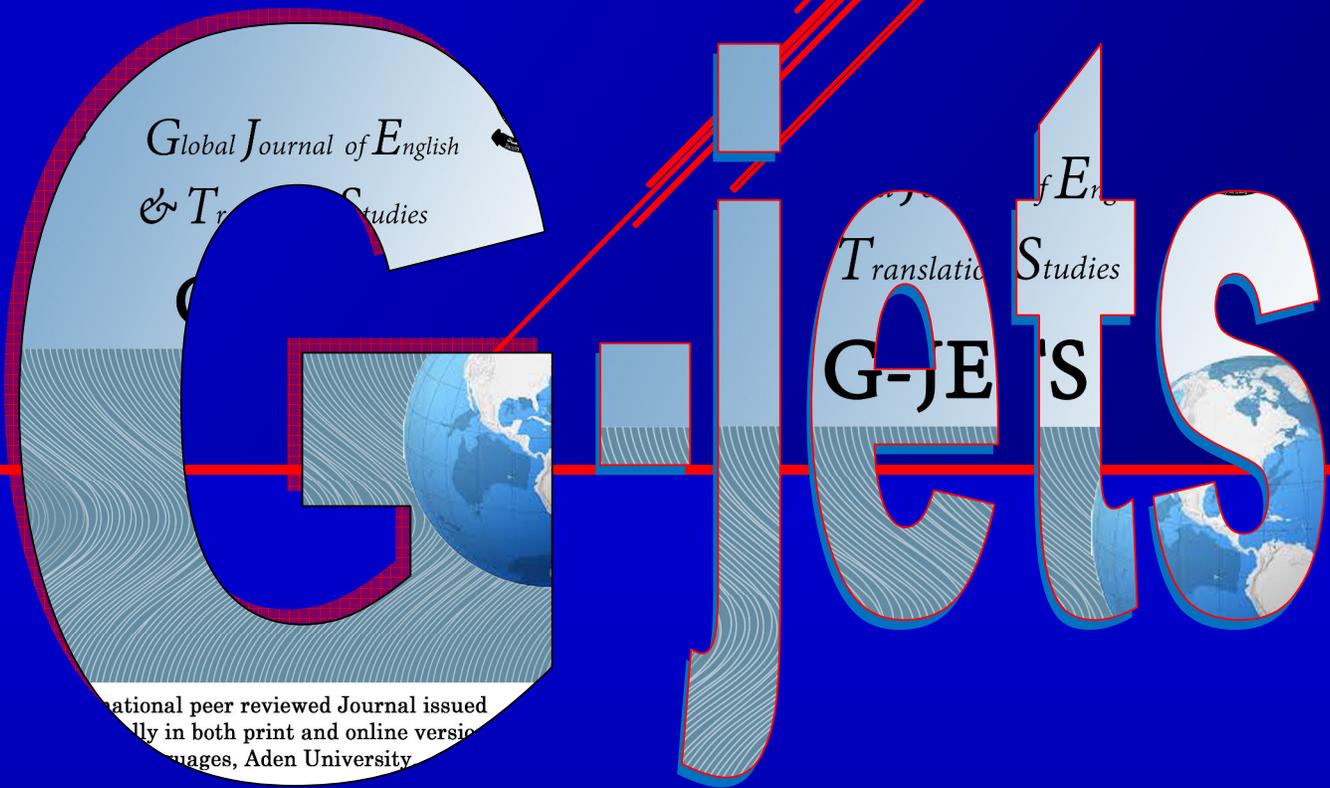


ISSN No. 2415-1777



The title 'G-JETS' is rendered in a large, stylized font. The letter 'G' is the largest and contains a globe and the text 'Global Journal of English & Translation Studies'. The letter 'J' is tall and thin, also containing a globe. The letters 'E', 'T', and 'S' are shorter and wider, with 'E' containing a globe and 'S' containing a globe. The letter 'T' is tall and thin, also containing a globe. The letter 'S' is shorter and wider, also containing a globe. The letter 'J' is tall and thin, also containing a globe. The letter 'E' is shorter and wider, also containing a globe. The letter 'T' is tall and thin, also containing a globe. The letter 'S' is shorter and wider, also containing a globe. The letter 'G' is the largest and contains a globe and the text 'Global Journal of English & Translation Studies'. The letter 'J' is tall and thin, also containing a globe. The letters 'E', 'T', and 'S' are shorter and wider, with 'E' containing a globe and 'S' containing a globe. The letter 'T' is tall and thin, also containing a globe. The letter 'S' is shorter and wider, also containing a globe. The letter 'J' is tall and thin, also containing a globe. The letter 'E' is shorter and wider, also containing a globe. The letter 'T' is tall and thin, also containing a globe. The letter 'S' is shorter and wider, also containing a globe.

Global Journal of English and Translation Studies
Online Version: ISSN 2415-1777

An international peer reviewed journal issued semi-annually in both print and online versions by the Faculty of Languages, University of Aden.

Issue No.3 January 2017

Chief Editor

Prof. Dr. Gamal Mohammed Ahmed Abdullah

Director of Editors

Dr. Abdalnasser Mohammed Ali Naqeeb

Editorial Secretary

Dr. Nabil Mohammed Saleh Muhayam

Editorial Board

Dr. Adel Abdulkhaliq Abdulkareem Dulul

Dr. Raba'a Taleb Thabe Abadel

Dr. Mohammed Ali Abdullah Albadri

Editorial Correspondence

Journal email, editor and Executive

Editorial Board emails.

Website: <http://langfaculty-aden.com>

Email: globaljets001@gmail.com

Tel: +967 02-236476

About the Journal

The Global Journal of English and Translation Studies (G-JETS) is a high quality open access peer reviewed research journal published by the Faculty of Languages- Aden University. It is a brainchild of a group of academicians and research scholars in the field of English language and translation studies. It aims to provide a platform for the researchers, academicians, professionals, practitioners and students to impart and share knowledge in the form of high quality empirical and theoretical research papers, case studies, literature reviews and books reviews to make them available to scholars of English language studies and translation. It is a biannual publication (January & June). The journal publishes in both print and online versions. The G-JETS was founded in 2014 by Prof. Dr. Gamal Mohammed Ahmed Abdullah, dean of the Faculty of Languages- University of Aden.

Editorial

As promised previously, the Global Journal of English and Translation studies (G-JETS) continues in presenting new additions to its new issues. This time, we are pleased to introduce the third issue of the journal with both print and online versions. Each version has its own ISSN No. Though the current and difficult situation in which the country goes through, the G-JETS proves its commitment towards its subscribers, readers and followers and appears this time in its prints and online versions. The online version is a new and wonderful addition to our print version which appeared in the previous two issues. We are issuing this issue with an online version hopefully that it will facilitate access to the G-JETS globally. Here, we seize this opportunity to express our thanks to our subscribers, readers and publishers for their continuous support and participation. We promise you that the G-JETS will continue presenting its best in every issue it appears in.

Chief Editor

Prof. Dr. Gamal Mohammed Ahmed Abdullah

Notes for Contributors

Editing Requirements

1. Paper should be in Times New Romans font with 12 font size on 4A paper printed at single side.
2. Margins should be 2.5 cm on four sides.
3. The soft copy of the paper should be in MSW 2007 OR any latest MSW version. Or log to our website to get a template.
4. A PDF format copy should also be included.
5. Contributors should follow APA referencing system.
6. The paper should not exceed 5000 words.
7. The paper should be acknowledged with a declaration that it is an original work and has not been published anywhere else.
8. Abstract should not be more than 200 words.
9. A short bio-note of the author should include name, institutional affiliation and brief career history.

Selection Procedures

All submitted papers will be screened by peer reviewing committee. Final selection will be made if the paper is selected by reviewers for publications. The editors reserve the right to make necessary editing of the selected papers for the sake of conceptual clarity and formatting.

Plagiarism Alert

It is warned to the contributor(s) that they must adhere to academic ethics with respect to acknowledgment of original ideas of the others. It will be the contributors' sole responsibility of any plagiarism.

Disclaimer

Articles and views published in this journal DO NOT necessarily reflect the views or policies of the editorial board or the journal. The authors will be solely responsible for the legal complications arising out of the contents of their articles or research papers.

Subscription Rates		
Period	Individuals	Institutions
One Year	\$20	\$50
Three Years	\$50	\$120
Five Years	\$80	\$200

Contents

Highlighting EFL Instructors' Methods of Assessing Students and Providing Feedback at the University of Aden.....	1
Asst. Prof. Dr. Shafiqah Anwar Fakir	
A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with British National Corpus (BNC).....	18
Asst.Prof. Dr. Abdunasser Mohammed Ali	
Asst. Prof. Dr. Adel Abdulkhaliq Abdulkareem	
Investigating Gender Bias in Crescent English Course for Yemen- Pupil's Book 2.....	39
PhD Candidate Intisar Saleh Ahmed Albundoq	

Issue No.3, Jan. 2017: Dr. Abdunasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with British National Corpus (BNC)

Asst.Prof. Dr. Abdunasser Mohammed Ali

University of Aden

Asst. Prof. Dr. Adel Abdulkhaliq Abdulkareem

University of Aden

Abstract

Corpus linguistics has opened a new window for course books evaluation especially in word counts, word repetition, word frequency, readability and many other statistics. The present study undertook the analysis of the Crescent English Course for Yemen (CECY) in terms of word counts, word repetition and word frequency in comparison with the most frequent words in the British National Corpus (BNC). The AntConc 3.2.4w software and the website "Compleat Lexical Tutor, v.4" (<http://www.lextutor.ca>) were used as the main tools for corpus analysis. The results of the analysis showed that Crescent English Course for Yemen (CECY) books 4, 5 and 6 do not cover the most frequent words found in the BNC. The percentage was less than 60%. In terms of word count, the analysis showed that the course book No.4 contains 1333 word families in a 2208-word type, course book No.5 contains 1388 word families in a 2237-word type course book No.5 contains 1450 word families in a 2354-word type and regarding all course books together, the analysis clearly revealed that course books No.4, 5 and 6 contains 2262 word families in a 4193-word type. The analysis also demonstrated that one timer frequent words is the highest. It is counted as 47.34%, 46.21% and 47.92 % in the three course books respectively. It showed that 41.5% is the percentage of one timers frequent words in the three course books together. The percentage decreased gradually in two timers, three timers and on. The proportion of the six timers and more is only 14.4 %, 18.48% and 16.74% respectively. The study concluded with recommendations and implications for further research.

Key words: Crescent English Course for Yemen (CECY), frequency, BNC, type, token and word family, course book.

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

Introduction

It has been widely accepted that vocabulary knowledge represents the backbone for language proficiency. Laufer (1986) advocates that "without adequate lexis there is no proper language competence or performance" (p.70). Course books in EFL classrooms represent the main source for learner's vocabularies especially in countries where English is taught as a foreign language. Consequently, course books vocabulary and frequency should be evaluated carefully. The Crescent English Course for Yemen (CECY) used in Yemeni secondary schools started in the school year 1993/1994 in the 7th class and continued to be introduced gradually, and now there is a need to have accurate statistics about it in terms of total number of words in each level, common words between these levels, different words, and frequencies and whether the words are among the most common words according to the BNC. In other words, there might be few number of empirical studies related to CECY in terms of vocabulary statistics or corpus based studies.

Currently computer technology has become the corner stone of most sciences and disciplines. It has revolutionized and sped up the steps of developments of most disciplines. Applied Linguistics is not a way of such impact. This impact started in the mid 1960s by the appearance of corpus linguistics. The advancements in corpus linguistics tools have permitted greater analysis of textual data especially word frequencies. The Crescent English Course for Yemen (CECY) used in Yemeni secondary schools' coverage of the most frequent words found in the BNC is inadequate and limited leading to low outcomes of English language learning in Yemen. This study uses the AntConc toolkit and the Compleat Lexical Tutor website to evaluate the Crescent English Course for Yemen (CECY) in terms of frequencies and compare it with the most frequent words in a famous corpus in the field of corpus linguistics, The British National Corpus (BNC).

Objectives of the study

This study aims at analyzing the Crescent English Course for Yemen (course books 4, 5 and 6) in terms of word frequency compared to the most frequent words in the BNC, size of words and how they are distributed in the three levels and lexical variation.

Issue No.3, Jan. 2017: Dr. Abdunasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

Statement of the research problem

The number of words as well as the frequency of the words in any EFL textbook represents a very important aspect in terms of teaching, textbook selection and learning. EFL literature emphasizes the importance of size and frequency aspect in textbooks. Kirsner (1994) showed that there are strong effects of word frequency on the speed and accuracy of lexical recognition processes (speech perception, reading, object naming, and sign perception) and lexical production processes (speaking, typing, writing, and signing), in children and adults as well as in L1 and L2. In the Yemeni context, Crescent English Course for Yemen (CECY) is used in teaching English in Yemeni elementary and secondary schools since 1994. Many studies have been conducted to this course but no one has evaluated it in terms of frequencies of words, chunks and collocations in comparison to the frequencies in famous corpora such as BNC to identify its strengths and weaknesses in terms of preparing Yemeni students to use English language. This study will try to identify the size and the most frequent words in the Crescent English Course for Yemen (CECY) in comparison to the frequencies in BNC.

Research Questions

This research paper tries to answer the following questions:

1. To what extent does the Crescent English Course for Yemen used in Yemeni Secondary Schools cover the most frequent words of English in comparison to the most frequent words found in BNC?
2. How many words are included in the Crescent English Course for Yemen? And how are they distributed in the different levels?
3. What are lexical variety and lexical density of the course books of the CECY?

Significance of the study

This study tries to present a better understanding of the Crescent English Course for Yemen (CECY) used in secondary schools in the light of words frequencies and word statistics. The significance can be materialized in the following points:

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

1. In evaluating the Crescent English Course for Yemen (CECY), the study helps Yemeni English teachers to have a clear picture of the course in terms of vocabularies and their frequencies. Ellis (1997) declares that textbook evaluation helps teachers move beyond impressionistic assessments and it helps them to acquire useful, accurate, systematic, and contextual insights into the overall nature of textbook material. Also Sheldon (1988) shows the importance of course books by saying that they represent "the visible heart" of any ELT program (p.237). Rahimpour (2013) also suggests that "[i]n order to use the textbooks effectively, it is essential for the practitioners to evaluate the materials since evaluation plays a key role in education and can provide valuable information" (p.766).
2. Course books evaluation helps in better decision making regarding the course books under use in terms of continuing using them or choosing new ones. This study might present insights and statistics on the vocabulary content of the course books under use that would help in decision making regarding the use of this course book.
3. The statistics of this study might serve as a reference to the designers of the course book to reconsider this course book or any new course books.
4. Teachers and learners will know how many words are included in the course and the difference between a level and the next to help them make clear vision regarding teaching and learning. Teachers will know what words should be taught and even what words should be used in tests and examinations and what pedagogical practices are suitable or what activities level can be used to consolidate learning. Teachers and learners will know the size of words that should be taught or learned a day. Knowing the low frequent words in the course book will help teachers to pay much attention and focus to those words in their teaching.
5. It might help in decisions regarding designing examinations and tests.
6. It might help teachers and learners to focus on the most frequent words in their teaching.

Issue No.3, Jan. 2017: Dr. Abdunasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

7. Evaluate the suitability of words in the course books for Yemeni learners in terms of adequacy for preparing students to use English and for undergraduate studies.

Methodology

In what follows a detailed description of the methodology of the study will be given. This will include preparation of the corpus (CECY), the software and the BNC.

Preparing the corpus and procedures

The corpus of this study consists of 6793 words which is the total number of words in the pupils' books of the secondary schools in Yemen (Crescent English Course for Yemen, CECY, 4, 5 and 6). The text books of all these levels were downloaded as a pdf format files from the website of the ministry of education <http://www.yemenmoe.net> and firstly converted into word document file then converted into text file format in order to be compatible with the software (AntConc) and the website Compleat Lexical Tutor, v.4 (<http://www.lextutor.ca>) used in the study. For the purpose of processing, only the actual words in lessons, the words in the introductory sections and appendices were removed and only lessons were processed. Some words are presented in a picture form in the course books that cannot be converted into word document file. These words are retyped manually and included in the corpus. Figures are replaced by the word number, contractions are replaced by full words, single letters are removed except 'a' and 'I' from the texts before the analysis process is implemented.

The Crescent English Course for Yemen (CECY)

The Crescent English Course for Yemen (CECY) was prepared by two British authors, Terry O'Neill and Peter Snow and was published by Garner Publishing Limited Oxford University Press for Arab World. It consists of six levels. The Crescent English Course for Yemen (CECY) 1,2,3,4,5 and 6 for the grades 7th, 8th and 9th in the elementary school and 1st, 2nd and 3rd grades or 10th ,11th and 12th grades of the secondary school. Each level text book consists of pupils' book and workbook. Pupils' books of the 1st, 2nd and 3rd grades or the 10th, 11th and 12th grades of the secondary school are only the main concern of this study.

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

The software

This study has used the free corpus analysis software AntConc 3.2.4w. This free toolkit software was first released by Laurence Anthony in 2002. It can run on any computer using Microsoft Windows, Macintosh OS X and Linux. It is available in different versions in the following <http://www.laurenceanthony.net/software/antconc>. This toolkit software processes txt, html, htm and xml file format and performs word frequency (wordlist, word cluster, word collocations, concordance, concordance plot, and key word list). This software is used to find out frequencies of types, types and word families of the corpus under study.

The Compleat Lexical Tutor, v.4

The Compleat Lexical Tutor, v.4 (<http://www.lextutor.ca>) is a free website designed by Tom Cobb to create concordancers, word profile, frequencies, comparing texts, making exercises and reading text and many other corpus analysis services and tools. It presents resources for learning, teaching and researching in the three main sections, learners, researchers and teachers. This website is used to draw the comparison between the corpus under study and the BNC and COCA. This website is used for drawing the comparison with the most frequent words of the BNC. This website is used on the 20th of January 2016.

The BNC

The British National Corpus (BNC) is a synchronic, monolingual, general and sample corpus containing 100 million of written and spoken words from different disciplines used in natural language use. The collection of the corpus began in 1991 and finished in 1994. The latest version of the BNC was released in 2007. It is an online free corpus available for learners, teachers and researchers.

Limitation of the study

This study is limited to evaluating only word frequency of the Crescent English Course for Yemen (CECY) used in secondary schools in comparison to the frequencies of the most frequent words in BNC. The study excluded the supplementary readers (science and literary readers).

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

Literature review

A review of the related literature is given below. This includes word frequency and research, previous studies related to the research topic.

Word frequency and research

Word frequency means the times a word appears in a text or a corpus. This occurrence or repetition in a text shows its importance and centrality in the text. In another simpler way, high frequency words are those words that most people use them frequently in their language. The words "the" and "and" are examples of high frequent words that appear in most English conversations or written texts. Frequency depends on the discipline or topic. A frequent word in a discipline might be a low frequent in another but there are words that are very frequent in most disciplines. The British National Corpus (BNC) and The Corpus of Contemporary American English (COCA) present the high frequent words in English among millions of words used in English. The table below shows the top ten frequent words in English among 450 million words in BNC and COCA.

Table No.1: The top 10 frequent words in BNC and COCA (20.04.2015).

<i>Rank</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
<i>BNC</i>	<i>The</i>	<i>Of</i>	<i>And</i>	<i>To</i>	<i>A</i>	<i>In</i>	<i>That</i>	<i>It</i>	<i>Is</i>	<i>Was</i>
<i>COCA</i>	<i>The</i>	<i>And</i>	<i>Of</i>	<i>To</i>	<i>A</i>	<i>In</i>	<i>That</i>	<i>I</i>	<i>It</i>	<i>Is</i>

Course books, teachers and even learners should pay much attention to this high frequency words because they are almost found in most texts and form a basis for language use and understanding. Nation (2014) explains the value of the top frequent words:

The ten most frequent words of English typically cover 25% of the words in any text and the 100 most frequent words cover around 50%. Have a look at a page and see how often the word occurs (It occurs in almost every line in this paragraph!). By itself, that word covers 7% of any written English text. The

Issue No.3, Jan. 2017: Dr. Abdulnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

most frequent 1,000 words cover around 80% of the words in most texts. In some languages the coverage figures are even higher than this. (p.13)

Many studies show the importance of high frequent words for language learning. For example, Schonell et al (1956) found that the most frequent 1,000 words in spoken English provide coverage of 94% of the running words in informal conversation. Similarly, Carroll et al. (1971) found that the first 1,000 words of English cover 74% of written text. (Nation, 2007, Nation & Waring, 1997) found that the most frequent 2,000 word families (all the forms of a given word) enable readers to comprehend approximately 80% of the average novel or approximately 83% of the average newspaper article. Nation (2001a) refers to the size of high frequent words needed for secondary school learners to move to the university level by saying that "[t] [he role of words in second language learning evidence that counting the 2,000 most frequent words of English as the high frequency words is still the best decision for learners going on to academic study"(p.15). However, Nation (2001b,p.11) distinguishes four types of words in a text: high-frequency words, academic words, technical words, and low-frequency words. (Nation, 2001a, p.17) presents a distribution of commonly used words in different text types:

Table No.2: distribution of commonly used words in different text types				
levels	conversation	fiction	Newspaper	Academic text
1st 1000	84.3%	82.3%	75.6%	73.7%
2nd 1000	6.0%	5.1%	4.7%	4.6%
AWL	1.9%	1.7%	3.9%	8.5%
Other	7.3%	10.9%	15.7%	13.3%

Other studies that show the importance of high frequency words in learning and acquisition are (Balota & Chumbly,1984; Forster & Chambers,1973). These studies found that high- frequency words are learned more rapidly than low-frequency ones. Barry and Seymour, (1988) found that high frequency words are spelled more accurately. Luce, (1986)

Issue No.3, Jan. 2017: Dr. Abdunasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

and Savin, (1963) also found that high frequency words are better for auditory word recognition. Similarly, Kirsner (1994) showed that there are strong effects of word frequency on the speed and accuracy of lexical recognition processes (speech perception, reading, object naming, and sign perception) and lexical production processes (speaking, typing, writing, and signing), in children and adults as well as in L1 and L2.

Regarding language pedagogy, word frequency comes in the heart of course books selection, course books evaluation, teaching materials and language testing. For example, Biber and Reppen (2002) concluded that frequency information should not be excluded in designing teaching materials:

Given its importance in acquisition, we would argue that frequency should also play a key role in the development of materials and in the choices that teachers make in language classrooms. With the recent availability of comprehensive frequency-based grammatical descriptions, such integration of pedagogy and research has become feasible. (p. 207-208)

Based on the theories and findings, White (1998, pp.48-50) proposed seven elements in deciding the words taught or introduced first or earlier: (1) frequency, (2) coverage, (3) range, (4) availability, (5) learnability, (6) opportunism, (7) center of interest. Research in second language learning shows the value of high frequency words. Lotto and De Groot (1998) study manifested that high-frequency words are easier to learn and retrieve than low-frequency words.

Considering the number of words that L2 learners need to learn, it is better to know how many words a native speaker knows. Zechmeister et al. (1995) and Nation and Waring, (1997) point out that an educated native speaker's words range between 17,000 and 20,000 word families. Similarly, Goulden et al. (1990) speaking about English native speakers contend that "well-educated adult native speakers of English have a vocabulary of around 17,000 base words" (p.341). Regarding children, Nation and Waring (1997, p7) indicate that native speakers five years-old children start school with an amount of 4,000-5,000 words and increase this amount by 1000 words a year and graduate from university with about 20,000 word families. Regarding L2 learner amount of words, Francis and Kucera (1982) refer that

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

L2 needs 2,000 word families to produce and understand L2. Laufer (1992) and Nation (2001) estimate the amount to 3,000 word families. In an aspect Sánchez (2009). assumed that "the most frequent 800-1,000 words were to be learnt by elementary level students, while intermediate and advanced students would be presented the next 800-1,000 or 2,000 words in the frequency list"(p.865). Laufer (1998) contends that knowing less than 3000 word families results in weak reading ability.

Several studies were done regarding vocabulary analysis in terms of frequencies and repetition. Davis and Face (2006) who studied the words frequency in Spanish textbooks in relation to the most frequent 1000 words. They found that only 10-50% of the words in textbooks are among the most frequent 2000 lemma in the language. Jimenez and Macebo (2008) studied the EFL textbooks words for secondary and elementary schools in relation to the most frequent fifty words. Ito and Bauman (1995) studied the rate of words learned by learners in Japan. They found that learners learn only one word per hour. Another similar study is by Waring and Takaki (2003) who found that students learn the average of 3.6 words per class hour in a three-month period.

Data analysis and discussion

In what follows are the analysis and discussion of the results found in the corpus understudy. The following tables show the results of the analysis and a discussion of the results will be given below the table.

Table No.3: Course Book No.4			
Freq. level	Families (%)	Types (%)	Tokens (%)
K-1 Word	685 (51.39)	1196 (54.17)	9332(81.63)
K-2 Words	291 (21.83)	387 (17.53)	802(7.02)
K-3 Words	129 (9.68)	155 (7.02)	303(2.65)
K-4-K-20	228 (17.25)	255 (11.36)	464 (3.88)

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

Off-List:	??	215 (9.74)	531 (4.64)
Total	1333+?	2208 (100)	11432 (100)

Table No.4: Course Book No.5			
Freq. level	Families (%)	Types (%)	Tokens (%)
K-1 Words:	737 (53.10)	1292 (57.76)	11763(84.66)
K-2 Words:	294 (21.18)	397 (17.75)	901(6.48)
K-3 Words:	130 (9.37)	149 (6.66)	264(1.90)
K-4-K-20	591 (10.36)	255 (11.84)	535 (3.85)
Off-List:	??	134 (5.99)	432 (3.11)
Total	1388+?	2237 (100)	13895 (100)

Table No.5: Course Book No.6			
Freq. level	Families (%)	Types (%)	Tokens (%)
K-1 Words:	742 (51.17)	1315 (55.86)	11979(83.63)
K-2 Words:	342 (23.59)	467 (19.84)	1033(7.21)
K-3 Words:	148 (10.21)	183 (7.77)	352(2.46)
K-4-K-20	218 (15.3)	262 (11.13)	662 (3.9)
Off-List:	??	127 (5.40)	298 (2.08)
Total	1450+?	2354 (100)	14324 (100)

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

Table No.6: Course Books No.4,5 and 6			
Freq. level	Families (%)	Types (%)	Tokens (%)
K-1 Words:	882 (38.99)	1908 (45.50)	33074 (83.41)
K-2 Words:	545 (24.9)	817 (20.20)	2736 (6.90)
K-3 Words:	290 (12.82)	375 (8.94)	919 (2.32)
K-4-K-20	545 (23.29)	1063 (25.36)	2921 (7.37)
Off-List:	??	413 (9.85)	1260 (3.18)
Total	2262+?	4193 (100)	39650 (100)

Table No.7: Frequency distribution						
Course book	1	2	3	4	5	6 and more
	Timers %	timers %	Timers %	timers %	timers %	Timers %
Book 4	47.34	17.56	10.02	6.29	3.85	14.4
Book 5	46.21	16.26	9.15	5.92	3.88	18.48
Book 6	47.92	16.96	9.14	5.87	3.37	16.74
Books 4,5,6	41.5	16.61	8.28	6.47	4.57	22.57

The results shown in the tables above, show in column one the frequency level (K-1, K-2, K-3 and K-4-K20) meaning the first most frequent 1000 words, the second most frequent 1000 words, the third most frequent 1000 words, the fourth–twentieth most frequent 1000

Issue No.3, Jan. 2017: Dr. Abdunasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

words in the BNC. Column two shows the word family percentage in the course book in relation to the BNC. Column two shows the word family percentage. Word family is the word and its inflected and derived forms. Column three shows the percentage of types. Type is a word that is counted only once even if repeated many times in the corpus. The fourth column shows the percentage of tokens. Token is the word that is counted every time it appears in the corpus. *It means what it means but I mean my meaning.* This sentence has one-word family which is *mean* and seven-word type (it, mean, what, but, I, my, meaning) and ten tokens (all words).

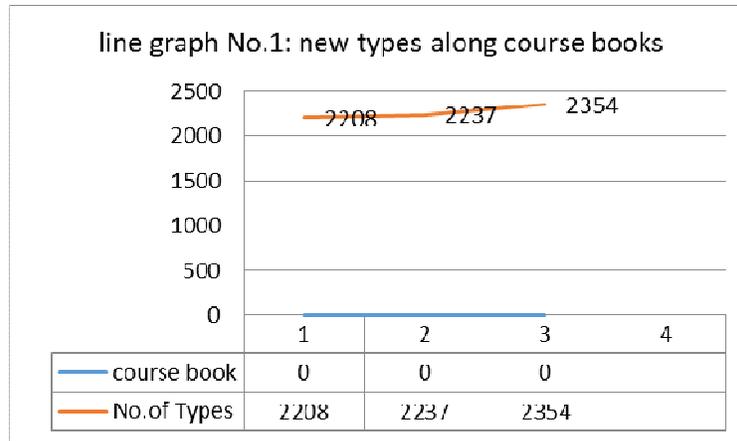
Regarding the frequency in comparison to the BNC, the analysis of the course books shows that almost half of the word family is only appeared among the most frequent 1000 words in the BNC. This is counted as 51.39%, 53.10% and 51.17% in the course books 4, 5 and 6 respectively. The percentage decreased gradually regarding the second 1000 and on. The percentage of the types of the course books corpus does not exceed 60% of the most frequent 1000 words in the BNC. This result is surprising because it is likely to be insufficient for reading improvement and proficiency. The first most frequent 1000 words is very essential for reading comprehension and conversation as well. Studies such as Schonell et al. (1956) found that the most frequent 1,000 words in spoken English provide coverage of 94% of the running words in informal conversation. Other studies such as Carroll et al. (1971) and (Nation, 2001) and (Balota & Chumbly, 1984; Forster & Chambers, 1973) have similar results. This leads that students will not be able to pass the “all-or-nothing” threshold (80% of known words in a certain text) suggested by Nation (2001) while reading or speaking. Hiebert (2005) refers that the 100 most frequent words account for about 50% of the words in a typical text; the 1,000 most frequent words for about 70%; and the 5,000 most frequent words for about 80%. This clarifies that the Crescent English Course for Yemen (CECY) does not fully equip students with familiar words needed for daily life usage.

In respect of frequency distribution shown in table No.7, one timer frequent words is the highest. It is counted as 47.34%, 46.21% and 47.92 % in the three course books respectively. It also shows that 41.5% is the percentage of one timers frequent words in the three course books together. The percentage decreased gradually in two timers, three timers and on. The

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

proportion of the six timers and more is only 14.4 %, 18.48% and 16.74% respectively. And only 22.57% of the six timers and more in the three course books together. This is another surprising results. This shows that the course books lack high frequent words. This indicates that the frequency of exposure is too low to be learned easily and not sufficient for quick and smooth learning. Humans learn through exposure to input (Krashen Input Theory) and frequency is an important element of exposure in foreign language learning. Low frequent words cause reading difficulty, reading comprehension and inadequate learning because of lack of commonness and familiarity. A word is learned faster and effortlessly if it is repeated many times in the course books. Studies found that high- frequency words are learned more rapidly than low-frequency ones. A word should appear at least 7-8 times (extreme opinions require 5-16 occurrences) for students to remember them, (Oxford and Scarcella 1994.). Barry and Seymour, (1988) found that high frequency words are spelled more accurately. Similarly, Luce (1986) and Savin, (1963) found that high frequency words are better for auditory word recognition. Also, Kirsner (1994) has shown that there are strong effects of word frequency on the speed and accuracy of lexical recognition processes (speech perception, reading, object naming, and sign perception) and lexical production processes (speaking, typing, writing, and signing), in children and adults as well as in L1 and L2. Nakata (2006) acknowledged that vocabulary acquisition requires continual repetition in order for effective vocabulary learning (p. 19). Also Freebody and Anderson (1983) in an experiment found that low frequency produces negative effect on comprehension. Research in second language learning shows the value of high frequency words. Lotto and DeGroot's (1998) study manifested that high-frequency words are easier to learn and retrieve than low-frequency words.

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)



The line graph above, clearly shows a very slight increase of new types or words from course book 4 to course book 5 and to course book 6. Only 29 new types were added to course book 5 which is considered very low. In turn, 146 new types were the difference between course book 5 and 6. This is considered a reasonable increase of new types. A total of 175 new types is the difference among the three course books which represent a very low amount regarding three years of study difference. This indicates an unbalance of word increase among the three course books which negatively affects the size of words needed for communicative potential of learned words. Students finishing course books 6 should go to university level which requires a larger stock of vocabulary. Studies show that 3000-word family is necessary for starting academic study in the university level. Laufer (1998) contends that knowing less than 3000 word families results in weak reading ability. Some other studies talk about a bigger number of 3000-word family that is needed for graduate study. The three course books all together don't reach the 3000-word family. The analysis clearly shows that course book No.4 contains 1333 word families in a 2208-word type, course book No.5 contains 1388 word families in a 2237-word type course book No.6 contains 1450 word families in a 2354-word type and regarding all course books together, the analysis clearly shows that course books No.4, 5 and 6 contains 2262 word families in a 4193-word type. This word size in the three course books seems not enough to prepare students for university study.

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

Lexical variation or lexical density is the ratio between the number of types and the number of tokens in a corpus. Considering the lexical variation of the course books, the following formula is used:

$$LV = \frac{\text{total number of types} \times 100}{\text{total number of tokens}}$$

Applying the formula to course book No.4: $LV = \frac{2208 \times 100}{11432} = 19.31\%$. Book No.5:

$LV = \frac{2237 \times 100}{13895} = 16.09\%$. book No.6: $LV = \frac{2354 \times 100}{14324} = 16.43\%$. And considering all course

books (4,5 & 6) together as one corpus: $LV = \frac{4193 \times 100}{39650} = 10.57\%$. This indicates that the type

-token ratio is ranges between 16% to 19.31% in the three course books separately and the proportion decreases to 10.57% in the three course books together. Which is not a higher one. This shows a high repetition because the number of word families and types is low in relation to three course books for three years of study. Further, the repetition is found high in function words rather than content words. Another reason for this is the analysis processed the course books as whole book not as a text. This leads that students will not be able to pass the “all-or-nothing” threshold (80% of known words in a certain text) suggested by Nation (2001).

Conclusions and recommendations

To conclude in respect to the research objectives, the analysis has shown that the Crescent English Course for Yemen (CECY) (books 4,5 & 6) used in Yemeni secondary schools provides inadequate frequent words since only almost half of the word family is only appeared among the most frequent 1000 words in the BNC. The proportion is counted as 51.39%, 53.10% and 51.17% in the course books 4,5 and 6 respectively. The percentage decreased gradually regarding the second 1000 and on. The percentage of the types of the course books corpus does not exceed 60% of the most frequent 1000 words in the BNC. The proportion of the types in the (CECY) does not exceed 60% of the most frequent types in the

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

BNC. This result represents a defect of the (CECY) because it fails to provide the most frequent words used by native speakers.

With respect to frequency distribution in the course books, the analysis has shown that one timers frequent words is the highest and the proportion decreased gradually in tow timers and on reaching only less than 20% in the six timers frequent words. This is another defect of the (CECY) because it demonstrates little repetition or recycling of the words in the course books.

Considering the number of words added from one level to another, the analysis also has shown that is not high resulting in a small vocabulary size. It showed that the vocabulary size is only 4193 which is not enough to prepare students to undergraduate studies.

Recommendations

Based on the results of the study, the following recommendations are forwarded to the authorities in charge of teaching English in Yemeni secondary schools in particular the ministry of education.

- ✓ Updating or substituting the Crescent English Course for Yemen (CECY) with another course that presents the most frequent words in English that help in improving Yemeni secondary schools students' proficiency of English.
- ✓ Including corpus linguistics specialists in committees responsible of selecting English course books.
- ✓ Doing more research on the Crescent English Course for Yemen (CECY).
- ✓ Holding an evaluation workshop on The Crescent English Course for Yemen (CECY) including different relevant specialists.
- ✓ Teaching English should start from an earlier stage in order to increase the number of words of Yemeni secondary school student that might help and prepare them for undergraduate studies.

Implications for further studies

Corpus based studies open up many future research. The corpus analyzed in this study should be focused upon in a more quantitative or qualitative fashion. A study might be

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

undertaken to know the most frequent collocations, grammatical structures. Subsequent research could analyze the course books with their supplementary sections and work books. A comparison and contrast of the corpus under study with other course books in any other country. In terms of course book evaluation, many other aspects can be undertaken such as design and layout and language type.

Issue No.3, Jan. 2017: Dr. Abdunnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

References

- Balota, D. A., & Chumbley, J.L. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10,340–357.
- Barry, C., & Seymour, P.H.K. (1988). Lexical priming and sound-to-spelling contingency effects in non-word spelling. *Quarterly Journal of Experimental Psychology*, 40(A), 5–40.
- Biber, D. & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24, 199-208.
- Carroll, J.B., Davies, P. and. Richman, B (1971). *The American Heritage Word Frequency Book*. New York: American Heritage Publishing Co.
- Cobb, T. *Web Vocab profile* [accessed 28 Dec. 2015 from <http://www.lex tutor.ca/vp/>], an adaptation of Heatley, Nation & Coxhead's (2002) *Range*.
- Davies, M. & T. L. Face. (2006). *Vocabulary Coverage in Spanish Textbooks: How Representative is it?* Selected Proceedings of the 9th Hispanic Linguistics Symposium. Eds. N. Sagarra & A. J. Toribio. Somerville, MA: Cascadilla Proceedings Project. 132-143.
- Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal*, 51(1), 36-42 <http://dx.doi.org/10.1093/elt/51.1.36>
- Forster, K., & Chambers, S. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12,627–635.
- Francis, W. & Kucera, H. (1982). *Frequency analysis of English usage*. Houghton Mifflin Company, Boston.
- Freebody, P., & Anderson, R. C. (1983). Effects on text comprehension of different proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, 15, 19-39.

Issue No.3, Jan. 2017: Dr. Abdunnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

- Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics 11*, 341-362.
- Ito, T. & J. Bauman. (1995). *The Acquisition of Basic Vocabulary by College-Age Japanese Students in an Intensive EAP Program*. Paper given at the 34th Annual Meeting of the Japan Association of College English Teachers, Tokyo.
- Kirsner, K. (1994). Implicit processes in second language learning. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp.283–312). San Diego, CA: Academic Press.
- Laufer, B. (1986). Possible changes in attitude towards vocabulary acquisition research. *IRAL*, 24(1), 69-75.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud and H. Béjoint (Eds.). *Vocabulary and Applied Linguistics*, (pp.126-132). London: MacMillan.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied Linguistics*, 19, 255–271.
- Lotto, L. & De Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, 48 (1), 31-39.
- Luce, P.A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Per-Caption and Psychophysics*, 39,155–158.
- Nakata, T. (2006). Implementing optimal spaced learning for English vocabulary learning: Towards improvement of the low-first method derived from the re-activation theory. *The JALT CALL Journal*, 2(2), 3-18.
- Nation, P., (2001a). *Learning vocabulary in another language*. Cambridge: CUP.
- Nation, P. (2001b). How many high frequency words are there in English? In M. Gill, A.W. Johnson, L.M.
- Nation, P. (2007). *How large do learners' vocabularies have to be? Presentation handout for JALT (Japan Association of Language Teachers) National Conference, Tokyo.*

Issue No.3, Jan. 2017: Dr. Abdalnasser Mohammed Ali & Dr. Adel Abdulkhaliq Abdulkareem: A Corpus Based Study of the Secondary School Crescent Course Book Vocabulary in Comparison with the British National Corpus (BNC)

Nation, P. (2014). *What do you need to know to learn a foreign language?* New Zealand: Victoria University of Wellington.

Nation, P., & Waring, R. (1997). *Vocabulary size, text coverage & word lists*. Retrieved on December 12, 2007, from <http://www1.harenet.ne.jp/~waring/papers/cup.html>

Oxford, R. L. & Scarcella, R. C. (1994). Second language vocabulary learning among adults: State of the art in vocabulary instruction. *System*, 22, No.2.

Rahimpour, S. (2013). *The textbook evaluation. Proceeding of the Global Summit on Education 2013* (e-ISBN 978-967-11768-0-1) 11-12 March 2013, Kuala Lumpur. Organized by WorldConferences.net.

Sánchez, A. (2009). *Vocabulary in EFL Textbooks. A Contrastive Analysis against Three Corpus-Based Word Ranges*.862-875.

Savin, H.B. (1963). Word-frequency effects and errors in the perception of speech. *Journal of the Acoustic Society of America*, 35,200–206.

Schonell, F. J., Meddleton, I.G., & Shaw, B.A. (1956). *A study of the oral vocabulary of adults*. Brisbane: University of Queensland Press.

Sheldon, L. (1988). Evaluating ELT textbooks and materials. *ELT Journal*, 42/2.

Waring, R. & M. Takaki. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language* 15 (2): 130-163.

White, R. (1998). *The ELT curriculum*. Oxford: Basil Blackwell.

Zechmeister, E. B., Chronis, A. M., Cull, W. L., D'Anna, C. A. & Healy, N.A. (1995) Growth of a functionally important lexicon. *Journal of Reading Behavior*, 27(2),