Gönderim Tarihi: 12.06.2018

Kabul Tarihi: 05.10.2018

A CORPUS ANALYSIS OF THE EIGHTH GRADERS' TEXTBOOK

Sekizinci Sınıf İngilizce Ders Kitabı Üzerine Derlem Temelli Çalışma

Nalan KIZILTAN

Prof. Dr. Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi
İngilizce Öğretmenliği Bölümü
kiziltan@omu.edu.tr

ORCID ID: 0000-0002-7427-363X

Doğan BAYDAL

Öğr. Gör. Ondokuz Mayıs Üniversitesi,
Yabancı Diller Yüksekokulu
Mütercim-Tercümanlık (İngilizce) Bölümü
doganbaydal@omu.edu.tr

ORCID ID: 0000-0001-9392-4779

Çalışmanın Türü: Araştırma

Öz

Yabancı dil ya da ikinci dil öğretiminde hedef dildeki derlemin dikkate alınması gerektiğinden dolayı, yabancı dil ders kitaplarında sözcük kullanımı yabancı dil öğrenenler için önem arz etmektedir. Belirlenen hedeflere ulaşmada dil öğrenenlere yardım etmek için; yabancı dil ders kitaplarının hedef dil derlemi dikkate alınarak hazırlanmış olması beklenmektedir. Bu çalışmanın amacı ilköğretim 8. sınıf İngilizce dersi için hazırlanan ders kitabındaki sözcüklerin hedef dil derlemi açısından uygululuğunu tartışmaktır. Aynı zamanda kitaptaki sözcüklerin Avrupa Dilleri Ortak Çerçeve Programındaki sözcük seçimi kriterlerine uygunluğu tartışılmıştır. Bu çalışmadaki araştırma soruları şunlardır: 1. Ders kitabında en sık kullanılan sözcükler nelerdir? 2. İngilizce 'de en sık kullanılan 1000 sözcük, en sık kullanılan ikinci 1000 sözcük ve akademik sözcükler ders kitabında nasıl kullanılmıştır? 3. Ders kitabında kullanılan sözcükler, Avrupa Dilleri Ortak Çerçeve Programı A2 düzeyine uygun mudur? Bu çalışma yabancı dil öğretmenleri ve ders kitabı yazarlarına A2 düzeyine uygun sözcük seçiminde faydalı olacaktır.

Anahtar Kelimeler: Derlem, sözcük, sözcük öğretimi, ders kitabı

Abstract

Vocabulary use in textbooks for foreign language learners is significant since national language corpus must be taken into account in teaching a language as a foreign or a second language. Textbooks are supposed to have been prepared by the corpus to help the foreign language learners to achieve desired level. This study aims at discussing the corpus in a textbook prepared for the eighth graders in Turkey. Besides it discusses how well it coincides with Common European Framework of Reference for Languages (CEFR). This study is based on the following questions: 1. What are the most frequent words in the textbook? 2. How well are K1 Words, K2 Words and Academic Words used in the textbook? 3. Does the corpus in the textbook coincide with A2 Level according to CEFR? The study will help not only foreign language learners but also school teachers and textbook writers in order to know and use basic vocabulary required for A2 Level according to the criteria of CEFR.

Keywords: Corpus, lexicon, vocabulary teaching, textbook

1. INTRODUCTION

The value of vocabulary in language teaching and learning is very important. Laufer (1986) describes that if the learner does not have the knowledge of the vocabulary, learning a language cannot be achieved. Laufer (1986) further adds that vocabulary has not been given importance by the practitioners of English Language Teaching (ELT) up to now. Along with the growth of Communicative Language Teaching (CLT), accuracy on grammatical structures lost its importance and communicative competency was favoured, which prioritize the importance of vocabulary.

Lexical syllabus has come into existence as the natural outcome of the Communicative Language Teaching. According to lexical syllabus, the words with high level of frequency should be taught first because they occur with a high percentage in spoken and written English (Thurnbury, 2002). As well as lexical syllabus, the field of corpus linguistics has become popular because of the high frequent words which are measured by corpus software programmes. Word frequency is an important concept of corpus linguistics and it shows how many times a word or word phrase takes place in spoken and written language. Word frequency is a prerequisite of Communicative Language Teaching; it is because if the learner knows the most frequent words in the target language, s/he will be able to understand the meaning of the written and spoken language highly, which means that with a small number of words, fluency in speaking can be achieved. Furthermore, teachers and textbook writers can get benefit from word frequency in that the word frequency helps them to decide which words to be selected in foreign language teaching.

Nowadays, many textbooks focus attention on the importance of word frequency. A textbook Cutting Edge Intermediate (2005) emphasizes the importance of vocabulary by focusing on high frequent useful words and phrases. Then the most frequent words can be great helpers for English as a Foreign Language (EFL) teacher in order to teach learners the required vocabulary.

2. THEORETICAL FRAMEWORK

2.1. Corpus Studies

The compilation of written and spoken texts is called corpus. The data of the corpus are held in computer (O'Keeffe, McCarthy & Carter, 2007). Currently, the data of the corpus are uploaded on a website and it becomes available to be observed by the researchers. However, when the computers

were not that common as in today, the experts of the corpus linguistics used to observe the written materials and it was so hard to collect, organize and save the data for them. Nevertheless, the development of the technology made the computers to be available much more easily and as a result of this, different software programmes on corpus linguistics area were prepared by the specialists of linguistics. With the advent of these corpus software programmes, collecting, organizing, calculating and saving the language data have become simple and as well as experts of linguistics, all the people (teachers, students, etc.) have had the chance of accessing the data of any corpus.

It is important to state that a compilation of any language data cannot be considered as corpus because there should be some requirements that have to be fulfilled while creating a corpus. First, the corpus should be planned and the purpose of it should be for linguistics. Second, the corpus should be representative of the data to be analysed.

First, as Hunston (2002) states, planning, organization and linguistic purpose are the core elements of the corpus. These features differ the corpus from library; it is because the books and the data in the libraries are kept only to be read. Hunston (2002) further adds that it is important that the data of the corpus can be used for quantitative and qualitative analyses. For example, a researcher can compose a corpus which is taught in a secondary school. From this corpus, anyone can learn the total number of the most frequent words, which words collocate with each other mostly and so forth. This is called quantitative analysis. The researcher can also compare the corpus of a textbook with the reference corpus (e.g. British National Corpus) to make judgements about it. This observation is called qualitative analysis.

Second representativeness is another requirement a corpus should have because the linguists judge the usefulness of a corpus by looking at how representative it is. As O'Keeffe, McCarthy and Carter (2007) state, if you want to create a classroom corpus, the corpus should represent all the variables related to classroom such as: Which type of school is it? What qualifications does the teacher have? (experience, education, gender and so forth) What are the genders and ages of the students? What is the number of the students in the classroom? What are the backgrounds of the students? The variables change according to the purpose of the corpus.

2.2. Corpus Techniques

There are two main techniques in corpus studies. These are frequency and concordance. Frequency is the most common corpus technique used by

the corpus linguistics specialists. The corpus software programmes count the words or the word phrases in any written or spoken text and a total calculation is reached. O’Keeffe, McCarthy and Carter (2007) state that through word frequency lists, it is possible to observe all the words in order of frequency. Frequency lists are effective in identifying probable differences between the corpora. Below are the most frequent ten words of different corpora. Concordance presents all of the occurrences of a specific word in the context it belongs to. The searched word is usually presented in the middle and there are a few words on the left and on the right of it (Bauer & Nation, 1993).

2.3. Corpus and Foreign Language Teaching

Throughout the years, linguistic analysis was the core purpose of the corpus studies. However, with the first half of the 1980s, a lot of researches have been made about corpus and foreign language teaching and learning. John Sinclair, who is a professor in English department, made the first important study in this field. He created the COBUILD Project (Sinclair, 1987) which was about running the corpus data in computers and working with its implementations in language teaching. Römer (2010) states that the aim of the COBUILD Project was to present useful dictionaries and genuine teaching materials to the learners of English. The COBUILD Project provides language samples which are not invented but authentic. Students face real conversation data and written texts of native speakers of English. Römer’s (2010) statements make it clear that COBUILD Project is composed of language samples which are very crucial in language teaching and learning. Apart from COBUILD Project, there are many other projects aimed to apply corpus studies in foreign language teaching and learning. English Language Teaching (ELT), however, seems to be unaffected by these developments and projects in the implementation of corpus studies in language teaching because few teachers have the knowledge of corpus tools, concordance programmes and computer software programmes about corpus.

The general and specialized corpora have many direct and indirect applications in ELT. Below is Figure 1, which shows the relationship between corpus and language teaching and learning.

Figure 1. The use of corpora in language learning and language teaching (Römer, 2010:9)

Figure 1 shows that direct applications of using corpora in language teaching and learning include the learner and the teacher who actively use corpora and concordance by means of hands on activities. The indirect applications, on the other hand, are related with researchers, material designers and syllabus makers. Römer (2010) states that direct applications of using corpora in language teaching seek the answer of how to teach while indirect applications seek the answer to the questions of what and when to teach.

3. THE ROLE OF VOCABULARY IN LANGUAGE TEACHING AND LEARNING

Vocabulary has an indispensable role in language learning (Evelyn and Brown, 1995; Waring and Nation; 1997; Schmitt, 2000; Thurnbury, 2002; Nation, 2006; Karatay, 2007; Nazary, 2008; Wallace, 2008; Moghadam et al, 2012; Akbari, 2012; Michael et al, 2012; Toprakoğlu and Dilman, 2015). However, the importance of vocabulary has not been realized for a long time. Language teaching approaches, such as Direct Method and Audiolingualism have given more importance to acquiring grammar. The practitioners of these approaches have stated that knowing the grammatical

approaches is the core element of language teaching and learning. Many former EFL textbooks were written to help the students to gain grammatical competence. Teachers were also in favour of teaching the grammar of the target language because they thought that grammar has a system of rules and it generated in a systematic way, which allowed the students to learn it easily. However, the practitioners of these language teaching methods, textbook writers, teachers etc. did not take into account the idea that the aim of the language learning was not to gain grammatical rules, but to communicate efficiently in the target language. With the advent of Communicative Language Teaching (CLT), communicative competence rather than grammatical knowledge gained importance. The most important concept of CLT, that is lexical syllabus, stressed the idea that vocabulary is the core element of communication and fluency rather than accuracy should be focused on. The supporters of CLT claimed that knowing grammatical rules of the target language does not mean anything if the learner cannot express himself/herself while, for example talking to a tourist. In this context, knowing the core vocabulary of “greeting and meeting” helps a lot and it shows that even with small number of words, the aim can be achieved.

It is understood that the role of vocabulary is seen very important in language teaching and learning and the acquisition of vocabulary should be prioritised for effective language teaching. There is a strong relationship between the acquisition of vocabulary and four skills of language (reading, listening, speaking and writing) as well as grammar.

3.1. Criteria for Vocabulary Selection

As it is seen, the role of vocabulary in language learning and teaching is crucial. For effective writing, speaking, listening and reading in any foreign language, the learner needs to have good vocabulary knowledge. Because of this importance of vocabulary, the vocabulary of the textbooks and teaching materials should be selected carefully. In this sense, vocabulary selection gains importance.

The students come across a lot of words in textbooks, teaching materials, reading texts and so forth. However, not all of these words are equally useful. Some of the words may be out of date, old-fashioned, belong to a particular dialect, be foreign word or so forth. Thus, it is not useful to learn every word that the students encounter. It is a heavy burden for the students. These kinds of vocabulary items do not help students in an immediate environment to express themselves in the target language. Schmitt (2000) states that if the student wished to survive in any short

holiday in abroad, it would be enough for him to know the basic vocabulary that he/she needs like numbers from 1-10, wh questions (what, where, when, why and how) and vocabularies that include activities such as buying, asking for direction, reading signs and so forth.

As it is seen, vocabulary selection should be done in a systematic way in language teaching. Schmitt (2000) states that there are four criteria for choosing vocabulary in language teaching. These criteria are vocabulary frequency, words particularly useful in a specific topic area (technical vocabulary), vocabulary that students want to learn and vocabulary necessary for classroom management. West (1953) proposes two criteria for vocabulary selection: common words and readability. White (1989), Richards (2001) and Genç (2004) talk about six criteria while choosing vocabulary: frequency, coverage, range, availability, defining power and learnability. Ünlü (2012), likewise, proposes three criteria for vocabulary selection. These are vocabulary frequency, range and vocabulary repetition.

It is seen that, in all of these criteria, vocabulary frequency has been seen as the most important one by the researchers.

3.2. CEFR A2 Waystage Level and Vocabulary

Milton and Alexiou (2009) state that although the earliest developments of the CEFR gave place to the vocabulary lists (for example, the Waystage Level by Van Ek and Trim, 1990), these lists have been removed with the compilation of CEFR in 2001. Skills-based rather than knowledge-based criteria were adopted by CEFR. (Council of Europe, 2001). Nevertheless, CEFR gives information about criteria for vocabulary selection (which words are the right words in foreign and second language teaching?) and ways to develop vocabulary. Council of Europe (2001) presents a number of options in vocabulary selection as follows:

- To select key words and phrases a) in thematic areas required for the achievement of communicative tasks relevant to learner needs, b) which embody cultural difference and/or significant values and beliefs shared by the social group(s) whose language is being learnt;
- To follow lexico-statistical principles selecting the highest frequency words in large general word-counts or those undertaken for restricted thematic areas;
- To select (authentic) spoken and written texts and learn/teach whatever words they contain;
- Not to pre-plan vocabulary development, but to allow it to develop organically in response to learner demand when engaged in communicative tasks (Council of Europe, 2001:150-151).

The options of the CEFR in vocabulary selection give priority to key words and phrases, highest frequency words and the words from authentic spoken and written texts. Thus it is seen that corpus studies play a crucial role in vocabulary selection according to the CEFR. Corpus techniques present the key words and phrases and the most frequent words with the help of the corpus software programmes. Moreover, general corpora like British National Corpus, Brown Corpus and Cambridge International Corpus have words from both written and spoken language data, which make these corpora the primary sources to apply in selection of vocabulary.

As well as wordlists, the vocabulary range in CEFR is missing in quantitative means. Milton and Alexiou (2009:196) say that “the word lists at Threshold Level (CEFR B1) contain about 2000 words and the word lists at Waystage Level (CEFR A2) contain about 1000 words”. It is understood from this statement that the first versions of CEFR had the number of vocabulary size and, however, with the recent compilation of CEFR, can-do descriptors and skills are concentrated and the number of vocabulary size is excluded.

The A2 Level vocabulary range of the CEFR gives great importance to communicative needs and shows that the learner needs to have a vocabulary to deal with the simple survival needs. Although the vocabulary range of the CEFR does not show the number of the vocabulary range in the levels, because of its validity as a hierarchy, it is understood that the learner at the A2 Waystage Level has more vocabulary than the learner at the A1 Breakthrough Level and the learner at the Proficient User Level has more vocabulary than the students at the other levels. Contrary to the general vocabulary range information in the CEFR, the researchers estimate that vocabulary range of the A2 Level is around 1500 and 2000 words (Van Ek & Trim, 1990; Meara & Milton, 2003; Milton & Alexiou, 2009). By looking at this number, it can be concluded that it is necessary to form this number with the most useful words, that is to say, the most frequent words are needed in this context.

4. METHODOLOGY

This study has been based on usual linguistic research methodology by applying to corpus sample. Corpus based approach has many applications on different parts of the area of linguistics, such as grammar, lexicography, language acquisition, discourse and so forth (Rayson, 2002). These applications come as a result of five main steps. These steps are;

1. Question: A research question or model is devised

2. Build: Corpus design and compilation
3. Annotate: Computational analysis of the corpus
4. Retrieve: Quantitative and qualitative analyses of the corpus
5. Interpret: Manual interpretation of the results or confirmation of the accuracy of the model (Rayson, 2002:13)

As it is seen, according to the corpus based methodology, first of all a research question or research questions are formed, then the selected corpus is designed. After that, the data in the corpus are analysed by means of software programmes. After the analysis, qualitative and quantitative results are reached and these results are interpreted. In this study, an EFL textbook has been analysed by applying these steps.

To answer the research questions of this study, the textbook Upturn in English 8 has been analysed. Before computational analysis of the textbook, the book has been scanned and the pages of the book have been transferred to the computer. It is important to state that the quality of the scanning should be high because the PDF file is converted to Microsoft Word file by PDF Transformer programme and this programme needs high quality scanned pages. Ünlü (2012:40) states that “the image quality of the scanning should be at least 300 dpi for obtaining more accurately read texts by the programme”. With regard to this statement, the PDF file of the textbook has been converted to a Microsoft Word file by the converter programme “Abby Fine Reader 12”. Nevertheless, some parts of the textbook include visuals in which the texts are not clear or hand-written, thus, these parts have not been converted to Microsoft Word file and the texts in these parts have been written to Word file manually.

The computational analysis of the words in the textbook has been performed by two different corpus software programmes. One of them is “Compleat Lexical Tutor v.8” and the other one is “Vmatrix”.

5. FINDINGS

5.1. What are the most frequent words in the textbook?

The frequency list tells us information about the core vocabulary of the corpus in the textbook. Below is Table 1 which shows the most frequently used 100 words in the textbook.

Table 1. Most Frequent 100 words in Upturn in English 8

Word	Frequency	Relative Frequency (%)	Word	Frequency	Relative Frequency (%)
the	1749	6.25	s	59	0.21
and	831	2.97	be	58	0.21
a	605	2.16	answer	58	0.21
you	580	2.07	share	57	0.20
I	514	1.84	ask	55	0.20
to	514	1.84	simple	55	0.20
is	490	1.75	friends	54	0.19

in	489	1.75	does	53	0.19
your	348	1.24	help	53	0.19
it	347	1.24	example	52	0.19
do	346	1.24	very	52	0.19
of	342	1.22	photos	52	0.19
are	306	1.09	science	52	0.19
can	254	0.91	then	50	0.18
with	239	0.85	her	50	0.18
what	237	0.85	if	50	0.18
about	230	0.82	because	49	0.18
my	215	0.77	more	49	0.18
for	207	0.74	water	49	0.18
on	201	0.72	well	48	0.17
unit	194	0.69	some	48	0.17
like	151	0.54	take	47	0.17
at	143	0.51	many	47	0.17
write	136	0.49	doing	46	0.16
not	128	0.46	them	46	0.16
have	123	0.44	why	46	0.16
or	116	0.41	understand	45	0.16
how	116	0.41	listen	44	0.16
they	108	0.39	put	44	0.16
use	107	0.38	please	44	0.16
internet	99	0.35	one	43	0.15
we	96	0.34	short	43	0.15
am	94	0.34	first	43	0.15
read	94	0.34	so	42	0.15
an	94	0.34	world	42	0.15
he	87	0.31	his	42	0.15
would	83	0.30	will	42	0.15
make	83	0.30	work	41	0.15
going to	80	0.29	sports	41	0.15
that	80	0.29	hello	41	0.15
people	79	0.28	who	41	0.15
questions	76	0.27	prefer	41	0.15
there	71	0.25	than	41	0.15
when	70	0.25			
below	69	0.25			
but	68	0.24			
activity	68	0.24			
she	68	0.24			
information	67	0.24			
check	66	0.24			
go	65	0.23			
usually	65	0.23			
school	64	0.23			
sentences	63	0.23			
text	63	0.23			
lesson	60	0.21			
this	59	0.21			

As it is seen from Table 1, generally the prepositions, pronouns, determiners, conjunctions, auxiliary verbs and particles are on the top of the list. It can be said that their being on the top of the list results from their functions; in other words, these words are not repeated intentionally.

It is seen that the definite article “the” is the most frequent word in the corpus with relative frequency of 6.25%. It is followed by conjunction “and” with relative frequency of 2.97%, indefinite article “a” with relative frequency of 2.16 and personal pronoun “you” with relative frequency of 2.07%.

O’Keeffe, McCarthy and Carter (2007) state that if the personal pronouns “I” and “you” and the discourse markers such as yeah, ok, hello, er are reflected in the high rank of the frequency list of any corpus, it can be said that the corpus presents a speaker-listener environment, which is beneficial for the communicative environment of the students. O’Keeffe, McCarthy and Carter (2007) further add that if there is a predominance of the third person references with prepositions and conjunctions in any corpus, the corpus represents the “the world out there”. These kinds of words are functional rather than lexical; they have little help to the content of the corpus. With this respect, it is seen from the most frequent 100 words in the textbook that although the personal pronouns “I” and “you” are on the top of the list, the discourse-marking items are not among the most frequent words in the list. Only the word “Hello” is in the list of most frequent 100 words. This shows that the vocabulary of the textbook does not help the communicative environment of the students.

5.2. How well are K1 Words, K2 Words and Academic words used in the textbook?

After the computational analysis of the words in the textbook by “Compleat Lexical Tutor v.8”, four frequency lists have been reached. These frequency lists are K1 Words which include the most frequent 1000 words in English, K2 Words which include the second most frequent 1000 words, AWL Words which include the words with lower frequency but common in academic and university texts and Off-List Words which do not belong to any of these three categories. It is necessary to state that the main aim of this study is to analyse the vocabulary in the textbook in terms of K1+K1 Words, academic words and off-list words.

Graphic 1. Percentages of representation of frequency lists in the textbook

Graphic 1 shows the clear domination of the K1 Words with almost 76% of the words in the textbook belonging to this class. This is followed by Off-List Words with 12.09% and K2 Words with 9.1%. AWL Words List has the least percentage with 3.04%. It is clearly seen from the graphic that K1 Words and K2 Words form almost 85% of the words in the textbook. However, it should be kept in mind that K1 Words is generally made up of function words as is seen in Table 2.

Table 2: Vocabulary Profile of the words in Upturn in English 8

	Families	Types	Tokens	Percent
K1 Words (1-1000):	709	1273	22286	75.76%
Function:	...	...	(13112)	(44.58%)
Content:	...	...	(9174)	(31.19%)
> Anglo-Sax =Not Greco-Lat/Fr Cog:	...	...	(5070)	(17.24%)
K2 Words (1001-2000):	370	530	2677	9.10%
> Anglo-Sax:	...	...	(1027)	(3.49%)
1k+2k		...	...	(84.86%)
AWL Words (academic):	141	211	895	3.04%
> Anglo-Sax:	...	...	(30)	(0.10%)
Off-List Words:	?	1032	3557	12.09%
	1220+?	3046	29415	100%

It is seen from Table 2 that 44.58% of the K1 words are made up of function words: prepositions (of, at, in...), pronouns (he, they, anybody...), determiners (the, a, my...), conjunctions (and, that, when...), auxiliary verbs (am, is, are, have...) and particles (no, not, nor...). Only 31.19% of the K1 Words are made up of content words like nouns, verbs, adjectives and adverbs. Thus, it is quite natural that the percentages of the K1 Words are high in the textbook because of the high percentage of the content words. In this context, the words in the textbook correspond with the 1000 most frequent words in General English.

The same conclusion cannot be made with the K2 Words. K2 Words form only 9.10% of the whole words in the textbook. K2 Words are made up of content words and are more essential for the meaning than K1 Words. Thus it is clear that there should be an increase in the percentage of second 1000 most frequent words. The 8th grade English Language Learners in Turkey have been learning English for six years, therefore, it is assumed that most of the K1 Words (1-1000) have been learned or handled to a certain extent. In this regard, more prominence needs to be given to increasing the number of the K2 Words (1001-2000) within the textbook.

Another surprising result is that the percentage of the Off-List words is unusually high, contributing 12.09% to the total. Off-list words may be proper nouns, uncommon words, abbreviations, acronyms, words belonging to different dialects, special vocabulary, misspellings and so forth. However, some of the words such as amazing, bye, chat, dentist, e-mail, festival, guitar, jogging, mall, notebook, ok, photos, honey, hiking, tablet, teenager, theatre and upload are classified as Off-List Words despite their being very functional for communicative purposes.

The least percentage of words belongs to AWL Words with 3.04%. It can be stated that the vocabulary of the textbook includes very few academic words, which do not cause much difficulty for the students.

In conclusion, it can be said that the percentage of the K1 Words and Academic Words are accepted in number, both taking the students' needs into account. However, the percentage of K2 Words is very low and there should be an enormous increase in the number of these words because this list includes content words and they are highly functional in the communicative context of the students. In contrast, the percentage of

Off-List Words is very high, which causes problems for vocabulary retention of the learners because dealing with Off-List Words is time consuming and not meaningful. Thus, there needs to be a decrease in the

percentage of the OFF-List Words.

5.3. Does the corpus in the textbook coincide with A2 Level according to CEFR?

Council of Europe (2001) stresses selecting the words in authentic spoken and written texts in language teaching and learning. That is why; the vocabulary of the textbook Upturn in English 8 has been examined by comparing it to the British National Corpus. Likewise, Council of Europe (2001) highlights applying statistical data in selecting the words with highest frequencies and prioritizes word lists for this purpose. In this study, the most frequently used words in the textbook have been analysed by referring to wordlists.

Council of Europe (2001:112) presents the vocabulary range of A2 Level of CEFR as follows:

Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics, has a sufficient vocabulary for the expression of basic communicative needs and has a sufficient vocabulary for coping with simple survival needs (Council of Europe, 2001:112).

As it is seen, the A2 Level vocabulary range of the CEFR gives great importance to communicative needs and shows that the learner needs to have a vocabulary to deal with the simple survival needs. It has been stated before that vocabulary range of the A2 Level is around 1500 and 2000 words (Van Ek & Trim, 1990; Meara & Milton, 2003; Milton & Alexiou, 2009). Therefore, this vocabulary range needs to be consisted of most useful words for the learners, that is to say, the most frequent words are needed in this context. However, in the previous section, it has been shown that there is an inconsistency in the vocabulary profile of the textbook. Although 76% of the words in the textbook belong to the class of most frequently used 1000 words in General English, only 9.10% of the words belong to the class of most frequently used second 1000 words in General English. This is a very low percentage because most frequently used second 1000 words are made up of content words and very essential for the meaning.

Although it has been stated in CEFR that the learner needs an adequate vocabulary necessary for basic communication and survival needs, the vocabulary in the textbook has 12.9% Off-list words. Uncommon vocabulary, words belonging to different dialects and special vocabulary do not help the vocabulary learning of the learners because dealing with Off-List Words is time consuming and not meaningful.

The A2 Level of CEFR further claims that vocabulary is very

fundamental to achieve communicative objectives of the language teaching and for the communicative competency of the students. (Council of Europe,2001). To achieve these communicative objectives, the learner needs to take part in the active participation in communication. Thus, the vocabulary of the spoken language is very necessary. However, as it has been seen in Section 4.7, the interjections, which are the indispensable elements of the conversations, are highly underused relative to the British National Corpus sampler. There are only 12 types of interjection that are used in the textbook. These interjections are: hello, yes, hi, no, bye, oh, yeah, hey, goodbye, oh dear, wow and cheers. Bearing this importance of interjections for the communicative purposes of the language teaching, it is necessary that there should be more interjections in the dialogues of the textbook.

Apart from interjections, it is clearly seen that the vocabulary of the textbook does not show the examples of the authentic language. As it has been stated earlier, personal pronouns and singular determiners are underused when compared to the British National Corpus. However, the definite and indefinite articles, possessive pronouns and unit of measurements are highly overused relative to the BNC. There seems to be an inconsistency in the usage of the words in the textbook Upturn in English 8 with BNC. The vocabulary of the textbook does not reflect the vocabulary of the BNC.

In conclusion, it can be said that because of the low usage of most frequently used second 1000 words, over use of Off-list words, insufficient use of spoken language (e.g. interjections) and the inconsistency in the use of some parts of speech, the corpus in the textbook Upturn in English 8 is not appropriate for A2 level according to CEFR.

6. CONCLUSION AND SUGGESTIONS

This corpus based study has been conducted to analyse the use of vocabulary in an A2 level EFL textbook in terms of the use of high frequent words and appropriateness of the words to A2 Level of the CEFR It also aims taking attention of textbook writers and teachers in selecting vocabulary in foreign language teaching.

The results of the frequency of the words have shown that generally the prepositions, pronouns, determiners, conjunctions, auxiliary verbs and particles are on the top of the list because of their functions. Although the personal pronouns “I” and “you” are on the top of the list of most frequent 100 words, they do not help to the communicative competence of the

students because they are not supported by the discourse markers and interjections in the textbook.

The results of K1 Words, K2 Words and Academic Words have shown that there is a clear domination of the K1 Words (0-1000) on the vocabulary of the textbook. Nearly 76% of the words in the textbook belong to this class. Thus, it can be claimed that the words in the textbook correspond with the 1000 most frequent words in General English. However, the words in this class are generally made up of function words (e.g. prepositions, pronouns, determiners, conjunctions, auxiliary verbs, and particles), thus their over use results from their function, in other words they are not recycled on purpose. Nevertheless, this high percentage disappears when K2 Words (1001-2000) are analysed. K2 Words form only 9.10% of the whole words in the textbook. The percentage of the K2 Words is very low although they are made up of content words and are more essential for the meaning than K1 Words. It is clearly seen that there should be an increase in the percentage of second 1000 most frequent words. Another interesting result is that Off-List words are unusually used too much with the percentage of 12.09%. There should be a decrease in the percentage of Off-List Words because Off List Words may be proper nouns, uncommon words, abbreviations, acronyms, words belonging to different dialects, special vocabulary, thus, cause problems for vocabulary retention of the learners. Dealing with Off-List Words is time consuming and does not seem so meaningful despite the fact that some of them help the communicative competency of the students, such as the Off-List words in the Unit “Cooking” and Unit “Tourism”. Contrary to Off-List Words, AWL (Academic) Words constitute only 3.04%. of the whole words, which is the positive side of the vocabulary profile of the textbook.

When it comes to the appropriateness of the corpus in the textbook for A2 Level of CEFR, the results have shown that because of the low use of most frequently used second 1000 words, high use of Off-list words, insufficient use of spoken language (e.g. interjections) and the inconsistency in the usage of some parts of speech, the corpus in the textbook Upturn in English 8 is not appropriate for A2 level according to CEFR.

7. BIBLIOGRAPHY

- Akbari, R. (2012). Vocabulary in L2 teaching: Some statistical findings. *TESOL Arabia Perspectives*, 19(2), 7-11.
- Bauer, L. & Nation, P. (1993), Word families. *International Journal of Lexicography*, Vol. 6 No. 4, 253-279.
- Council of Europe. (2001). *Common European framework of reference for*

- languages: *Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Genç, B. (2004). New trends in teaching and learning vocabulary. *Çukurova University Journal of Social Sciences* 13 (2), 117-126.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Karatay, H. (2007). Kelime öğretimi. *Gazi Eğitim Fakültesi Dergisi*, Cilt 27, Sayı: 1, 141-153
- Laufer, B. (1986). Possible changes in attitude towards vocabulary acquisition. *IRAL*, 24(1), 69-75
- Meara, P. & Milton, J. (2003). *X_Lex, the Swansea levels test*. Newbury: Express.
- Michael F. G., Diane, A. & Jeannette, M. (2012). *Teaching vocabulary to English language learners*. New York, NY: Teachers College Press.
- Milton, J. & Alexiou, T. (2009). Vocabulary size and the Common European framework of reference for languages. In B. Richards et al. (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 194–211). Basingstoke: Palgrave.
- Moghadam, H. S., Zainalb, Z. & Ghaderpourc, M. (2012). A review on the important role of vocabulary knowledge in reading comprehension performance. *The 8th International Language for Specific Purposes (LSP) Seminar - Aligning Theoretical Knowledge with Professional Practice. Procedia - Social and Behavioral Sciences* 66: 555 – 563.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening. *Canadian Modern Languages Review*, 63(1)59-82.
- Nazary, M. (2008). The role of L1 in L2 acquisition: Attitudes of Iranian University students. *Novitas-ROYAL*, 2(2), 138 – 153
- O’Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press
- Rayson, P. (2002). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. Unpublished Ph.D. thesis. Computing Department, Lancaster University.
- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press.
- Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In: Campoy-Cubillo, M. et al. (eds.): *Corpus- based approaches to English Language Teaching*. London: Continuum, 18-38
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, England: Cambridge University Press
- Thornbury, S. (2002). *How to teach vocabulary*. Harlow: Longman.
- Topkaraoğlu, M. & Dilman, H. (2015). The impact of vocabulary instruction on vocabulary size levels of students. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15 (2), 349-363.

- Ünlü, F. (2012). *Loading, distribution, and repetition patterns of the 2000 high frequency words of General English in an EFL course book*. Unpublished MA Thesis. Department of English Language and Literature, Karadeniz Technical University.
- Van Ek, J. A. & Trim, J. L. M. (1990). *Threshold 1990*. Cambridge: Cambridge University Press
- Van Ek, J. A. & Trim, J. L. M. (1990). *Waystage 1990*. Cambridge: Cambridge University Press.
- Wallace, C. (2008). Vocabulary: The key to teaching English language learners to read. *Education Digest*. 73 (9), 36-39
- Waring, R. & Nation, P. (1997). Vocabulary size, text coverage, and word lists. In *Vocabulary: Description, Acquisition and Pedagogy* N. Schmitt and M. McCarthy (eds.). Cambridge University Press, Cambridge: 6-19.
- West, M. (1953). *A general service list of English words*. London: Longman, Green and Co.
- White, R. V. (1989). *The ELT curriculum: Design, innovation and management*. Cambridge: Blackwell.