A Corpus-Based Study for the Identification of ESP Learning Needs of Business Students in Pakistan

Haroon Shafique[1], Behzad Anwar[2]

[1]University of Gujrat, Gujrat
[1] Haroon.linguistics@gmail.com
[2]University of Gujrat, Gujrat
[2] Behzad.anwar@uog.edu.pk

Abstract

General English language learners can generally communicate well, but their vocabulary in specific fields is normally limited. This study aims to demonstrate the use of a specific language corpus for identifying the learning needs of English for Specific Purposes (ESP) students in Pakistan. For this purpose, two hundred fundamental textbooks from the field of business studies are selected and processed by using appropriate software, and a list of technical words is generated. The data extracted from the corpus helps to identify the learning needs of business students in terms of technical vocabulary. The study shows that the corpus is an effective tool to identify the ESP needs of learners. The study suggests the dire need to adopt corpus-based teaching and learning methodologies to cater to the ESP learning needs of the learners. Moreover, this corpus can serve as a basis for designing a lexical syllabus for ESP learners in Pakistan.

Keywords: Corpus Analysis, English for Specific Purposes, technical terms of Business, learning needs, Corpus as a tool

A Corpus-Based Study for the Identification of ESP Learning Needs of Business Students in Pakistan

English is the standard language of the world and the most prestigious language in Pakistan, among other languages (Akram & Mahmood, 2007). It is the language of science, education, business, law, government, and technology. The students of Pakistan face numerous issues while dealing with English during studentship and afterward. In Pakistan, business administration is one of the most common fields, as it is offered in every university. The students of the Business Administration Program in Pakistan are offered at least three modules of English during their undergraduate program. However, it is believed that the offered modules do not fulfill the ESP needs of the students, and students are weak in the technical vocabulary of the field. Consequently, the students face various problems. This research aims to provide corpus-based assistance to the researcher to identify the ESP needs of business students in Pakistan.

The current research offers corpus-based data for the understanding of corpus as a tool and its pedagogical implications in an ESP setting. The undertaken research provides suitable data and the right type of subject specialization for the English language learners of Pakistan. However, the scope of this research is limited only to the field of Business Administration.

A corpus is a collection of written texts that may belong to single or multiple authors in a particular language variety or language in use. Linguistically, the corpus can be defined as a large and structured body of text. Moreover, the electronically stored data of language in use is regarded as the corpus (Hunston, 2006). To obtain a specific purpose or to know the answers to certain questions, a corpus of written or verbal discourse is usually formed (Leech, 1992). A corpus is used to testify to the theories and hypotheses related to language that already exist and even to develop new language theories and hypotheses. A corpus is based on the textual data of a single language or it may contain the text data in more than one language, which is called a multilingual corpus (Hunston, 2006).

Corpus emerged in the 1960s and right after its emergence, the academic use of corpus in terms of the teaching and learning process started. However, most of the academic applications were not in direct relation with corpus-based studies as they only dealt with grammar, and dictionary wordlists (Kennedy, 2014). A corpus is not merely compiling language data, rather it represents a language, a language variety, or different linguistic features (Biber et al., 1998). It is noteworthy that a corpus is not merely the collection of language based on isolated utterances and randomly selected words or sentences occurring naturally. However, it is the accumulation of linguistic data occurring in such circumstances where the language users acknowledge to be doing something to determine how the system works. Hence, the text used as a corpus should be purposeful.

According to Sinclair (2004), the corpus is a collection of linguistic data collected and stored in electronic form based on external criteria to represent particular features of a language or a variety of it. The collected piece of language or variety serves as a corpus for research. However, for such research, the amount of linguistic data is of grave importance as the corpus should be large enough to be considered for representing a language or its variety based on some external criteria for any research.

According to Biber (2009), there are two main approaches in corpus studies. The first approach is corpus-based while the second one is corpus-driven. Primarily, these two are the main

methods for the analysis of language variation. In corpus-based analysis, the structures and forms of language are analyzed with the purpose to investigate the systematic patterns of language variation. Then pre-defined linguistic features are examined. While corpus-driven analysis is inductive as it deals with the existence of words and concepts like clause and phrase, has no concern with it. In a language, the variation patterns of the complete text of spoken registers, including classroom lectures, debates, friends' face-to-face conversations, etc., or written registers, e.g., academic discourse, newspaper articles, text messages, etc., are analyzed (Biber, 2009).

Corpora and ESP

The transition period from second world war to scientific and technological world is considered to be the origin of ESP as a field (Rehman, 2015). There have been multiple factors i.e., globalization of industry, business and education, oil crisis, second world war and change in teaching methodology that led the development of ESP as a field however, it is quite difficult to define ESP because of its application in multiple areas of studies in varied situations. Strevens (1988) attempts to define ESP with help of its varied characteristics. To him, the goal for designing an ESP syllabus is to cater the needs of learners based on appropriate lexis, semantics, syntax and discourse of a particular language. Alfehaid (2011) claims that the role of ESP is not only to provide context-based knowledge to the learners but also the development of language skills that could help them to excel in their field. Corpora play a vital role in the field of English for Academic Purposes/English for Special Academic Purposes (EAP/ESAP). For analysts, the advent and the use of corpora have made it possible and easy to examine language patterns, most frequent words, and frequent phrases in various domains (Hussain, 2019).

Primarily, corpus has two roles in language learning i.e., direct and indirect (Leech, 2014). Both the roles are associated with ESP. The direct role corpus serves in the field of ESP is pedagogy. So, corpus serves as an approach and pedagogic tool within the classroom for language learning. In this way, corpus helps as a tool for teaching as well as learning any language. The students or teachers need to have hands on expertise in the use of corpus for the direct use of corpus as a tool. Contrarily, the indirect use of corpus is the provision of research-based language learning material drawn out of corpus. The huge data contains repeated linguistic patterns which provide evidences of academic registers for language learning in a classroom. Such evidence-based registers of language offer the Students of ESP the needful data for learning second language. Granger (2009) believes that learner corpora play a vital role for designing corpus-informed materials for the development an ESP course. Such investigation of corpus helps the language teachers to identify the lacking of the learners which determines the linguistic patterns which require more focus on the part of teachers and students. Nesi (2015) claims that even a small corpus can offer helpful information for EAP/ESP but it needs to be contextually rich. Such context-based corpus can be mined on internet within specific domains of fields or topics associated with particular linguistic features.

The benefits of corpus-based ELT cannot be neglected. With the help of corpus, Idiomatic expressions with contextual meanings can be explored (Shehzad, 2005). A non-native learner of English can learn how a native speaker of English use text structure, phrasal verbs, idioms and vocabulary with respect to different situations (Liu & Jiang, 2009). The investigations of corpus informed studies can help in attaining native like proficiency, contextual knowledge and linguistics efficiency (Gardner & Devies, 2007). Unlike the available teaching materials which exhibits few

examples of the usage of language, but corpus provides various aspects of language and facilitates them with the best material for learning (Kessler, 2016).

Research Questions

1. Whether or not English courses provide a suitable subject-specific vocabulary to the students?
2. What kind of technical vocabulary needs to be incorporated into business English?

Literature Review

The role of the corpus in the field of ESP has proved to be effective and convenient. Corpus is considered to be a significant tool for material designing in ESP. A corpus offers authentic and credible data for teaching and learning a language in an ESP setting. So, the role and impact of the corpus are not neglectable in ESP. Despite the verdict of researchers on the implementation of corpus-based learning in the field of ESP, it is challenging for teachers to incorporate corpora into the ESP classroom (Leech, 1997). Timmis (2015) believes that corpus has three, major objectives regarding ESP. The first is to carry out corpus-based research that assists the process of teaching and learning in an ESP classroom. Second, is the use of data-driven learning, which deals with the presentation of accurate data to the learners within a classroom to dig out the required language patterns through the corpus. Third, is the use of corpus as a resource for the teachers for preparation materials for teaching in the classroom.

Three main uses of corpora in ESP have been identified by Timmis (2015): first, in conducting research that underpins effective language teaching and learning. Second, corpora can be used as a way of presenting authentic language data to learners in an approach referred to as "data-driven learning (DDL)." Third, corpora can be a resource for teachers to use in preparing classroom teaching activities and materials. The researchers in the field of ESP believe that the direct use of corpora makes the learners, language researchers. In such a context where learners collect the data, compile the specialized corpus and generate the results and understand the structure and patterns of the language data, they become researchers. Such practices promote a learner centered approach in an ESP setting.The corpus is also helpful in providing data based on frequency. The extracted data based on frequency can be used to teach in an ESP classroom. Mindt (1996) compared the corpus of native English speakers and the textbooks taught in Germany to the ESP learners' knowledge regarding the use of future time expressions, modal verbs, and conditional clauses. The research tried to gauge the difference in frequencies in the use of said patterns in both corpora. The results revealed a difference in the frequencies of modal verbs and conditional clauses in the textbook and native English corpora. The researcher claims that a syllabus, not designed through a corpus does not match the actual use of native English, written or spoken, which leads to the problem. He is of the view that an ESP syllabus must be designed based on empirical evidence. In this regard, the corpus provides accurate evidence of data which should be the priority for teaching English in the ESP setting. The disparity between the use of linguistic features between native speakers and German textbooks is the same between business corpora and course outlines for business students in Pakistan

Fuentes' (2002) research on the usage of corpus for investigating speaking skills in the Spanish ESP setting deals with the productivity of corpora in the classroom. For the study, two groups were made i.e., experimental and controlled. The students of the experimental group were offered training in the use of one million-word specialized corpora spread over two weeks to investigate compounds, collocations, and clusters within Business English. Both groups were expected to give oral presentations on any of the topics from the field of business. The data taken from the presentations showed that the experimental group made better use of collocations and clusters with the help of corpora while the controlled group could not perform well in terms of the use of business technical vocabulary. The study suggests that a corpus is a useful tool for ESP learners for learning technical discourse and grammar, which is also the objective of the research in hand: to identify the exact learning needs through corpora.

Another study conducted by Fuentes (2004) deals with the use of corpora for individual reporting and group discussion by tourism students. A specialized corpus consisting of twenty-five thousand words was built from business reports and reviews on the internet. The students of tourism majors belonging to the third semester were divided into experimental and controlled groups to gauge the efficiency of the corpus. The students of both groups had to deliver an individual oral report and have a group discussion on an entirely unknown topic. Fuentes (2004) reported a significantly effective use of the specialized terms in oral reporting as well as in group discussions by the experimental group. The experimental group was described as being more confident, and spontaneous while performing the tasks. Moreover, the discussion and reporting of the experimental group had more pace and delivered longer reporting. It was because of the corpora provided to the experimental group. In this way, the exposure of corpora to the curriculum designers before designing a course or students within a classroom can help in specializing the technical vocabulary in the relevant field.

The students of tourism also participated in a study carried out by Fuentes (2007). The students had to find the register of six tourist advertisements. The students were divided into two groups for the exploration of technical vocabulary and subject-specific lexical items. The controlled group was offered textbooks while the experimental group had access to a corpus based on 60 advertisements for tourism. Both groups were given two hours to find suitable text types and register for the tourism industry. Both groups were offered reading comprehension passages. The experimental group performed better compared to the control group. The students from the experimental group succeeded in answering all five questions, while the students taking help from the textbooks were unable to give satisfactory answers. The research proposes the use of corpora for the identification of suitable technical vocabulary to perform better in a particular field, which is also in line with the research at hand regarding the technical vocabulary of business students.

The study of Ng et al. (2013) is based on the development of an engineering corpus for pedagogic purposes. The textbooks of engineering technology in Malaysia are taken for the construction of corpus. A wordlist of engineering technology was created to see the significance of specialized lexical data in the corpus. The frequency of technical vocabulary was checked in Concordancer to regard it as a specialized corpus. All the technical words found in the corpus were not available in the academic word list which showed the specificity of the data. The research conducted by Ng et al (2013) is similar to the current research, as the engineering corpora in the previous study and the business corpora in the current study have subject specificity, while AWL and business course outlines do not carry specialized lexical data.

The study conducted by Lizhu (2014), confirmed that corpus study is helpful for the better understandability of aviation English. For this purpose, a small corpus was designed to compare verbal nominalization in general English and aviation English. The corpus provided evidence that verbal nominalization is a rich feature of aviation English. The researcher also suggested creating an aviation corpus for teaching and learning aviation English. Chen and Huang (2017), suggested that with such a corpus, the students of aviation can learn field-specific linguistic data, grammar, and vocabulary. Aviation English provides professional features that cannot be provided by general English. Also, it is highly important to develop an aviation corpus and provide hands-on expertise for the autonomous learning of ESP learners. The current study also deals with the development of business corpora as the general English or the curriculum does not contain subject-specific grammar, vocabulary, and data.

Kurtul and Kirkgoz (2020) conducted a study to design a lexical syllabus for electrical and electronic engineering at Nigde University. The corpus was compiled with help of textbooks and online sources. The research aimed to provide accurate and authentic data for learning subject-specific lexical data and linguistic structures. A total of ninety-four students were selected and divided into two groups for the study. A pre-test was conducted to analyze the present situation analysis and the current level of English of the learners. The controlled group was instructed using the conventional method, while the experimental group was instructed through a corpus-lexical syllabus for a tenure of twelve weeks. The post-test exhibited a significant difference in the learning of both groups. The researchers suggested developing a corpus-based lexical syllabus for the ESP students for better understanding of technical vocabulary. The literature review is about situations your work among existing works. But this lit review seems more definitional. It does not relate the current study with existing studies..

The previous research by Mindt (1996), Fuentes (2007), Lizhu (2014), Kurtul and Kirkgoz (2020), Chen and Huang (2017) on corpus-based learning needs, on the one hand, deals with controlled and experimental groups based on comparisons between pre-test and post-test of experimental and controlled groups regarding corpus-extracted technical vocabulary, and, on the other hand, deals with the comparison of the specialized corpus with AWL. However, there has been little research on the specialized vocabulary needs of ESP learners in Pakistan. The research in hand deals with the identification of vocabulary learning needs of ESP students in Pakistan. Moreover, the previous studies were based on engineering, tourism, food technology, and aviation corpora, whereas the current study deals with the business corpus of ESP learners of Pakistan.

Methodology

This research is carried out by employing the mixed method of research that advances the systematic integration of quantitative and qualitative data within a single investigation (Cresswell & Cresswell, 2005). The basic premise of this methodology is that such integration permits a more complete and synergistic utilization of data than separate quantitative and qualitative data collection and analysis. Further, in this research, a sequential explanatory method of research is utilized. The results drawn by the quantitative part of the research assist the qualitative portion which serves to be the purpose of sequential explanatory research. Moreover, the explanatory sequential approach is a sequential approach and is used when the researcher is interested in following up the quantitative results with qualitative data. In this way, qualitative data is used in the subsequent interpretation and clarification of the results drawn from quantitative data analysis. Sequential explanatory research design is highly useful for this research as the data is quantified

with help of a corpus and later on the quantified data is authenticated with qualitative data. A questionnaire is a research instrument that consists of a set of questions to collect data from the respondents. In this research, the respondents are given open-ended questionnaires for the collection of data.

The data is quantified, and a list of one hundred technical terms has been generated with their frequency in the corpus. The criterion to choose the technical terms out of the corpus is frequency. The terms with higher frequency are placed at the top of the list. The data is quantified first and then analyzed qualitatively to draw results. For this purpose, all the technical terms have been analyzed through expert opinion. First, all the technical terms are checked with the help of two doctoral experts from the field if the drawn technical items serve as technical terms of the business field. Expert opinion is a research instrument used in multiple areas of scientific research which guarantees its methodological validity. So, all the technical terms are analyzed qualitatively by experts in the field. Out of the list, twenty-five technical terms are chosen by using systematic random sampling. For the further qualitative analysis of the technical vocabulary, a few technical terms are analyzed by using the nth number i.e., 20. All the chosen technical terms are analyzed and elucidated for qualitative analysis.

Since three modules of English are taught in the BS program, two groups of students are made to check the understandability of technical terms. Twenty-five students are chosen from 3rd semester and the rest of the twenty-five students are chosen from the 5th semester. For the research, the students of both semesters with a CPGA score of 3.2 or more were chosen as respondents.

For the analysis, all fifty students are given open-ended questionnaires. The participants of the research are briefed about the questionnaire and formal consent has been taken from the respondents by offering them consent forms. The respondents have been made assured of anonymity in this research. Moreover, all the research ethics have been taken into account for conducting the research.

Theoretical Framework

In this research, needs analysis is incorporated as a theoretical framework. The term "need" in needs analysis refers to necessities, wants, and demands (Martins, 2017). Needs analysis is a synonymous term of needs assessment, and both terms are prevalent in ESP and EAP. It is of grave importance to know the actual needs of the learners before designing any course. The very purpose is rightly served by needs analysis, as the learning needs of the learners are identified to enhance the effectiveness of the curriculum. It is a procedure for identifying, classifying, and validating the required priorities of the learners.

Previously, in the early stages of ESP, need analysis was conducted to assess the communication needs of the learners. Also, it was carried out in the 1960s and 1970s to evaluate the teaching techniques to have set results. However, the term "needs analysis" has different connotations presently as compared to the past. Since needs analysis aims to collect authentic information about their needs, the tasks of needs analysis have varied significantly (Otilia, 2015). Numerous factors require attention in the process of needs analysis, i.e., the learning environment, learning strategies, learning styles, the competence of the learners, behaviour towards learning, and most importantly the target situation (Basturkmen, 2014). In this research, the needs of the ESP learners are analyzed by conducting a deficiency analysis, which reveals present needs and

shortages of needed items in the target situation. This approach has been developed by West (1994) to find learning needs, requirements lack, and deficiencies.

Corpus building

The corpus for this research is compiled by using two hundred textbooks from the field of Business Administration. The data is gathered from two hundred books recommended in HEC's course outline for the BBA program. The books are downloaded and converted into Notepad files for further proceeding. The data is screened, filed, encoded, and then compiled for further analysis. All the unnecessary data is removed from the corpus, like images and figures, to present accurate data for the analysis.

The corpus collected from the textbooks comprises more than sixteen million words. The particular corpus has been chosen by the criteria suggested by Nwogu (1997) being accessible, well-reputed, and representable. First of all, the chosen data must be accessible to make it readable for the software. In light of the first criterion, the data needs to be accessible. Secondly, the data chosen for corpus collection has a good reputation as the books have been written by local and foreign authors. Moreover, the books have been recommended by HEC, which is one of the most reliable institutes in Pakistan. Lastly, the data represents the field of business administration. All the core and fundamental books of the field have been part of the corpus which proves the representativeness of the corpus. The collected data carry all the characteristics of a suitable corpus, taking into account Nwogu's (1997) criteria. Moreover, the size of the corpus is sixteen million eighty-four thousand and ninety-seven words (1684,497), which is sufficient to represent linguistic features in a particular linguistic domain.

Data Collection

The data for this research is collected in different phases. In the first phase of data collection, the corpus is built and, out of the compiled corpus, the subject-specific terms are drawn and enlisted. Now, in the second phase of data collection, the subject-specific terms are chosen through systematic random sampling, and a questionnaire is given to the undergraduate students of the concerned field to identify deficiencies. The number five is chosen through systematic random sampling of the list, for the compilation of the questionnaire. The data collected from questionnaires reveals the current situation of ESP learners regarding technical terms.

Research Tools

For corpus-based research, certain tools are required to extract the data from a huge language corpus. For this purpose, Antconc 3.2.4w (2011) is used for the analysis of corpus in this research. Anthony is a toolkit designed by Laurence Antony to draw the intended data from the corpus (Anthhony, 2005). This software hosts a comprehensive set of tools containing a concordancer, wordlist generator, keyword frequency generator, and tools for collocation, cluster, and lexical bundle analysis (Antony, 2004). The reason to use Antconc for this research is to find concordances of required words and to see the functionality of those required words in the context.

Data Analysis and Discussion

The respondents of the study are 3rd and 5th-semester students from undergraduate business administration programs who have secured a 3.2 or above CGPA. The students from both groups are chosen through convenience sampling. The purpose of this study is to testify if current

Business English courses cater to the ESP learning needs of the students by providing required technical terms. Also, what kind of technical vocabulary needs to be inculcated in their courses? The questionnaire is open-ended and contains twenty-five technical terms selected through systematic random sampling out of the list of one hundred terms with the highest frequency in the corpus. The list of the top one hundred technical vocabulary items is given below.

Table 1

List of Technical Terms in Business Corpus

Technical Terms	Freq.	Technical Terms	Freq.	Technical Terms	Freq.
business cycle	689	sloan management	256	asset approach	158
cost volume	674	income effect	244	limit order	154
cash flow	641	interest parity	240	money security	151
income identity	602	control protocol	236	financial leverage	151
financial capital	588	interest arbitrage	236	interest bearing	150
trade imbalance	564	median income	232	price discrimination	148
supply chain	561	cash equivalents	230	deferred income	148
variable cost	555	competitive strategy	227	cash forecasting	145
business tree	543	accrued interest	227	value proposition	142
output gap	539	direct wages	224	contingency approach	141
price level	528	credit worthiness	222	market structure	141
Cost volume	504	distribution strategy	216	realizable value	140
fund market	487	inelastic demand	214	fixed scale	140
cost structure	476	demand shift	208	chronological order	137
cost curve	452	sales force	206	reconciliation statement	134
price floor	450	trade deficit	201	safety stock	131
price index	450	fixed asset	201	marketing stimuli	130
equilibrium price	445	internal business	201	customer retention	130
income effect	433	progressive tax	198	secondary market	127
Market segmentation	430	business ethics	196	product positioning	127
differential cost	421	vertical marketing	195	financial distress	125
sales reps	404	price sensitivity	194	pocketing order	124
cost drivers	401	stock valuation	194	price sensitivity	120
market clearing	395	strike price	183	asset liability	117
price signal	382	expenditure output	182	price war	114
price leadership	380	marketing pitfall	180	compound interest	111
cost center	374	direct marketing	173	middle income	111
open market	366	Macroeconomy	168	cash dividends	110
forward market	354	build customer	168	face value	108
Logistics	352	preferred stock	165	price ceiling	108
spot market	352	regressive tax	162	historical cost	103
chain management	309	tax base	162	book value	101
cash disbursements	290	excess supply	159	market penetration	101
opening income	275				

All the technical terms are quantified and then analyzed qualitatively to see if they are technical terms from the field of business administration. In the process of the quantification of technical terms, the terms with the highest frequency in the corpus are placed on top, and a list is generated in descending order from higher frequency to lower frequency. For instance, the term "business cycle" is used six hundred and eighty-nine (689) times in the corpus with the highest frequency. On the other hand, the technical term with the lowest frequency is market penetration with one hundred and one (101) frequency in the corpus. All the technical terms are enlisted with their occurrence in the corpus. After the quantification of the data, the list of subject-specific terms drawn through the business corpus is analyzed qualitatively. For this purpose, all the technical vocabulary has been examined by expert opinion. First, all the technical terms drawn from corpora are checked with help of two doctoral experts from the field of Business Administration, if drawn technical items serve as technical terms in the business field. In this regard, the doctoral experts have excluded only two terms that do not serve as the technical terms of the Business field, i.e., yield and actuals. Further, an nth number, i.e., 20, is chosen out of a hundred technical terms for the qualitative illustration.

The given examples are the extractions of technical terms drawn from business corpora for qualitative analysis. The examples have been chosen by using the nth number which is 20 out of one hundred most frequently occurring technical terms taken from the corpus. The first technical term is "market segmentation," meaning dividing the customers concerning age, gender, area, needs, education, and income. In the given example, market segmentation is a technical vocabulary from the field of business, which has also been verified by two doctoral experts.

- Bases for Segmenting Consumer Markets **Market segmentation** divides a market into well-defined slices.

In the second example, the term "median income" has been taken from the corpus. The term "median income" is the separation between the two groups having income above the limit or below the limit. The example suggests that the term median income serves as a technical term for business. The specific term has a frequency of 243 in a sixteen-million-word corpus.

- The poverty line is set at 60% of the **median income** of each member nation in a particular year.

The third example deals with the subject-specific item “marketing pitfall”, meaning the possible difficulties or mistakes an entrepreneur or a company could face while running a business. The frequency of the term in business corpora is 180 out of sixteen million business corpus. The subject-specific item is also approved by the doctoral experts because of the way of it is used in the corpus.

- **Marketing pitfall** brings to life possible marketing situations or dilemmas that might arise due to cultural differences or lack of knowledge.

In the fourth example, the term "fixed scale" has been used. The term has been used 140 times in the corpus, which suggests the wide use of the term in the field of business. The term suggests certain factors under which a scale is developed for the operations of a business.

- The firm is operating under a **fixed scale** of production.

The given example deals with the technical term "market penetration," which has a significant frequency in the corpus. The term "market penetration" is a scale to assess how much a particular product or service is consumed as compared to the total market of the particular product. The under-discussion term has also been analyzed by doctoral experts in business administration.

- It is immediately clear that zip codes closest to the selected store have the highest **market penetration**.

An open-ended questionnaire is given to the respondents of the study having twenty-five subject specific terms drawn through systematic random sampling. Out of one hundred subject specific terms, twenty-five technical terms are the part of questionnaire, where the respondents have to explain the technical terms and make sentence regarding their understanding of the technical terms. For the selection of the subject specific terms out of the list, an nth number i.e., 5 is chosen. After the systematic random sampling, the questionnaire contains the terms i.e., financial capital, output gap, cost curve, market segmentation, price signal, logistics, sloan management, median income, credit worthiness, trade deficit, vertical marketing, marketing pitfall, regressive tax, financial leverage, cash forecasting, fixed scale, customer retention, price sensitivity, cash dividends and market penetration out of one hundred subject specific terms. The questionnaire is distributed among two groups i.e., 3rd semester and 5th semester students. Each of the group has twenty-five students to check the understandability of the ESP learners regarding subject specific terms. The ESP learners are given 40 minutes to explain the subject specific terms. The open-ended questionnaires are assessed with help of two doctoral experts of the field.

For the quantitative analysis of questionnaire, the average of each technical term is derived and later on the accuracy percentage of the group is drawn. For instance, the first technical term, financial capital is explained by 4 learners which is 16 percent while the term sloan management is explained by none of the ESP learns with zero percentage from the ESP learners of 3rd semester. Likewise, the term trade deficit is rightly elaborated by 5 students with an average of 0.2 while the term marketing pitfall is again described by none of the students. The term logistics has the highest average among all the subject specific items explained by the group of 3rd semester. Out of twenty-five students, nine students explained the term correctly. On the other hand, cash dividends, customer retention and fixed scale are explained by 4 percent, 8 percent and 4 percent of the learners respectively. Moreover, the terms like sloan management, marketing pitfall, price sensitivity, financial leverage, market penetration are not elaborated by any of the learners. So, the average score of 3rd semester ESL learners is 2.28 which is 9.12 percent.

The same questionnaire having twenty-five subject specific terms is distributed among the ESP learners of 5th semester. The first technical term is explained by 12 learners which is 48 percent of the group. The term which is explained by most of the students is customer retention having 64 percent of the group regarding the particular subject specific item. The least explained technical terms are sloan management, management pitfall, credit worthiness and market penetration with 0 percent, 4 percent, 0 percent and 4 percent respectively. Another significant term which is elaborated by 14 learners with 52 percent is cost curve. The accumulative average of the group is 5.28 which is 21.1 percent of the group which suggests that only 21.1 percent ESP learners of 5th semester are able to explain subject specific item even after studying three modules of English. The results are in line with the results of Lizhu, (2014) and Chen and Huang (2017)

which suggest the need to develop specialized corpus for learning of professional and technical vocabulary of a particular field. Likewise, the results drawn from this research are based upon the difference between general English and Specialized technical English based on corpus which differentiates between the two as suggested in the research of Mindt (1996).

The results drawn from the questionnaire reveal that only 9.12 percent technical terms are explained by the students of 3rd semester. Whereas the students of 5th semester are able to explain 21.12 percent technical vocabulary. The questionnaire also shows that most of the technical terms are not explained by the learners which illustrates the deficiency and the need to learn technical terms on the part of learners. The results reveal the dire need to fulfill the accurate needs of ESP learners by providing them suitable ESP content assisted by corpus.

Conclusion

The research deals with the identification of ESP learning needs of undergraduate students in Pakistan with the help of corpus. Through this research, the needs of English language learners in the field of business are identified. The data reveals that the students of the 3rd and 5th semesters explained the technical vocabulary at 9.12 percent and 21.12 percent, respectively, which suggests the inability to understand the technical terms from business corpora. The results also indicate that the current business courses do not cater to the learning needs of the learners in terms of technical vocabulary. The corpus-based analysis in this research provides the needed technical vocabulary extracted from business corpora which can be highly effective during studentship as well as in target situations. Moreover, it will be helpful for the researchers to work in this area and for syllabus designers to employ a corpus-based method for finding suitable data for teaching English to ESP students in Pakistan.

References

Akram, M., & Mahmood, A. (2007). The status and teaching of English in Pakistan. *Language in India, 7*(12), 1-7.

Alfehaid, A. F. T. (2011). *Developing an ESP curriculum for students of health sciences through needs analysis and course evaluation in Saudi Arabia* [Doctoral dissertation, University of Leicester].

Basturkmen, H. (2014). *Ideas and options in English for specific purposes*. Routledge.

Biber, D., Douglas, B., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Biber, D., Leech, G. & Conrad, S. (2002*). Longman student grammar of spoken and written English.* Longman.

Boulton, A. (2016). Integrating corpus tools and techniques in ESP courses. *ASp. la revue du GERAS*, (69), 113-137.

Chen, Z., & Huang, M. (2017). Corpus Construction for ESP. *International Journal of New Developments in Engineering and Society*, *1*(3).

Creswell, J. W., & Creswell, J. D. (2005). Mixed methods research: Developments, debates, and dilemmas. *Research in organizations: Foundations and methods of inquiry*, *2*, 315-326.

Fuentes, A. C. (2002). Exploitation and assessment of a business English corpus through language learning tasks. *ICAME journal*, *26*, 5-30.

Fuentes, A. C. (2003). The use of corpora and IT in a comparative evaluation approach to oral business English. *ReCALL: the Journal of EUROCALL*, *15*(2), 189.

Fuentes, A. C. (2004). The use of corpora and IT in evaluating oral task competence for tourism English. *CALICO journal*, 5-24.

Hunston, S. (2002). Pattern grammar, language teaching, and linguistic variation. *Using corpora to explore linguistic variation*, 167-183.

Hunston, S. (2002). *Corpora in applied linguistics*. Ernst Klett Sprachen.

Hussain, G. (2019). Preposition Phrase Expressions: A Corpus-Based Study of Lexical Bundles in Pakistani Textbooks. *CORPORUM: Journal of Corpus Linguistics*, *2*(2), 69-83.

Kurtul, N., & Kirkgoz, Y. (2020). *Design and Implementation of a Lexical Syllabus: A Corpus-based Approach for Electrical and Electronics Engineering Students.* Lambert academic publishing.

Kennedy, G. (2014). *An introduction to corpus linguistics*. Routledge.

Lizhu, Z. (2014). Application of Small Corpora in ESP Teaching: A Case Study of Aeronautical English Teaching. *Journal of Beijing University of Aeronautics and Astronautics Social Sciences Edition*, *27*(2), 89-

Leech, G. (1992). Corpora and theories of linguistic performance. *Directions in corpus linguistics, 1992*, 105-122.

Leech, G. (1997). Introducing corpus annotation. *Corpus annotation: Linguistic information from computer text corpora*, 1-18.

Leech, G. (2014). Teaching and language corpora: A convergence. In *Teaching and language corpora* (pp. 1-24). Routledge.

Lewis, M. (2002). *The Lexical Approach: The state of ELT and the way forward.* Thomson Heinle.

Martins, H. F. (1997). Revisiting needs analysis: a cornerstone for business English courses. *Journal of English Language & Translation Studies*, *5*(1), 57-63.

Mindt, D. (1996). English corpus linguistics and the foreign language teaching syllabus. *Using corpora for language research*, 232-247.

Nesi, H. (2015). ESP corpus construction: a plea for a needs-driven approach. *ASp. la revue du GERAS*, (68), 7-24.

Ng, Y. J. (2013). Development of the engineering technology word list for vocational schools in Malaysia. *International Education Research*, *1*(1), 43–59.

Otilia, S. M. (2015). Needs analysis in English for specific purposes. Annals of the Constantin Brâncuşi. *University of Târgu Jiu, Economy Series, 1*(2), 54-55.

Rahman, M. (2015). English for Specific Purposes (ESP): A Holistic Review. *Universal Journal of Educational Research*, *3*(1), 24-31.

Sinclair, J. M. (1987). *An account of the COBUILD project*. Collins ELT.

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. *Vocabulary and language teaching*, *140*, 60.

Shehzad, W. (2005). Time Related Idiomatic Language: A Corpus-based Approach to TEFL with Reference to MICASE. *Journal of Asia TEFL*, *2*(4).

Strevens, P. (1988). *ESP after twenty years: A reappraisal*. State of the Art.

Sun, Y. C., & Wang, L. Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, *16*(1), 83-94.

Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. Routledge.

Wang, X.X. (2007). *An analysis of English Majors' needs in BE course design.* (Unpublished master's thesis), Northeast University, Changchun.

West, R. (1994). Needs analysis in language teaching. *Language teaching*, *27*(1), 1-19.