Lexical Analysis of Intermediate English Course Books: A Corpus Based Study

Kalsoom Jahan, Muhammad Asim Mahmood and Wardah Azhar

Department of Applied Linguistics, Government College University, Faisalabad, Pakistan

ABSTRACT This research was conducted to analyse vocabulary included in intermediate English course books recommended by Punjab Curriculum and Textbook Board. It aimed to compare the frequency of words used in English course books with British National Corpus, and academic word list to identify the effectiveness of intermediate English course books in learning vocabulary. For this purpose a corpus was developed from the intermediate English course books. Data was analysed by using compleat lexical tutor and antconc 3.5.7 version. The results show that presentation of vocabulary items in intermediate English course textbook is very complex and very less coverage has been identified while tracing British National Corpus families and academic word list coverage.

INTRODUCTION

In Pakistani context, literature has been an important aspect of English syllabus course outline and there are multiple ways of using different literary genres such as novel, poetry, drama, and short stories in English language classrooms. At intermediate level, the students have to face literature for language learning, even for the practice of academic reading and writing government has not prescribed any book. Punjab Curriculum and Textbook Board has just included topics for writing essays, letters, applications, stories and a few other grammatical items. So the stuff for practicing grammatical items and prose writing remains in sufficient. Though literature has its own importance, as Khatib et al. (2011), assert that in an ELT environment literary genres can be used in various ways for learning language skills.

The presentation of lexical items in textbooks is the most important aspect of textbook development. In literature based textbooks, sometimes text demonstrates such vocabulary that is out of use these days. Lipinski (2010) observed that the use of low frequency vocabulary items in textbooks tends to be hardly used in daily life. So the textbook developers should avoid low frequency words in order to reduce load from learners' minds.

Address for correspondence:
E-mail: kalsoomjahan1@gmail.com

Nation (2013) has observed that the issue of controlling and presenting vocabulary items has been neglected in the process of material development for language teaching and learning. Nation and Beglar (2007) found that learning of high frequency words is closely related to text coverage in a text. Nation (2006) asserts that 1,500-2,000 high frequency words need more attention and should be learnt quickly in learning vocabulary. However, English textbooks published in Pakistani contexts, the high frequency words of English language spoken and written by the native speakers are not sufficiently investigated. So, it evoked the researchers to investigate the intermediate English textbooks vocabulary items and its coverage in most frequently used English lexicon.

Laufer (1997) recommends that in L2 reading, ninety five percent lexical threshold is required to enable a learner to apply L1 reading strategies. Researches has reflected that a vocabulary size with 3,000 words is needed for the comprehension of the text, but to attain ninety eight percent coverage of the text, a vocabulary size of 5,000 word families is needed. According to Hirsh and Nation (1992), to enjoy a pleasurable reading of the text and exact guessing of unknown words in the context 5,000 word families are needed as vocabulary size. Hu and Nation (2000), studied the effect of coverage level (80%, 90%, 95%, and 100%) in four different texts. They used fictional texts for unassisted reading comprehension and observed an increase in the

density of known words supplemented the process of reading comprehension. The findings of their study concluded that some readers required ninety five percent while most of the readers needed ninety eight percent lexical coverage for adequate unassisted reading comprehension. According to these results, the researcher took interest in analysing English course books of Intermediate whether these course books are presenting the same coverage of lexical items for a better reading comprehension or based on just randomly selected lexical items.

Purpose of the Study

The purpose of the study is to find out the patterns of presenting lexical items in intermediate English course books (IECBs) recommended by Punjab Curriculum and Textbook Board, and to compare the frequency of words used in English course books with British National Corpus (BNC), and Academic Word List (AWL) to identify the effectiveness of IECBs in reading comprehension.

Research Objectives

The objectives of the study are to:
- To trace out the frequency of BNC in intermediate English course books.
- To find out the frequency of AWL in intermediate English course books.
- To investigate the effectiveness of lexical choices in reading comprehension.

Research Questions

Research questions of the study are:
- What is the frequency of BNC in intermediate English course books?
- What is the frequency of AWL in intermediate English course books?
- What is the effectiveness of IECB lexical choices in reading comprehension?

Literature Review

Presenting vocabulary in textbook is a difficult task. Different researches have established criteria of including words in textbooks at a particular level so that the students may get adequate comprehension of a text. Nation (2006) maintains that high frequency words have a

close relation with text coverage and it also effects the difficulty level of the text. Nation and Beglar (2007) add that for successful language learning and having good command over reading comprehension, the frequency of words in a particular text of a particular level should never be neglected. Nation (2006) has presented the idea of vocabulary size in a text for unassisted reading comprehension. He took fourteen thousand word family list form British National Corpus (BNC), and checked the frequency of word families in a novel titled 'Lady Chatterley's Lover' and found that eighty eight percent was covered in first two lists of BNC. Similarly, in other case, he analysed the speeches of children and found that 250 most frequent words have covered seventy five to eighty percent of children's total language production. Therefore, it can be assumed that first two thousand most repeated words cover a major part of any discourse. Learning of first two thousand most frequent words can help in language learning and can improve reading and listening comprehension of the students.

According to McEnery et al. (2006) Corpus is a collection of data that is machine readable, authentic, and is sampled to be representative of a particular language or language variety. O'Keeffe et al. (2007) define corpus as a collection of texts, spoken or written, which can be stored in computer. Biber et al. (2007) mention that it a collection of available text based on the principle qualitative and quantitative analysis. Introduction of corpus-based studies have changed the way of vocabulary studies. Moreover, corpora have been used broadly to provide more exact descriptions of language use, learning and teaching. According to McEnery et al. (2006), a large number of scholars have also utilised corpus data to look critically at current syllabi and textbooks for Teaching English as a Foreign Language (TEFL).

Repetition of lexical items in a text is a very important aspect of English course books. Several researches have been conducted to investigate the effectiveness of repetition of vocabulary items in a text and its effects on reading comprehension. Nation (2001) found that word repetition was a useful condition in presenting vocabulary items in the text and had favourable effects on vocabulary learning and teaching. Word repetition in a text affects the learning vocabulary process on various grounds as, types of repetition, spacing of repetition and

number of repetition. Huckin and Coady (1999), mention that there are not fixed number of repetition of lexical items that can guarantee its learning. However, Webb (2007) suggests that there must be at least 10 repetitions of unknown words in a text. Hirsh and Nation (1992), have separated the texts according to their topics such as newspapers of same and different events. Hwang and Nation (1989) suggested the categories of text according to their themes. Gardner (2008) adds to it saying that repetition of unknown words is more favourable when lexical items are presented in unrelated texts. In this way, it provides an improved form of text for vocabulary teaching and learning. It can be used to enhance the effect of incremental acquisition of the most repeated vocabulary items while reading.

English course books which consist of fiction-oriented reading texts such as telling short stories to enhance the narrative ability and vocabulary acquisition, and this aspect is highly prominent in intermediate English course books of Punjab, Pakistan. Therefore, this point also evoked the researchers to analyse the English course books lexically to investigate the usefulness of these books in vocabulary learning.

In Pakistan, English course books are used in formal setting and mostly the teaching techniques are reading based. As intermediate English course books are literature based so direct vocabulary instructions are the most commonly used. Ellis (1990) and Nation (1993) have recommended direct vocabulary instructions with ELT literature. Mehrpuor and Rahimi (2010) have investigated the impact of specific and general vocabulary knowledge on listening and reading comprehension and found that knowledge of lexical items can improve the comprehension level of the text. Beglar et al. (2012) has also investigated the vocabulary knowledge in relation with reading literary texts and comprehension level. They also found vocabulary knowledge favourable in reading comprehension.

Yang and Dai (2011) found that, in Asian countries, memorization and repetition of unknown words are the basic techniques utilised for vocabulary learning. The students find equivalent word in native language to learn new lexical items given in a text. However, these types of activities, which desexualise a word from the text, are not much helpful in learning new words of a language.

Therefore, it is clear from above studies that presentation of vocabulary items in a text is very important in learning vocabulary which plays a vital role in reading comprehension. A valid threshold of vocabulary items and its coverage in a text is the most important point which must be kept in consideration while developing a textbook. In Pakistan, intermediate English course books are based on literature like Book-1 is based on short stories, Book-2 consists of prose items, Book-3 deals with poetry and drama, and Book-4 is a novel. So the researchers thinks that it is valuable to investigate the patterns of presenting lexical items in intermediate English course books and its comparison with the most frequent words of native lexicon.

In Corpus studies, Nation (2006), has divided vocabulary into twenty, 1,000 or in some recent researches into twenty 1,000 word families. British National Corpus (BNC) frequency based list which comprises of ninety percent formal written and ten percent spoken text. Nation (2006) claims that BNC word list presents the learners' vocabulary largely according to its range and frequency. BNC list provides more detailed and accurate evaluation of the vocabulary burden of the text, as these lists give coverage to a large amount of word families.

In order to categorize the most valuable, useful and frequent words in educational contexts, a variety of vocabulary lists have been collected from corpora. Coxhead (2000) mentions that corpus based studies can be easily accessed by the teachers and researchers. Moreover, it is easy to compile corpus based material for language classrooms for the assistance of teachers of any discipline. According to Coxhead (2000, 2011), the AWL word forms account for ten percent of the tokens in the Academic Corpus that is a representative text from the academic domain. Brezina and Gablasova (2015), have added that it should be eminent that West's General Services List GSL (1953) functioned as the non-academic standard for the creation of the AWL. Moreover, Coxhead (2000) omitted the 2000 GSL words from the academic corpus. However, taken together with the first 2000 high-frequent word families from GSL, AWL provides the coverage of eighty six percent of the above mentioned corpus. In contrast to the Academic Corpus, the AWL word forms an account for almost 1.4 percent of tokens from fictional collection that counts as a collection of non-academic texts.

This proposes that the majority of the word families, in the AWL, occur more frequently in academic texts and less frequently in fictional books.

AWL is a list which includes 570 academic word families. Overall, it has a good coverage of academic texts regardless of subject area. Coxhead (2011) adds that the goal of AWL has been set for EAP (English for academic purposes) and in many countries it is being widely used in classrooms with a wide range of materials. It has become a major source for the researchers.

Chung and Nation (2003) have observed that various studies have been conducted to explore the frequency, range and distribution of the AWL words that are commonly used in applied linguistics' books and research articles. Khani and Tazik (2013) add to it saying that AWL helps in developing and analysing science-specific middle school textbooks. Matsuoka and Hirsh (2010) mention that in college English as well as medical and agriculture research articles, the percentage of the AWL vocabulary varies from study to study.

In this study, intermediate English course book is going to be analysed by making a comparison with BNC and AWL to investigate the effectiveness of this course in presenting vocabulary and its role in reading comprehension. This study investigates the questions related to Intermediate English course, its lexical coverage with twenty 1,000 word families of BNC and 570 word families of AWL. Moreover, its effects on reading comprehension are also the part of this investigation.

RESEARCH METHODOLOGY

Research Design

In this study, analytical research design was used for lexical analysis of intermediate course books. A comparison between English course books' vocabulary and British National Corpus (BNC) and academic word list (AWL) was executed quantitatively.

Population

All lexical items included in English textbooks, recommended by Punjab Curriculum and Textbook Board for intermediate level, were the population of the study which involve short stories, dramas, novel, prose and poetry.

Collection of Intermediate English Textbook Corpus

The corpus consists of four English textbooks recommended by Punjab Curriculum and Textbook Board. These four books are based on English literature as it includes short stories, one-act plays, poetry, and prose. These books are prepared and published by Government of Punjab, Pakistan under the supervision of ministry of education (curriculum wing), Islamabad. For the development of corpus, it involved the collection of texts from all the four English textbooks in electronic form.

All indices, appendices, and pictures were removed from the text to improve the reliability of word counting. All exercises were included in the text as it involved students in repeating the words in different contexts. All proper nouns were also included in the text as it could affect the text coverage and exclusion could shrink the text. In addition as, BNC corpus has not listed the proper nouns in 1st to 20th 1000 most frequent words, so in this research all proper nouns have been filtered in off-list.

Electronic version of these English textbooks was retrieved from Punjab Curriculum and Textbook website. In case, if the books were not available electronically, the content of the book was scanned on dpi 300 resolution and then converted into MS Word through Free OCR Software. The MS Word text was then converted into plain text that is notepad file. Due to some possible alterations, which were related to the images included in the textbook, some sections and parts of the chapters were typed or edited manually. Spelling mistakes, which occurred during conversion process, were also corrected with the help of spell checker.

The Instrument

In this study, vocabProfile and range from the website of compleat lexical tutor was used as an instrument. Vocabprofile is a program which is used for vocabulary analysis in relation with different renowned world corpuses. This program breaks the text into families, types and token and correlates it from 1st 1,000 to 20,000 BNC word families, and even more than 20,000 thousand families. BNC lists, according to the classification of word families are available at the website of Compleat lexical tutor (Webb and Nation

2008; Nation and Webb 2011). Vocabulary items from many texts can be compared to a provided specific text or a single big corpus by the using of Rang program. It provides with frequency results which tells that how much same vocabulary exists in the sub-corpora and frequency of words in total.

Thus, for the vocabulary analysis of Intermediate English course books, a comparison of English course books vocabulary was made with the two categories of word lists: (1) the BNC 1st to 20th 1000 word families; and (2) Coxhead's (2000) AWL containing 570 word families. According to Nation and Webb (2011), word family is the most sensible unit in receptive knowledge, as is in listening and reading. Little learning is required when a learner knows the one family member of the word then it is easy to recognise the other family members of that word in the text. The rationale for selecting the concept of a word family as a counting unit is that Bauer and Nation (1993) have supported the idea that if the meaning of the based word is inferred, then the meaning of the whole family of that word can be easily known to the learner. For example, grouping of the headword appreciate with its family members goes as appreciates, appreciated, unappreciated, appreciation, appreciable, appreciably, appreciating. So, if the learners can identify the meaning of the head word, they are assumed to infer the meaning of other family members as well.

As word family is an important part of vocabulary learning and teaching, most frequent words are counted on the basis of considering word family as a unit.

Data Analysis

In this part of the study, twenty 1,000 word-family list from BNC was used. BNC is a corpus with 100 million tokens including ten percent spoken and ninety percent written English texts. All of the information about word type, token and lemma lists of BNC, including range, frequency and dispersion.

In this study, the way has been applied to check whether the word family lists are properly ordered in IECBs. From the first 1,000 to the twenty 1,000, the number of types, tokens, and families found in IECBs should decrease. That is, when a correlation proceeds with BNC, the first 1,000 word family list should include more types,

tokens, and families than the second 1,000 family list. Same is the case with the next one, like two 1,000 family lists should include more types, tokens, and family lists than the three 1,000 word lists, and the case should move on till twenty 1,000 words family list.

The purpose behind developing a list of lexical items included in IECBs was to determine the higher frequency of words introduced to the learners of English at intermediate level. It is assumed that the range and frequency of a word affects the learning processes of vocabulary among students. Laufer et al. (2004) found that wide range and higher frequency improves vocabulary of the learners and lower frequency and narrower range results affect the scoring of the students in vocabulary tests. So, it is clear through this study that the frequency and rang of the word in the text affect the vocabulary learning process.

RESULTS

The Frequency of BNC Lexical Items in English Course Books of Intermediate

IECBs consist of four intermediate English course books. The major part of the corpus is covered by Book-1. Moreover, Book-4 has contributed less numbers of tokens as compared to the other books included in this corpus (see Table 1).

Short stories = Book-1
Prose and Heroes =Book-2
Drama and poetry = Book-3
Novel = Book-4

Table 1: Total numbers of tokens (running words) in IECBs

	Book-1	Book-2	Book-3	Book-4
Total numbers of tokens (running words)	35850	33578	27865	18525
Total numbers of tokens in IECB	115,818			

The Coverage of BNC in the University English Textbooks

Table 2 shows a clear description of family and token coverage in four sub corpora. It is very clear that only Book-2 achieved the target

Table 2: Word families and cumulative text coverage at each word level in four sub-corpora of IECBs

Wordlist (1000)	Book-1		Book-2		Book-3		Book-4	
	Families and textcoverage %		Families and textcoverage %		Families and textcoverage %		Families and textcoverage %	
K-1	853	78.09	883	77.36	798	78.21	775	79.23
K-2	587	84.88	687	85.80	483	84.04	422	85.69
K-3	373	88.22	396	89.15	284	87.12	264	88.35
K-4	230	89.82	262	90.95	157	88.90	147	89.66
K-5	149	90.72	192	92.28	120	89.81	102	90.51
K-6	102	91.33	131	93.23	83	90.56	70	91.09
K-7	75	91.80	102	93.86	61	91.16	53	91.46
K-8	63	92.29	76	94.29	28	91.46	42	91.74
K-9	63	92.62	79	94.71	37	91.70	30	91.95
K-10	28	92.81	45	94.97	38	91.92	23	92.11
K-11	31	93.03	38	95.16	25	92.13	28	92.29
K-12	22	93.19	43	95.40	12	92.20	18	92.41
K-13	28	93.33	23	95.51	13	92.25	19	92.53
K-14	5	93.35	15	95.60	8	92.29	10	92.59
K-15	10	93.42	17	95.69	12	92.38	8	92.64
K-16	4	93.44	7	95.73	3	92.41	3	92.66
K-17	3	93.46	7	95.81	-	-	5	92.69
K-18	2	93.47	5	95.84	1	-	3	92.71
K-19	7	93.62	3	95.87	6	92.49	4	92.75
K-20	1	93.64	-	-	1	92.50	-	-
0ff list	?	100.00	?	100.00	?	100.00	?	99.99

of ninety five percent coverage at level K-10 that is considered to be high, as According to Nation (2006), the text covering ninety eight percent in the first five 1,000 most frequent words from BNC is easy to read. So, if Book-2 has coverage of ninety five percent at level K-10, it means it is suitable for intermediate level but the whole corpus of Book-2 is not covering ninety eight percent in twenty 1,000 words from BNC. Moreover, Book-1, Book-3, and Book-4 have not achieved the target of ninety five percent even at level K-20, that is showing that the text is too difficult and suitable for intermediate learners. If off lists are counted, Book-4 has not hundred percent coverage, and it contained a few words from Latin and French and those are not the part of BNC. Overall, Book-2 is showing better presentation of lexical choices as compared to the other books included in IECB.

Frequency and Distribution of Academic Words in the Intermediate English Course Books

Table 3 indicates that the highest AWL coverage is observed in Book-2 (2.59%) and the lowest in Book-1 (1.14%). However, the coverage found in Book-3 (1.22%) and Book-4 (1.69%) is not more than two percent. Overall, the AWL

coverage in four sub corpora did not rise up to three percent. Token coverage in the corpora is below average to provide commonly used academic words that the students should know in reading intermediate English textbooks. As per counting the number of AWL families, the highest numbers of families are counted in Book-2 (278), and the lowest numbers of AWL families are counted in Book-3 (146). However, the numbers of AWL families in any book did not rise up to 280.

Table 3: The number of AWL vocabulary occurrences in tokens, types and families

TIECB	AWL Families (%)	AWL Types (%)	AWL Token (%)
Book-1	173 (10.44)	221 (4.23)	410 (1.14)
Book-2	278 (15.41)	424 (7.40)	869 (2.59)
Book-3	146 (10.16)	187 (4.46)	339 (1.22)
Book-4	162 (12.10)	199 (5.24)	313 (1.69)

Frequently Used AWL Lexical Item in Intermediate English Course Books

Table 3 shows that only 173 AWL families occurred in Book-1, with author (26), consume (18), theme (16), tape (14) and all others have

less than 10 frequency. Noting point is that most of the words have been found to occur only once or twice in the text, so the concept of repetition is highly ignored.

Only 278 AWL families have been found to be repeated in Book-2, with method (36), medical (26), culture (14), chemical (13), economy (13), physical (10), and the rest have less than 10 frequency. Though in this book a large number of families stand at the frequency of seven, eight and nine but still it is a noting point that most of the words occurred only once or twice in the text, so the concept of repetition is highly ignored.

One hundred and forty six (146) AWL families appeared in Book-3, with theme (25), image (15), create (13), attribute (10) and the frequency of the rest is less than 10. A low frequency of words is observed in the text and it is a noting point that most of the words have occurred only once or twice in the text, so the concept of repetition is highly ignored.

In Book-4, only the word the "chapter" counts the frequency of 18 but the frequency of the rest of the families is less than 10. A low frequency of words is observed in the text and it is a noting point that most of the words have occurred only once or twice in the text, so the concept of repetition is highly ignored.

DISCUSSION

The Frequency of BNC Lexical Items in English Course Books of Intermediate (IECBs)

The first research question was dealing with the frequency of BNC lexical items included in intermediate English course books. So to answer this question the VocabProfile from compleat lexical tutor was used in the sub corpora of IECBs.

The Coverage of BNC in the University English textbooks

In Book-1 the frequency of BNC families, types and token was dropping down consistently till K-8 but inconsistency begins from K-9 onward. Then, it was clear up to K-20. The lexical coverage up to K-20 level was 93.64 percent, which showed that the text was not easy and had included complex vocabulary. For, the

1st thousand words are the most frequent running words of daily use and as the level increases, the level of difficulty also rises. The inconsistency also enhances the difficulty level of the text. So it is not an easy text to be read by the students independently. The number of families and the accumulative text coverage have been summarised in Table 1. The cumulative text coverage of BNC has reflected that K-1 had covered only 78.09 percent from Book-1, and till the level of K-5, mean from 5,000 families the text has coverage of 90.72 percent. Even 19,000 word list covered only 93.64 percent of the lists. These results reflect that the text is too difficult and has complex vocabulary distribution. There is a wide difference in vocabulary threshold which is needed to read and to comprehend the Book-1. Laufer (2000) suggested that to know 5,000 word families ninety five percent coverage would be needed but Nation (2006) proposed that knowing 8,000 word families for ninety five percent coverage would be sufficient. This observation also correlated with the type of the text.

Considering Nation's (2006) point of view it revealed that there was a significant difference in the vocabulary threshold which was required to read and comprehend the Book-1 investigated for this study. The results had shown that, the percentage of the word families remained inconsistent after K-8.

In Book-2, the frequency of BNC families, types and tokens was decreasing consistently till K-8 but inconsistency started from K-9 onward. Then, it was detected up to K-20. The target of ninety five percent started from K-8 mean 8,000 word families as indicated by Nation (2006), so this text was not suitable for the beginners but for the intermediate level. Here again the difficulty of the text also related with type of the text. The lexical coverage up to K-20 level was 95.87 percent, which reflected that the text included complex vocabulary. As the level increased the level of difficulty raised. The inconsistency also enhanced the difficulty level of the text. So, it was such an easy text as the student could read it independently. The number of families and the accumulative text coverage is summarized in Table 1. The cumulative text coverage of BNC has reflected that K-1 has covered only 78.09 percent from Book-1, and till the level of K-5, mean from 5,000 families the text has cov-

erage of 90.72 percent. Even 19,000 word list covered only 93.64 percent coverage. These results reflected that Book-2 had difficult text and complex vocabulary distribution, as claimed by Laufer and Ravenhorst-Kalovski (2010).

In Book-3, the frequency of BNC families, types and token was decreasing consistently till K-8 but inconsistency started from K-9 onward. Then it was observed up to K-20. The target of ninety five percent that was claimed by Nation (2006) was not achieved even at the level of 20,000. At the level of K-20 the cumulative text coverage was observed as 92.5 percent. Even no lexical item was included at the level of K-17, and at the point of K-19 only one word family was introduced from BNC. So, this text was not suitable for the beginners but for the advanced level learners, here again the difficulty of the text also related with type of the text. The lexical coverage upto K-20 level was 92.5 percent, which reflected that the text included complex vocabulary. The inconsistency also boosted the difficulty level of the text. So, it was not an easy text to be read independently by the learners.

In Book-4, the frequency of BNC families, types and token was dropping down consistently. It is observed only in the case of Book-4. The target of ninety five percent that was claimed by Nation (2006) was not achieved even at the level of 20,000. At the level of K-20 the cumulative text coverage was observed as 92.5 percent. Even no lexical item was included at the level of K-20. If off-list is also included, even then the text is not totally covered by BNC as it has shown that text has some lexical items which are not included in BNC. So, it makes the text more difficult than Book-1, Book-2 and Book-3. It can be suitable for advance learners but not for the intermediate level. The lexical coverage up to K-20 level was 92.75 percent which reflected that the text included complex vocabulary items.

Frequency and Distribution of Academic Words in the Intermediate English

The highest AWL coverage was observed in Book-2 (2.59%) and the lowest in Book-1 (1.14%). However, the coverage found in Book-3 (1.22%) and Book-4 (1.69) was not more than two percent. Overall the AWL coverage in four sub-corpora did not rise upto three percent To-

ken coverage in all corpora was below average to provide commonly used academic words that student must know in reading intermediate English textbooks.

As per counting the number of AWL families, the highest numbers of families were counted in Book-2 (278), and the lowest numbers of AWL families were counted in Book-3 (146). However, the numbers of AWL families in any book did not rise up to 280. The results have shown that AWL coverage in IECBs is not up to the satisfactory level which means that the text is lacking such academic words as are necessary in reading comprehension.

The results can be well observed in the light of the previous researches. Vongpumivitch et al. (2009) found in his research that the text coverage of the AWL word families in English textbooks corpus was 6.5 percent which was much lower than the 11.7 percent of the AWL words in a study on applied linguistics' research articles. Chen and Ge (2007) observed in their study ten percent of the AWL coverage in medical research articles which was mentioned (13.1%) by Chung and Nation (2003). However, in this study, the frequency analysis shows that 388 of the 570 AWL word families are included in IECBs. It's a very low coverage which indicates that reading the text is not suitable to the academic purposes at intermediate level.

Frequently Used AWL Lexical Item in Intermediate English Course Books

Only 173 AWL families occurred in Book-1, with author (26), consume (18), theme (16), tape (14) and all of the others had less than 10 frequencies. Noting point was that most of the words occurred only once or twice in the text, so the concept of repetition was highly ignored. Table 3 shows that only 278 AWL families have occurred in Book-2, with method (36), medical (26), culture (14), chemical (13), economy (13), physical (10) and the rest have less than 10 frequencies. Though, in this book, a large number of families were observed to occur seven, eight and nine times but still it is a noting point that most of the words occurred only once or twice in the text, so the concept of repetition was highly ignored.

Only 146 AWL families occurred in Book-3, with theme (25), image (15), create (13), attribute (10) and the frequency of the rest was less than

10. A low frequency of words was observed in the text and it was a noting point that most of the words occurred only once or twice in the text, so the concept of repetition was highly ignored.

In Book-4, 162 AWL families occurred, only the word "chapter" counted the frequency of 18 and the frequency of the rest was less than 10. A low frequency of words was observed in the text and it was a noting point that most of the words occurred only once or twice in the text, so the concept of repetition was highly ignored.

CONCLUSION

The results have shown that the frequency level of vocabulary in intermediate English course books is different as compared to those of the BNC textual coverage and word family lists in four subcategories of the main corpus. If the students have command over the 1st two 1,000 family words, they can comprehend the text of IECBs and can further benefit from other unknown words surrounded in the text. Though the starting 1st thousand coverage is not bad but moving further it is observed that only Book-2 achieved ninety five percent coverage in the text, but the rest cannot achieve, even Book-4 has such words as are not on BNC family word list.

Considering the intermediate English textbooks, it is evident that the students of intermediate need more than 20,000 word families to get ninety five percent coverage of all of the books included in the corpora. The intermediate students have to struggle a lot with these books having little vocabulary size of 2,000 words. Other factors such as those connecting with language proficiency, length of the text, difficulty, topic of interest, motivation, inferencing skills and L1 reading ability all play a part and can be addressed through instruction that supports students in improving reading comprehension.

Concerning the coverage and frequency of distribution of the AWL words, the results exposed that, although AWL vocabulary lists might provide some rules for teaching EAP, separate items occurred in the corpus of intermediate English course books. Moreover, AWL coverage is very low. For these reasons, it can be said that the intermediate English course books are not having sufficient vocabulary related to BNC. The distribution and selection of words from different levels is also very complex that has made the text difficult for reading comprehension.

RECOMMENDATIONS

It is recommended that intermediate English course books need revision. Amendments must be done according to the needs of the students. There might be other factors responsible for textual difficulty which the future researchers are invited to explore. In this regard lexical analysis of secondary school English course books might be helpful in determining the difficulty level of IECBs vocabulary. Furthermore, graduation English course books might also be lexically evaluated to identify the effectiveness of IECBs lexical coverage.

REFERENCES

Bauer L, Nation ISP 1993. Word families. *International Journal of Lexicography,* 6(4): 253-279.
Beglar D, Hunt A, Kite Y 2012. The effect of pleasure reading on Japanese university EFL learners' reading rates. *Language Learning,* 62: 665-703. Doi: 10.1111/j.1467-9922.2011.00651.
Biber D, Connor U, Upton T 2007. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure.* Amsterdam: John Benjamins.
Brezina V, Gablasova D 2015. Is there a core general vocabulary? Introducing the new general service list. *Applied Linguist,* 36(1): 1-21.
Chen Q, Ge G 2007. A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families. *Medical Research Articles,* 235-253.
Chung TM, Nation ISP 2003. Technical vocabulary in specialized texts. *Read Foreign Lang,* 15(2): 103-106.
Coxhead A 2000. A new academic word list. *TESOL Quarterly,* 34(2): 213-238.
Coxhead A 2011. The academic word list 10 years on: Research and teaching implications. *Tesol Quarterly,* 45(2): 355-362.
Ellis R 1990. *Instructed Second Language Acquisition.* London: Blackwell.
Gardner D 2008. Vocabulary recycling in children's authentic reading materials: A corpus based investigation of narrow reading. *Reading in a Foreign Language,* 20: 92-122.
Hirsh D, Nation ISP 1992. What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language,* 8: 689-696.
Hu M, Nation ISP 2000. Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language,* 13(1): 403-430.

Huckin T, Coady J 1999. Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition,* 21: 181-193.

Hwang K, Nation ISP 1989. Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. *Reading in a Foreign Language,* 6: 323-335.

Khani R, Tazik K 2013. Towards the development of an academic list for applied linguistics research articles. *RELC Journal,* 44(2): 209-232. Doi: 10.1177/0033688213488432.

Khatib M, Rezaei S, Derakhshan A 2011. Literature in EFL/ESL classroom. *English Language Teaching,* 4: 201. Doi: 10.5539/elt.v4n1p201.

Laufer B 1997. *The Lexical Plight in Second Language Reading.* Cambridge: Cambridge University Press.

Laufer B 2000. Avoidance of idioms in a second language: The effect of L1 L2 degree of similarity. *Studia Linguistica*, 54(2): 186-196.

Laufer B, Elder C, Hill K, Congdon P 2004. Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing,* 21(2): 202-226. Doi: 10.1191/0265532204lt277oa.

Laufer B, Ravenhorst-Kalovski GC 2010. Lexical threshold revisited: Lexical text coverage, learner's vocabulary size and reading comprehension. *Read Foreign Lang,* 22(1): 15-30.

Lipinski S 2010. A frequency analysis of vocabulary in three first year textbooks of German. *JSTOR,* 43(2): 167-174. Doi: 10.1111/j.1756-1221.2010.00078.x

Matsuoka W, Hirsh D 2010. Vocabulary learning through reading: Does an ELT course book provide. *Reading in a Foreign Language,* 22(1): 56-70.

McEnery T, Xiao R, Tono Y 2006. *Corpus-based Language Studies.* London, New York: Routledge.

Mehrpuor S, Rahimi M 2010. The impact of general and specific vocabulary knowledge on reading and listening comprehension: A case of Iranian EFL learners. *System,* 38(2): 292-300. Doi: 10.1016/j.system.2010.01.004.

Nation ISP 1993. Using dictionaries to estimate vocabulary size: Essential, but rarely followed, procedures. *Language Testing,* 10(1): 27-40.

Nation ISP 2001. *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press.

Nation ISP 2006. How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review,* 63(1): 59-82.

Nation ISP 2013. *Learning Vocabulary in Another Language.* 2nd Edition. Cambridge: Cambridge University Press.

Nation ISP, Webb S 2011. Content-based instruction and vocabulary learning. In: E Hinkel (Ed.): *Handbook of Research in Second Language Teaching and Learning.* Volume 2. New York: Routledge, pp. 631-644.

Nation ISP, Beglar D 2007. A vocabulary size test. *The Language Teacher,* 31(7): 9-13.

O'Keeffe A, McCarthy MJ, Carter RA 2007. *From Corpus to Classroom: Language Use and Language Teaching.* Cambridge: Cambridge University Press.

Vongpumivitch V, Huang JY, Chang YC 2009. Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes,* 28(1): 33-41.

Webb S 2007. The effects of repetition on vocabulary knowledge. *Applied Linguistics,* 28(1): 46-65.

Webb S, Nation P 2008. Evaluating the vocabulary load of written text. *TESOLANZ Journal,* 16: 1-10.

Yang W, Dai W 2011. Rote memorization of vocabulary and vocabulary development. *English Language Teaching,* 4(4): 61-64.