



Lexical Analysis of Two ELT Coursebooks: Corpus-informed & Conventional Coursebook

Mürüvvet SÖĞÜT^{1, a}, Ersen VURAL^{2, b},

¹Antalya Bilim University, Foreign Languages Dept., Antalya, Türkiye
²Akdeniz University, Faculty of Education, ELT Dept., Antalya, Türkiye

^aORCID: 0000-0003-4197-201X; ^bORCID: 0000-0002-2779-9153

✉Corresponding Author: ersenvural@akdeniz.edu.tr

Received: 20/02/2023 Accepted: 30/03/2023

Abstract: The objective of the present study is to examine the lexical profiles of two pre-intermediate level ELT coursebook titles commonly used in language classrooms. One coursebook is credited with being “corpus informed” while the other is published as a conventional coursebook with no special emphasis on its lexical selection. With the help of lexical frequency analyses, the study attempts to quantify the extent of their coverage of 2K high-frequency words from New General Service List (NGSL). Similarities and differences in the lexical frequency profiles of the books were investigated using corpus-based frequency analyses. For this purpose, a vocabulary profiler and frequency calculation tool were used. The resulting lexical profiling of the books revealed that the corpus-informed coursebook had a coverage of the most frequent 2K words of English (NGSL-1 and NGSL-2) at 74.2% level, while the conventional coursebook title had a coverage at 82.2% level. The lexical coverage of high frequency words in the corpus-informed book was counterintuitively found to be lower than the conventional coursebook. Again, the analyses of the type-token ratios and recycling rates of the lexical items in two coursebooks indicated that the corpus-informed book did not vary markedly from the conventional coursebook. From the analyses there emerged no indication of any “systematic approach” to lexical coverage and recycling of high frequency lexical items in the coursebook title published as ‘corpus-informed’ coursebook.

Keywords: *Vocabulary, Corpus, Coursebook, Frequency, Recycling*

1. Introduction

A sufficient level of comprehension and production of second language requires knowledge of both lexical and grammar forms. Using their knowledge of L2 form-meaning mappings at lexical and structural (grammar) level, learners comprehend and convey messages in L2. Although knowledge of the two types of mappings is essential, knowledge of lexical form-meaning mappings is usually considered more primary than that of grammar. As Lewis (1993) argues “vocabulary or lexis carries more of the meaning of a text than does the grammar” (p.17). Doughty and Long (2003) observe that most of the difficulties that arise while communicating in another language derives from gaps in learners’ knowledge of L2 lexicon. This can also be observed in the proportion of learners’ vocabulary errors that occur more frequently than grammar errors. This observation cannot be downplayed because sentences containing grammar errors are still intelligible to language learners, while sentences lacking lexical specificity or containing lexical errors impede communication in L2 seriously (Gass, 1988). That simple fact can reasonably justify why vocabulary learning has recently come to be seen an integral part of language learning.

With the shift of focus towards communicative competence in language development, knowledge of vocabulary has become central to language teaching. Consequently, a search for effective and systematic ways of vocabulary instruction arose in the field. Attention has now turned to developing

lexically oriented L2 instructional materials. The new generation of instructional materials which multiply in number and diversity confirms that paradigm shift. They are designed to enable learners to increase their exposure frequency to most frequent L2 lexical items in specifically created tasks and to enrich their learning experience by offering engaging content.

A new dimension to the lexically oriented instructional materials has come from digital learning environments. There is today a wide array of computer/mobile programs designed to get learners interactively engage in vocabulary learning tasks. Nevertheless, despite the exponential increase in the use of digital technologies within language learning contexts, the use of print medium in classroom remains steadfast. As time-honored materials, coursebooks keep providing input to second language learners as staple learning materials. Since the bulk of the L2 input most learners receive comes from content of coursebooks, they often get the criticism that the type and frequency of words in the content may be skewed so that they lead to misguided learning of L2 vocabulary. In more precise terms, since their content is relatively limited and artificially created lexical items in the coursebooks may not ideally represent the items occurring frequently in real language use. Based on such criticisms, it becomes evident that the content of coursebook titles needs to be screened constantly to ensure they represent real language. That need is most obvious when we consider the fact that frequency and currency of words in authentic language usage change over time. A coursebook title created twenty years ago would certainly contain words that were frequent and current at the time but have fallen out of use or been replaced by more current ones in the intervening years.

Given the importance of lexical selection for the content of coursebook titles, the onus is always on coursebook developers to ensure that learners encounter current and most frequent words in language. Especially novice learners' progress after their initiation to target language critically depends on careful selection and presentation of lexical items. For effortless processing of L2 input, they need to be methodically exposed to 1/ 2K high-frequency words in the beginning stage of learning. Nation (2006) argues that 'core words' in language should get more attention and be taught systematically. Taking such considerations into account, a judicious coursebook writer makes proper lexical decisions by arranging words hierarchically in the content of coursebook titles. More precisely, more frequently used words in language are prioritized in the coursebook over the less frequent ones.

Such a principled prioritization and systematic approach to lexical coverage in coursebook content draws heavily on corpus data, which is the most reliable reference source for analyzing the frequency of occurrence of language items in real language use. Since a corpus is formed by compiling spoken or written data from native speakers' authentic use of language, it provides solid evidence on how they authentically use language in different contexts. Material developers who are concerned with careful lexical selection for their coursebook contents need to rely on corpus information to determine lexical items' context of use, frequency of occurrence and collocational relationships.

Despite increased use of corpora in the creation of instructional materials (coursebooks in particular) there appears to be a lack of systematic approach in the coverage, selection, and distribution of vocabulary in coursebooks. Excepting the authors of 'corpus informed' coursebooks, it appears that many authors rely on their intuition for selecting lexical items in the content of their books. To date, a few coursebooks are published as corpus informed. Nevertheless, no substantiating data exist to reveal the extent corpus data is tapped for vocabulary selection in those coursebook titles. It is still not known whether the titles defined as 'corpus-informed' vary by some means from conventional coursebooks in terms of their lexical coverage. As indicated, relatively few studies exist in that area. The present study seeks to address this research gap by examining the lexical content of two coursebooks (corpus-informed and conventional) with respect to their representativeness of corpus data. It sets out to detect the high-frequency words from corpus in two different coursebook titles and to check whether they are recycled in a systematic way throughout the books.

1.1. Review of Literature

Words, usually defined in cognitive tradition as form-meaning mappings, are stored in brain in the form of semantic networks. They get semantically connected to each other by forming new associations as language users exploit the communicative potential of language to receive or convey meaning (Aitchison, 1994). For this to occur, one essential requirement is to afford learners repeated encounters with lexical items so that links in associative brain are formed and get entrenched in the long-term memory. The process of forming such richer and complex connections among related concepts in the brain explained by Craik and Lockhart (1972) as ‘the depth of processing hypothesis. According to the hypothesis when words are processed more deeply, they result in better storage in the long-term memory. Since resulting memory traces are usually long lasting and immune to forgetting, learners can internalize and remember those items with ease (Nation, 2009).

Besides depth of processing, a second condition for better entrenchment (mental storage) is how often a lexical item is encountered by a learner. Tyler (2012) proposes that learners need to encounter words multiple times in varied contexts to gain good receptive and productive knowledge and to process them fluently during online tasks. The number of studies on the influence of lexical encounter frequency increases as confirming evidence from corpus linguistics accumulates. Supported by advances in computer technology to collect and statistically analyze data on real language use, corpus studies indicate that word frequency underlies not only lexical items but whole language system.

Acquisition studies indicate that word frequency in language input has a strong impact on memorization, lexical access, word storage and retrieval (Nation, 1993; Schmitt, 2008). Likewise, neurological accounts of frequency effect on learning show that repetition of the same cognitive/mental operations influences the structure of specific synapses in the brain (Sanchez & Criado, 2012). Accordingly, when the connections involved in lexical processing get stronger, similar tasks subsequently reach a level of automaticity with which it becomes less effortful to retrieve words from memory. Ellis (2002) maintains that the number of times a task is practiced influences the speed of performance so that memory for words is strengthened with each repetition. Studies from different research domains find convincing evidence that frequency is a function of efficient storage and retrieval of high words in language.

By encountering words repeatedly, the threshold level for activation of language memory gets lower (Divjak, 2019), which means a language user has reached the minimum level of proficiency for language performance. From that stage onwards, it becomes less effortful for the learner to activate memory for past language experiences. A lower threshold for a word means less additional activation to recognize the word and swifter and less effortful retrieval of lexical form-meaning mappings. It is without doubt that the frequency of encountering a word is an important condition for retention and retrieval of it. Nonetheless, opinions differ regarding the number of times needed for effective vocabulary learning. Waring and Takaki (2003) maintains that at least 8 repetition is required, whereas Cameron (2001) suggest a minimum of 5-6 encounters will lead to learning of the new words. Schmitt (2008) states that even 8-10 exposures would only improve learners’ recognition of the words. The number of repetition suggested by Nation and Wang (2020) is at least ten. They also propose that the higher the number of repetitions, the better is the learning outcome.

Again, research findings from corpus studies increasingly indicate that most frequent words in the language must be prioritized to make sure that learners retain and retrieve those central lexical items in language (Horst, 2013). Acknowledgement of the importance of learners’ frequency of exposure to lexical items in language input generated some frequency lists to guide coursebook preparation practices. For example, West’s General Service List (1953) is typically used by teachers and course/textbook designers as a reference to make careful lexical selections. It is based on a written corpus of 5K words and consists of 2K headwords. An updated version of the list, New General Service List (Browne et al., 2013), was compiled based on a larger corpus with a better coverage in 2013. With the creation of NGSList, the coverage rate of GSL (84%) was increased to 92%, which means that

knowledge of words at that level will enable learners to comprehend most of the vocabulary used in many general English texts.

Lexical prioritization or limiting the range of words used in an instructional material has the potential to facilitate learners' experience of processing L2 input. Since 2K words from the high frequency bands would correspond to approximately 85% of the lexical items that make up any kind of text or discourse, concentrating on them would lead to rapid increases in reading and listening comprehension (Newton & Nation, 1997). In that regard, it could be said that taking word frequencies into account in the process of developing L2 instructional materials will undoubtedly save time and effort on the part of both learners and instructors.

Word lists showing high frequency items in corpus are essential both for the development and content analysis of language coursebooks. Today computerized frequency lists allow material developers and researchers alike to compare any content against computerized frequency bands. The resulting lexical profiling provides them with the proportions of frequent or infrequent words so that accurate judgements about lexical coverage are made. Content analyses of instructional materials using computer assisted lexical profiling usually provide invaluable information which would otherwise be impossible to obtain. For example, in one such analysis study Nordlund and Norberg (2020) identified a number of limitations existing in instructional materials they investigated. Their study found that lexical selection in the content of the coursebooks was not systematic and lexical items used were not recycled either. The coursebooks analyzed in their study were also found to misrepresent authentic language use in their content.

Studies examining the recycling rate of words in English teaching materials demonstrated that the number of times words occurred in textbooks did not provide sufficient exposure for language acquisition (Fujimori, 2005; Konstantakis & Alexiou, 2012; Nordlund & Norberg, 2020). Konstantakis and Alexiou (2012), for example, analyzed the lexical content of five EFL coursebook titles used in Greece with a corpus-based approach. They determined that the books contained proportionately higher number of words from mid- and low-frequency bands and fewer number words from the high-frequency bands on British National Corpus (74-85%). Based on that finding, they suggest that word frequency information needs to be taken into consideration by coursebook authors and teachers. In another study by Reda (2003), it was found that vocabulary selection of many ELT books is governed largely by the topic-based units and that they are commonly characterized by their limited range of lexical items as they include topics of similar nature.

Another study by Ünlü (2012) on loading, distribution, and repetition patterns of the 2K high-frequency words of general English in an EFL coursebook revealed that the coursebook contained only a limited number of most frequent words in English. The researcher concluded that vocabulary component in the textbook examined was neglected to a large extent.

The coursebooks reported in the studies above are described as regular (conventional) coursebook titles without any special emphasis on deliberate and careful vocabulary selection. Because of that, one would anticipate that vocabulary selection in the books is governed mostly by the requirements of the selected topics. To the best of our knowledge, no study has been conducted to test whether supposedly corpus-informed coursebook titles proportionately include higher number of frequent words in their content and to corroborate the claim that 'principled lexical selection' reportedly adopted for those books is in keeping with the lexical frequency bands determined by corpus. The present study attempts to answer that question by comparing a coursebook title credited with the feature of being 'corpus-informed' for its lexical selection to a regular coursebook without that feature.

The present study attempts to determine the extent of the vocabulary content of two types of coursebooks: a 'corpus-informed' English teaching coursebook titled 'Touchstone Level 3, Student's Book' (McCarthy et al., 2014), and a regular coursebook titled 'Headway Pre-intermediate Level, Student's Book' (Soars & Soars, 2007). The coverage and recycling of vocabulary in the books were

examined with respect to their representativeness of real language use from the corpus data. To this end, the two books were submitted to content analysis to reveal their coverage of most frequent 2K words from New General Service List (NGSL).

Four research questions guided the study:

RQ1 What is the coverage level of the most frequent 2K words of NGSL in a corpus-informed coursebook (Touchstone Level 3, Student's Book) and a regular coursebook (Headway, Pre-intermediate Student's Book)?

RQ2 What is the recurrence rate of words in corpus-informed and regular coursebook?

RQ3 What is the recycling rate of twenty select words in two textbooks?

To what extent does the frequency of occurrence of the twenty words in two textbooks reflect the corpus-based frequency ordering?

2. Method and Materials

In the study, two small-scale corpora emerging from the contents of the two English textbooks were investigated using a quantitative corpus linguistics methodology and systematic sampling method for deciding on twenty words to represent the ordering of the words on 2K most frequent-words list. The lexical content of each book was examined by means of a vocabulary profiler accessible online as Compleat Lexical Tutor (Cobb). With the help of the vocabulary profiler tool, information was revealed about the lexical content of the two books as well as the range and recycling rates of lexical items. The resulting lexical profiles of two textbooks were subsequently compared to show the extent of their representativeness of their authentic use.

2.1. Sampled Materials

The regular coursebook analyzed in terms of its content titled 'Headway Pre-intermediate Student's Book', which was published globally for quite a considerable length of time and reached large audience at different instructional settings. The coursebook contained no information to show that it was created as a corpus-informed coursebook. As the coursebook is known to be commonplace instructional material in language classrooms, for the present study it was believed to typify other regular coursebooks that do not claim to have put any special emphasis on corpus based lexical selection. While the new edition of the coursebook has been updated in terms of its methodology, its text contents and integration of authentic language / digital resources, it remains as a topic-based book. The content of the book is claimed to be based on a well-balanced grammar and skills syllabus.

The corpus-informed coursebook title examined in the study was Touchstone 3 (A2-B1 Level). The book is claimed to have been created using corpus information from The Cambridge English Corpus, which contains multi-billion words compiled from different written/spoken language samples in British and American English. In the introduction section of the coursebook title, information is provided informing that the authors determined the most useful frequent items and typical communicative uses in everyday situations after years of work on the Cambridge International Corpus. For the research purposes in the present study, that book was assumed to represent other typical corpus-informed coursebooks that claim to systematically include more frequent lexical items from real language use.

For the purposes of practicality and their representative property, the study focused only on 'reading spots' of the two books so that the results from those sections could be generalized to the whole content of the books. That would be a justifiable decision considering that a strong relationship exists between reading practice and vocabulary learning and that reading activities make up the bulk of the

classroom activities. In this regard, it should also be mentioned that coursebook sections dedicated to reading are prioritized over sections containing listening writing and speaking tasks.

2.2. Vocabulary Profiler

Evaluation of the lexical profiles of the content of the books was made possible by Compleat Lexical Tutor (Cobb), an online accessible tool which offers a various analysis tools for researchers, teachers, and learners to determine the lexical content of texts. In the study, the analysis of the frequencies of lexical occurrences and the ranges of lexical coverage in the content of two books was carried out using various modules of the tool in question.

For answering RQ1, Vocabulary Profiler Compleat (Cobb) tool was used to determine the lexical coverage of the two coursebooks. The tool was further used to reveal the lexical diversity of the contents of the books and to compare them against the lists of high-frequency words from corpus. The emerging lexical items in the textbook contents were divided into frequency bands from New General Service List (Browne et al., 2013), a commonly used corpus based word list containing most frequently used lexical items in authentic language use.

The corpus-based word lists in question were from the New General Service List, which is a list of 2K core English words considered to be the most useful for non-native English speakers to learn. The list provides a starting point for building a vocabulary in English and is used as a reference for language teaching materials and assessments. The list has two bands, NGSL-1 and NGSL-2 containing the most common 1K / 2K words respectively in a text or transcribed discourse. Using the Lexical Profiler Compleat tool, words in the textbooks belonging to the most frequent 1K English words (NGSL-1) and 2nd most frequent 1K English words (NGSL-2) were identified. Additionally, words belonging to the New Academic Word List (Browne et al., 2013) and off-list words not appearing in NGSL and NAWL, were identified and classified accordingly.

2.3. Frequency Calculation Tool

To answer RQ3, twenty words having a certain distance in between on NGSL1 and NGSL2 were selected as samples to reveal whether the frequency of occurrence of words reflect corpus-based frequency orderings. The frequency of occurrences of selected words in two books were determined using the frequency calculator at Compleat Lexical Tutor (Cobb), which gives the frequency count of each lexical item occurring in the text and lists them in the order of their recurrence frequency.

2.4. Extraction of the Coursebook Contents

The reading sections of the books were extracted to perform lexical analysis of the content of the books. The resulting data was evaluated by comparing them to the list of 2K high frequency words from New General Service List (NGSL1 and NGSL2) and New Academic Word List (NAWL). This required transcription of the given sections of the books into word processable data. As the textbooks came in PDF file format, the relevant sections were transformed to the suitable format using Abby Fine Reader software. The occasional spelling mistakes were checked and corrected. Words in short forms were changed into long forms. Proper names were excluded. Instructions preceding or following the targeted sections were eliminated.

The modified versions of the two coursebooks were saved on separate word documents so that the data would be ready for lexical analysis. The data on the files separately were entered into the relevant sections at Vocabulary Profiler Compleat and Frequency at Lextutor. With the help of the analysis tool, the lexical items in the contents were sorted into frequency bands they belong and computed scores the were obtained for items in each coursebook.

The resulting data yielded the percentages and the number of occurrences of the frequent items in the data. The words were assigned to one of each category: NGSL-1, NGSL-2, NGSL-3, NAWL and off-list words. In the process, lemmatization was implemented ensuring that different inflectional forms are reduced to the base forms. The resulting data, therefore, contained the lemma forms of the words. For example, the words ‘information’ and ‘informing’ were lemmatized as instances of the base word ‘inform’.

To answer the RQ3 type-token ratios and indices were determined using VP-Compleat tool. The type-token ratios (TTR) of the occurrences of words were used for the comparison of frequency of items. TTR equals to the number when the total number of word types is divided by the total number of tokens. The resulting score is a number that falls within 0 and 1. Calculation of type-token ratios made it possible to determine what extent the words in the textbook were recycled.

2.5. Systematic Sampling of Words

To answer the RQ3, a set of 20 words were systematically sampled from NGSL-1 and NGSL-2. The aim was to find out to what extent the frequency of 20 randomly selected words in two textbooks reflected their frequency ordering on NGSL. The words were sampled from the lists by selecting words keeping a certain distance in between. This was done to allow words from various frequency ranges to evenly enter the sampling list. The standing orders of the select items was expected to serve as points of reference representing frequency scales in authentic language use.

Table 1. Selected words and their frequency ordering on NGSL

Select words and their standings on the frequency lists based on their corpus frequencies	Inflected forms before lemmatization
make (48)	made, makes, making
home (161)	homed, homes, homing
buy (242)	bought, buying, buys
reach (380)	reached, reaches, reaching
shop (437)	shopped, shopping, shops
listen (573)	listened, listening, listens
forget (651)	forgets, forgetting, forgot, forgotten
marry (775)	married, marries, marrying
suffer (837)	suffered, suffering, suffers
maintain (942)	maintained, maintaining, maintains
hang (1083)	hanged, hanging, hangings, hangs, hung
quick (1183)	quicker, quickest
trend (1279)	trended, trending, trends
usual (1383)	-
afford (1486)	afforded, affording, affords
unique (1592)	-
faith (1684)	faiths
multiple (1730)	multiples
bother (1842)	bothered, bothering, bothers
rural (1961)	-

3. Results and Discussion

The RQ1 in the study addressed the coverage level of the most frequent 2K words of NGSL in a two coursebooks. The answer to RQ1 is provided in this section. The results from the analysis of lexical coverage of two book are successively presented and discussed in the light of previous discussions.

As can be observed in Table 1, the analysis of vocabulary in the corpus-informed coursebook showed that more than half of the lexical items (74.2%) belonged to the bands containing the most frequent 2K words. This percentage for the coverage level for the most frequent words may not be conducive for comprehension. Hirsch (2003) asserts that since smooth comprehension depends on a higher level of frequent word coverage, a text should include frequent words at 95% level.

Table 2. Comparison of Lexical Coverages in two books

Frequency Bands	Corpus Informed Book			Conventional Book		
	TOKENS (%)	TYPES (%)	Sum	TOKENS (%)	TYPES (%)	Sum
NGSL 1	2964 (80.4 %)	702(57.6)	57.%	3104 (86.6 %)	663 (66.1 %)	66.%
NGSL 2	287 (7.8 %)	202 (16.6)	74.%	225 (6.3 %)	162 (16.1 %)	82.%
NGSL 3	140 (3.8 %)	95 (7.8 %)	82 %	120 (3.3)	73 (7.2 %)	89.%
NAWL	38 (1.0 %)	30 (2.4 %)	84.%	13 (0.4 %)	12 (1.2 %)	90.%
OFF-LIST	256 (6.9 %)	187(15.3)	100	122 (3.4 %)	92 (9.1 %)	100
TOTAL	3686	1217	100	3585	1003	100

For the corpus informed coursebook, the total number of counted lexical items in its content was found to be composed of words from NGSL-3 at 7.8 % level and academic word list at 2.4 % level. This could indicate difficulty for learners for comprehension because words from those two bands would be infrequent and unfamiliar to learners.

The low frequency words in a text content can be ignored when they are not numerous. However, the findings on the table show that there are 15.3% off-list words classified as low frequency lexical items. This number, which is almost close to the number of NGSL-2 words, may have negative effect on comprehension, and render the coursebook intended for learners at pre-intermediate level unnecessarily challenging.

In most of the previous research studies, analyses of the vocabulary profiles of textbooks demonstrated very similar results in terms of the coverage of words from different frequency levels. It was usually found that the higher the frequency of a word is, the higher their coverage is in the textbook. The analysis of vocabulary in the conventional textbook, Headway Pre-intermediate, revealed that more than half of the lexical items (82.2%) belonged to the most frequent 2K words.

In the conventional textbook, the coverage of NGSL-3 words is at 7.2% level and the coverage of academic words is 1.2 % level. This number may be considered as appropriate for a material designed for pre-intermediate level learners. Presence of a limited number of academic words at the pre-intermediate level could be a good preparation for the next level.

The total number of off-list words in conventional coursebook is 92 (9.1%), which is lower than the proportion low frequency words in corpus-informed coursebook, however, it may still be high to afford learners an effortless comprehension.

The findings for the RQ1 show that Headway Pre-intermediate, in which there is no mention of overt reference to corpus for lexical selection, has a relatively high coverage of high frequency words from 1K-2K bands. The contrast in the cumulative percentages of types from NGSL-1 and NGSL-2 lists in two books is a clear indication of that. This contrast can be observed in the percentages found for the words from 1K-2K bands in two books: corpus-informed coursebook (Touchstone 3) 74.2%, in the conventional coursebook (Headway Pre-intermediate) 82.2%.

Possible reasons for the difference in their coverage of high frequency words between two textbooks could be: Headway textbook series have been available for longer time and used more commonly than Touchstone series worldwide. Therefore, more feedback on its content may have been collected to update and adapt the book accordingly. Another reason could be that 5th edition of Headway is a more recently published book than Touchstone 3. Thus, vocabulary content in the reading spots of the book might have been recreated according to more up-to-date topics.

RQ2 What is the recurrence rate of words in corpus-informed and regular coursebook?

The number of tokens/types as well as the type-token ratios was found as in the following table.

TTR shows the number of lexical recurrences with high TTR indicating higher number of word recycling.

Table 3. Comparisons of Recycling Rates of Two Coursebooks

	Corpus-informed Coursebook	Conventional Coursebook
Types	1217	1003
Tokens	3686	3585
Type-Token Ratio	0.33	0.28
Token per Type	3.03	3.57
Lexical Density	0.57	0.5

To answer RQ2 and see the recycling rate of the vocabulary items, types are divided by tokens to get type-token ratio. As can be seen on Table 3 the type-token ratio for the corpus-informed textbook is 0.33 and 0.28 for the traditionally prepared textbook. Since the type-token ratios of both textbooks are found to be below 0.5 and even close to 0, we can conclude that the repetition of vocabulary in both coursebooks is not very low. Nevertheless, it should also be noted that the recycling rate of lexical items in corpus-informed coursebook is counterintuitively lower than the conventional coursebook with a small margin. Normally, a corpus informed instructional material is expected to afford learners repeated encounters with words since increasing the frequency of exposure to lexical items is useful for learners as it helps them deepen their knowledge (Nation, 2001).

Most exposure frequency studies suggest a minimum of five repetitions as the threshold for retention (Thornbury, 2002; Webb, 2007). As can be seen in the following table, for the corpus-informed textbook the total proportion of types with 5 or more frequencies is 8.8%. The same value is 11.3 % in the conventional textbook.

Table 4. Frequency of Lexical Recurrences

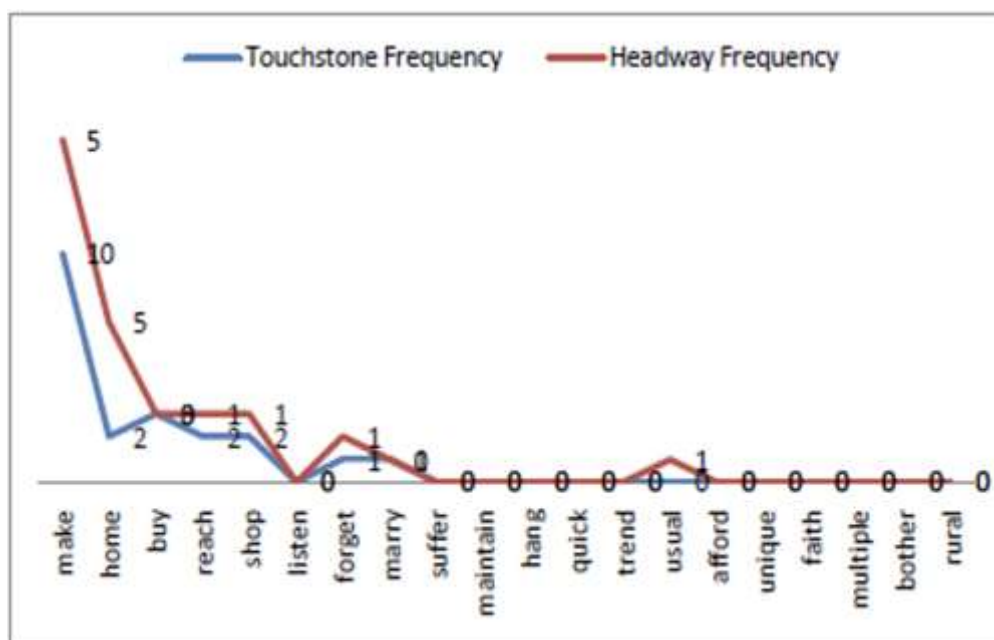
Frequency Values	Cumulative Coverage Corpus-informed Coursebook	Cumulative Coverage for Conventional Coursebook
10 +	4.30%	5.30%
5-10 Occurrences	4.50%	6%
3-5 Occurrences	10.8	10.5
2 Occurrences	17.20%	16.10%
1 Occurrence	62.80%	61.80%

Based on this it may be concluded that the recycling of types in the conventional textbook is slightly better than the corpus-informed textbook. The Proportion of the words which are recycled fewer than 5 times is 80% in corpus-informed course book and 77.9% in conventional coursebook.

This finding is in keeping another study by Ünlü (2012), who demonstrated that 60% of the types were not repeated and 9% of the lexical items were recycled 5 or more times in the book she analyzed. With similar results the researcher concluded that the book contained “inadequate number of recycled words with insufficient frequency values” (p.82).

RQ3 What is the recycling rate of twenty select words in two textbooks? To what extent does the frequency of occurrence of the twenty words in two textbooks reflect the corpus-based frequency ordering?

The orderings of the select items were checked based on the occurrence frequencies found in the corpus of the two books and subsequently were indexed on a line graph to visually compare them.



Graph 1. Frequency orderings of twenty select items

As can be seen on the Graph 1, the repetition and coverage and of 10 words from NGSL-2 group is lower than 10 words from NGSL-1 group in both books and the recycling rate of the words drops as their list standings based on occurrence frequencies decrease. Considering their distribution of words from 2K bands in authentic language corpora, this could be accepted as a normal distribution. Nonetheless the corpus-informed coursebook might have differed slightly from the conventional coursebook in this regard. One would expect to get an evenly declining line for the occurrence frequencies of select words in the corpus informed coursebook. Accordingly, it would not be wrong to conclude that that there is a lack of systematic and sufficient recycling for the randomly words in both textbooks.

4. Conclusion

It is likely that some English teachers and textbook writers may not be familiar with the basic corpus linguistics terms and corpus use in language teaching. Since most language teaching practice revolves around the confines of the coursebooks, teachers should be wary of the caveat that elements of authentic language like lexical items or word combinations are hard to be represented equally well in coursebooks in the strictest sense of the word. They should be made aware that the content for reading and listening for learners needs to be carefully supplemented by materials from authentic material for

native speakers. Furthermore, the teachers should confer corpus data for the selection of the authentic materials. They need to be familiar with corpus and its implications for their teaching practice, which requires corpus literary skills on the part of teachers.

Given the potentials of corpus for language teaching practice, it would be apt to say that coursebooks developers and teaching practitioners should reap the benefits of corpus appropriately. To ensure this, corpus literacy courses could be offered to stakeholders in language teaching field. Also, teachers should take initiatives and assume more responsibility in material development and adaptation of the textbooks by using corpus.

The present study found that there exists a need to include more high-frequency words in instructional materials. The developers of instructional materials and teachers must be informed about the importance of focusing on the core lexical items in language and of using frequent word lists for reference purposes. By tapping the easily accessible vocabulary profiler tools, it is quite practical to use high-frequency lists as a reference to check the coverage of frequent and infrequent words in a text. These tools can be used to check out low frequency words and to include more of high-frequency words. With the help of lexical content profiling, the stakeholders in instructional practice can easily decide on the lexical items. For example, teachers can rapidly make decisions about lexical items and include or exclude lexical items based on students' level and needs rather than strictly following the textbook content page by page.

Additionally, the coursebook could be supplemented with materials that reflect the most useful vocabulary and structures of English. Teachers may try to provide sufficient vocabulary repetition for learners during the lessons. In this respect, graded readers, which are lexically organized can be used to afford learners more frequent encounters with targeted words from frequency words lists. Since graded-readers are written based on the most frequent vocabulary, using them could be an effective and practical way of presenting high-frequency words to the students.

More importantly, the spotted inadequacies in the lexical selection of instructional materials, particularly coursebooks, should urge teachers to become active content developers rather than passive consumers of the content.

Conflict of Interest

“No conflict of interest was reported by the authours.”

The Rate Contribution

“The authors state that they have contributed equally to the article”

Ethical Statement

“The data, information and documents in the study were obtained in line with the official ethical guidelines.”

Financial Support

“The study was not funded by any institution or organization”

Remarks

The article was based on a MA thesis completed in 202

REFERENCES

Aitchison, J. (1994). *Words in the mind : an introduction to the mental lexicon / Jean Aitchison.* Blackwell.

- Browne, C., Culligan, B., & Phillips, J. (2013). *A new general service list: the most important words for second language learners of English (version 1.01)*. <http://www.newgeneralservicelist.org/>
- Cameron, L. (2001). *Teaching Languages to Young Learners*.
<https://doi.org/10.1017/CBO9780511733109>
- Cobb, T. *VocabProfiler*. <https://www.lexxtutor.ca/cgi-bin/range/texts/index.pl>
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684.
[https://doi.org/https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/https://doi.org/10.1016/S0022-5371(72)80001-X)
- Divjak, D. (2019). *Frequency in language: Memory, attention and learning* [doi:10.1017/9781316084410]. Cambridge University Press.
<https://doi.org/10.1017/9781316084410>
- Doughty, C. J., & Long, M. H. (2003). The Scope of Inquiry and Goals of SLA. In *The Handbook of Second Language Acquisition* (pp. 1-16).
<https://doi.org/https://doi.org/10.1002/9780470756492.ch1>
- Ellis, N. C. (2002). Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188. <https://doi.org/10.1017/S0272263102002024>
- Fujimori, J. (2005). The lexical composition of two oral communication textbooks *The Language Teacher*, 29(7), 5-19.
- GASS, S. M. (1988). Integrating Research Areas: A Framework for Second Language Studies 1. *Applied Linguistics*, 9(2), 198-217. <https://doi.org/10.1093/applin/9.2.198>
- Horst, M. (2013). Mainstreaming Second Language Vocabulary Acquisition. *Canadian Journal of Applied Linguistics*, 16, 171-188.
- Konstantakis, N., & Alexiou, T. (2012). Vocabulary in Greek young learners' English as a foreign language course books. *The Language Learning Journal*, 40(1), 35-45.
<https://doi.org/10.1080/09571736.2012.658222>
- Lewis, M. (1993). *The Lexical Approach: The State of ELT and a Way Forward*. Language Teaching Publications. <https://books.google.com.tr/books?id=nNRZAAAAMAAJ>
- McCarthy, M., McCarten, J., & Sandiford, H. (2014). *Touchstone Level 3 Student's Book*. Cambridge University Press. <https://books.google.com.tr/books?id=TrSAgAAQBAJ>
- Nation, I. (2001). *Learning Vocabulary in Another Language* Cambridge University Press
<https://doi.org/10.1017/CBO9781139524759>
- Nation, I. (2006). How Large a Vocabulary is Needed For Reading and Listening? *The Canadian Modern Language Review*, 63(1), 59-82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2009). *Teaching vocabulary: Strategies and techniques*. Cambridge University Press.
- Nation, P. (1993). Using dictionaries to estimate vocabulary size: essential, but rarely followed, procedures. *Language Testing*, 10(1), 27-40. <https://doi.org/10.1177/026553229301000102>
- Nation, P., & Wang, K. (2020). Graded Readers and Vocabulary.
<https://doi.org/10.26686/wgtn.12560366.v1>
- Newton, J., & Nation, I. S. P. (1997). Vocabulary and teaching. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 238-254). Cambridge University Press.
- Nordlund, M., & Norberg, C. (2020). Vocabulary in EFL teaching materials for young learners. *International Journal of Language Studies*, 14, 89-116.
- Reda, G. (2003). English coursebooks: prototype texts and basic vocabulary norms. *ELT Journal*, 57(3), 260-268. <https://doi.org/10.1093/elt/57.3.260>
- Sanchez, A., & Criado, R. (2012). Lexical frequency, textbooks and learning from a cognitive perspective. A corpus-based sample analysis of ELT materials. *RESLA*, 2012, 77-94.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research - LANG TEACH RES*, 12, 329-363.
<https://doi.org/10.1177/1362168808089921>
- Soars, J., & Soars, L. (2007). *New Headway: Pre-Intermediate Third Edition: Student's Book: Six-level general English course for adults*. OUP Oxford.
<https://books.google.com.tr/books?id=rLuwSAAACAAJ>
- Thornbury, S. (2002). *How to teach vocabulary / Scott Thornbury*. Longman.

- Tyler, A. (2012). *Cognitive Linguistics and Second Language Learning*.
<https://doi.org/10.4324/9780203876039>
- Ünlü, F. (2012). *Loading, distribution, and repetition patterns of the 2000 high frequency words of general English in an EFL coursebook* [Karadeniz Technical University,]. Trabzon.
- Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15.
- Webb, S. (2007). The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics*, 28(1), 46-65. <https://doi.org/10.1093/applin/aml048>
- West, M. (1953). *A General Service List of English Words*. Longman, Green & Co.