165 تھے۔	33,017	
16 کر	326,898	6,137	66 پھر	71,254	1,338	116 کیلئے	43,656	820	166 غیر	32,924	
17 ایک	313,159	5,879	67 اب	71,005	1,333	117 محمد	43,209	811	167 ایسے	32,772	
18 ہیں	311,926	5,856	68 جائے	70,088	1,316	118 مجھے	43,066	808	168 شخص	32,555	
19 نہیں	299,022	5,613	69 ہونے	69,072	1,297	119 جن	42,935	806	169 دل	32,356	
20 ان	289,344	5,432	70 ہوں	68,215	1,281	120 پیش	42,421	796	170 دوسرے	32,068	
21 ہے۔	276,073	5,183	71 پر	65,303	1,226	121 خلاف	42,382	796	171 style	31,985	
22 وہ	270,370	5,076	72 ہے	63,505	1,192	122 نظر	42,117	791	172 استعمال	31,620	
23 کیا	247,339	4,643	73 دو	62,738	1,178	123 انہیں	42,004	789	173 زندگی	31,397	
24 جو	229,018	4,299	74 ہوا	62,613	1,175	124 صاحب	41,990	788	174 عوام	30,933	
25 جو	195,154	3,664	75 اسے	60,498	1,136	125 آج	41,561	780	175 گھر	30,662	
26 و	185,307	3,479	76 رہا	59,906	1,125	126 حاصل	41,155	773	176 موجود	30,548	
27 نہ	176,937	3,322	77 ہے	59,711	1,121	127 وجہ	40,844	767	177 ایسا	30,502	
28 ہی	162,911	3,058	78 دے	59,598	1,119	128 رہی	40,481	760	178 امریکہ	30,397	
29 اللہ	162,267	3,046	79 حکومت	59,375	1,115	129 سال	40,438	759	179 جہاں	29,938	
30 آپ	157,868	2,964	80 ملک	59,248	1,112	130 ہونی	40,376	758	180 the	29,903	
31 gt	143,220	2,689	81 گا	59,016	1,108	131 تم	40,123	753	181 کرے	29,788	
32 اپنے	142,776	2,680	82 علیہ	58,339	1,095	132 وسلم	39,921	749	182 اپنا	29,763	
33 lt	141,960	2,665	83 میں	58,133	1,091	133 دن	39,906	749	183 امریکی	29,584	
34 کرنے	141,580	2,658	84 طرف	57,102	1,072	134 ن	39,811	747	184 جاری	29,555	
35 ہیں۔	131,932	2,477	85 انہوں	56,526	1,061	135 صدر	39,626	744	185 امام	29,549	
36 جس	127,974	2,402	86 کرنا	55,557	1,043	136 مطابق	39,583	743	186 تین	29,526	
37 تھا	123,870	2,325	87 دی	55,487	1,042	137 ا	39,049	733	187 ایم	29,511	

D File | C:/Users/Core%20M/Desktop/urdu%20wordlist%201.pdf



42	ساتھ	107,386	2,016	92	لوگوں	53,481	1,004	142	کی	37,378	702	192	سکتا	28,920	543
43	بعد	104,701	1,966	93	کام	53,358	1,002	143	باعر	37,373	702	193	ہیں	28,843	541
44	تک	104,299	1,958	94	پہلے	52,714	990	144	شعور	36,956	694	194	بیان	28,403	533
45	یا	104,058	1,953	95	گئی	52,474	985	145	خاد	36,855	692	195	جبکہ	28,021	526
46	کسی	102,231	1,919	96	ہوتا	52,221	980	146	کم	36,597	687	196	میں	27,985	525
47	لیے	101,572	1,907	97	زیادہ	52,040	977	147	بلکہ	36,446	684	197	ودنوں	27,656	519
48	گیا	98,440	1,848	98	گئے	51,568	968	148	علی	36,090	678	198	جباو	27,471	516
49	کجا	97,189	1,824	99	جا	50,859	955	149	span	35,749	671	199	سامنے	27,453	515
50	اپنی	95,738	1,797	100	کے	50,386	946	150	کرتا	35,494	666	200	رسلو	27,407	514

Frequency word lists of the languages continuously evolve and grow with the passage of time (Henriksen, 1999). Similarly, word list of Urdu language is dynamic and can be upgraded with the ever changing needs of the L2 learners and time. The rationale behind using large corpus to retrieve these lists is that the words in the list should reflect all aspects of life from social to academic. As the proposed word list has been driven from a large corpus of Urdu language, it is supposed to be productive and potential for the L2 learners of the Urdu language and will encompass all domains of Urdu culture.

### 3. Analysis and Discussion

Descriptive research of a language is useful from many aspects. Apart from helping the foreign language learners (Gleason, 1961), it helps the translators to translate books for the foreign learners with lexical sensitivity, for the grammarians who intend to publish books on lexis and in language teaching for native speakers as well as for the foreign learners of the language.

#### 3.1 Structurally, Urdu is a different language from English

The significance of such descriptive studies becomes all the more important when the non-native learners of the Urdu language constantly make comparisons of the target language with their mother tongue. Realistically, such convergence of the target language and the mother tongue is not potential as “sunny day” and “دھوپ والا دن” have difference of meaning when used in their respective cultures. Conversely, these attempts of convergence by English learners may lead to unsuccessful handling of the Urdu language.

As language is culture, expressed by language and no expression of language can occur without suitable word combinations. The above description of lexicon of the Urdu language is potential as it is replete with cultural information. As the corpus data has been collected from various domains of Urdu culture, the vocabulary list also reflects the word associations with full cultural meanings as the word روشنی (light) is occurring with word choices like القرآن الکریم (The Holy Quran) حدیث (Hadees), etc. connoting to some extent the Muslims background of the Urdu lexicon.

The above description of Urdu collocations shows that some words accept more word choice than the rest. For example a few Urdu adjectives take more Urdu nouns in association in comparison with Urdu nouns that accept Urdu verbs. For example Urdu noun “صدر” (president) take more nouns in association in comparison with Urdu verbs.

Predictably, Urdu vocabulary described in the above section, can be classified the most frequent as they have been retrieved from large corpus data. Without considerable awareness of frequent wordlist, as supported by Lewis (2004), it is difficult to handle overall language skills of Urdu. These are the vocabulary items that shape the linguistic expressions of non-native learners, natural and fluent even though they lack grammatical strength sometimes. However, it is an established fact that Urdu as L1 and as L2, lexical development differ significantly. The first difference in the case of Urdu as L2 is the poverty of input of the word combinations in terms of quality and quantity that make the bulk of Urdu

lexicon. The L2 learners of Urdu have to face lack of contextualized data that can only be presented in the form of lexical collocations. This situation puts foreign learners of Urdu language in an extremely difficult position to create syntactic, semantic and morphological harmony among word associations of Urdu language and they are at a loss to integrate words into acceptable combinations. Learners of Urdu as a foreign language usually focus on the single words items and disregard their association and context. Consequently, they often refer to their mother tongue to produce word combinations in the target language that results in unacceptable lexical choices.

Summing up the corpus data, it can be safely claimed that Urdu is a very rich language. Its lexicon structure is very diverse. Urdu language offers many word combinations of lexical collocations that are constituted with the help of single words. The corpus data of Urdu lexical lexicon can be very useful in the description of Urdu language at the level of lexis.

#### **4. Concluding Remarks**

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

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

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

Vocabulary handling in any L2 program is a key indicator in overall comprehension of the language (Bauer, 1998). Computers and soft wares have contributed in improving this situation. The retrieved frequency list of Urdu language is supposed to be very helpful for the Chinese business men who are working in various capacities in Pakistan. The use of corpus and soft wares will also lead towards of the learners in the foreign language class rooms. The undertaken research is an innovation in the art of foreign language learning and fills a critical crevice in vocabulary building of Urdu language for the Chinese business community in Pakistan.

#### **References**

- Bauer, L.(1998). Vocabulary. London and New York: Routledge.
- Biemiller, A. (2004). Vocabulary instruction (pp. 159-176). New York: Guilford Press.
- Cacoullos, R. T., & Walker, J. A. (2009). Grammatical variation and collocations in discourse Language
- Gleason, H.S. Jr. (1961). An Introduction to Descriptive Linguistics. New Delhi: Oxford and IBH Publishing Company
- Granger, S. (1998). The computer learner corpus: London: Longman.



- Henriksen, B. (1999). Three dimensions of vocabulary development, OUP.
- Hoey, M. (2000) 'A world beyond collocation London
- McCarthy, M. J. (2001). Issues in Applied Linguistics. Cambridge: Cambridge University Press
- Mukoroli, J. (2011). Effective Vocabulary Teaching Strategies: Longman
- Reppen, R. (2009). English language teaching and corpus linguistics: London: Continuum Press.
- Moon, R. (2006). Introducing Metaphor, Routledge
- Neuman, S. B., & Dwyer, J. (2009): Vocabulary instruction in pre-K. OUP
- Qian, D.D. (2002). Investigating the Relationship between Vocabulary Knowledge and Academic Reading Performance: Continuum press.
- Schmitt, N. (2000). Vocabulary Learning Strategies. Cambridge University press

