APPLYING CORPUS LINGUISTICS TO ENGLISH TEXTBOOK EVALUATION: A CASE IN VIET NAM

Huynh Thi Thu Nguyet[1*], Nguyen Van Long[2]

1. Department of English - National Taiwan Normal University
No. 162, Section 1, Heping East Road, Da'an District, Taipei City, 106, Taiwan
2. University of Foreign Language Studies - The University of Da Nang
No. 131, Luong Nhu Hoc Street, Cam Le District, Da Nang, Viet Nam

Abstract: Looking at textbook evaluation from a corpus linguistics perspective, this paper compares two sets of textbooks used at senior high school in Vietnam and evaluate the effectiveness of the new one, centering on lexical resources at word level, particularly individual words and phrasal verbs. As for the comparison of the wordlist in general, the two corpora, taken from the two sets of textbooks, were analysed by Antconc software to extract the wordlist, then the two wordlists are compared by Venny 2.1.0 to see the similarities and differences. The research reveals a quantifiable evaluation of the lexical resources, tapping into the mutual and exclusive words, as well as examining lexical complexity of the two sets of textbooks. Unlike conventional textbook reviews focusing on grammar, this study is one of the first attempts to evaluate textbooks efficiency from corpus linguistics perspective, which in turn contributes to the improvement of the current English textbooks in Viet Nam, as well as a source of consideration for curriculum design worldwide.

Keywords: Corpus linguistics, textbook evaluation, lexical resource, phrasal verb, word complexity.

1. Introduction

In the era of educational reform since 2000, the National Foreign Languages Project 2020 was enforced from 2008 in order to enhance English competence of Vietnamese. It provides comprehensive actions to obtain its goals, such as establishing new benchmarks for teachers' language proficiency, training and retraining teachers, applying new teaching methodologies, introducing a new set of English textbooks (Prime Minister, 2008). The effectiveness of this project is still insignificant as there have been numerous shortcomings in planning and implementation. Therefore, the government must adjust the plan and extend it to 2025 (Prime Minister, 2017).

In the light of this Project, since the school year 2019, the new set of textbooks has been officially used in general education to replace the old one after five years of pilot implementation. Textbooks play a vital role in classrooms as they provide input into lessons in the form of texts, activities, explanations, etc., which are beneficial to both teachers and students in teaching and learning process (Harmer, 2007; Hutchinson & Torres, 1994). While there have been numerous studies evaluating textbooks used in general education from various perspectives in other countries (Kornellie, 2014; Litz, 2005; Quero,

* Tel: +886-928-370439, Email: httnguyet@ufl.udn.vn

2017), this field of research is still in its infancy in Viet Nam. Although the Ministry of Education and Training (MOET) has called for feedback from both experts and practitioners on the use of textbooks, the comments are quite subjective which are mostly limited to discussion in newspapers or at workshops. Similarly, research on book review in Viet Nam just pays attention to grammar or tasks (Ngo & Luu, 2018) instead of lexical resources. Given that Corpus linguistics is quite novel in Vietnamese context, and the need for an evidence-based evaluation of the new English textbooks, this small-scale study is conducted to compare the two sets of textbooks and evaluate the efficacy of the new one by employing corpus linguistics' approach, focusing on lexical resources at word level, particularly individual words and phrasal verbs. The goal of this study is to provide a quantitative evaluation of the lexical resources, which can contribute to the improvement of the current English textbooks.

2. Literature review

2.1. A Corpus-based approach to Language Planning Policy (LPP)

Language planning today mainly focuses on three major aspects, which are status planning, corpus planning, and acquisition planning. The earliest reference to status and corpus planning was made by Heinz Kloss in 1969 while acquisition planning was introduced by Cooper in 1989 (as cited in Hornberger, 2006). Hornberger (2006) refers to these major aspects of language planning:

> We may think of status planning as those efforts directed toward the allocation of functions of language/literacies in a given speech community, corpus planning as those efforts related to the adequacy of the form or structure of languages/ literacies; and acquisition planning as efforts to influence the allocation of users or the distribution of languages/literacies, by means of creating or improving opportunity or incentive to learn them or both. (p. 28)

Types	Policy planning approach (on form)	Cultivation planning approach (on function)
Status planning (about uses of language)	Officialization Nationalization Standardization of status Proscription	Revival Maintenance Spread Interlingual communication – international, intranational
Acquisition planning (about users of language)	Group Education/School Literary Religious Mass media Work	Reacquisition Maintenance Shift Foreign language/second language/literacy
	Selection Language's formal role in society *Extra-linguistic aims*	**Implementation** Language's functional role in society *Extra-linguistic aims*
Corpus planning (about language)	Standardization of corpus Standardization of auxiliary code Graphization	Modernization (new functions) Lexical Stylistic Renovation (new forms, old functions) Purification Reform Stylistic simplification Terminology unification
	Codification Language's form *Linguistic aims*	**Elaboration** Language's functions *Semi-linguistic aims*

Figure 1: Language Policy and Planning Goals: An Integrative Framework (Hornberger, 2006)

Corpus linguistics data is generally defined as a body of naturally occurring texts that is (a) representative of a specified type of language; (b) relatively large in terms of word count; and (c) machine-readable (Fitzsimmons-Doolan, 2015, p. 107). Corpus linguistics studies are those that 'analyze corpus linguistics data by applying both quantitative and qualitative techniques to the analysis of textual patterns using computers' (Fitzsimmons-Doolan, 2015, p. 107). Though corpus linguistic approaches are being applied to an increasing number of areas of linguistic study at an escalating pace (Baker, 2009, 2010), exceptionally few Language Planning Policy studies have employed corpus linguistics approaches. In Vietnam, corpus linguistics is still in its infancy, and its application in foreign language planning policy is not academically documented.

2.2. National Foreign Languages Project 2020 and Textbooks innovation

The National Foreign Languages Project 2020 (NFLP), which has been recently renamed just as The National Foreign Languages Project, was enacted by Decision 1400/QĐ-TTg dated 30th September 2008, whose goals are:

> by 2020 most Vietnamese students graduating from secondary, vocational schools, colleges and universities will be able to use a foreign language confidently in their daily communication, their study and work in an integrated, multi-cultural and multilingual environment, making foreign languages a comparative advantage of development for Vietnamese people in the cause of industrialization and modernization for the country. (Prime Minister, 2008)

The general goals of the Project include to thoroughly renovate the tasks of teaching and learning foreign languages within the national education system, and to apply a new program on teaching and learning foreign languages at every school, level and training degree, which aims to achieve by the year 2025 a vivid progress on professional skills, language competency for human resources, especially at some prioritized sectors (Nguyen, 2013). This will enable them to be more confident in communication, further their chance to study and work in an integrated and multi-cultural environment with a variety of languages. The goals also make using foreign languages as an advantage for Vietnamese people, serving the cause of industrialization and modernization for the country (Nguyen & Ngo, 2018). According to Nguyen and Ngo (2018), the decision is the basis for comprehensively reforming basic education, improving the structure of the national education system; consolidating the teacher training system, innovating comprehensive contents and training methods, implementing preferential policies for the physical and spiritual motivation for teachers and education managers; innovating content, teaching methods, examinations; investigating and evaluating the quality of education; expanding and improving the efficiency of international cooperation in education, developing and application of educational methods of some advanced education systems.

In the framework of NFLP, high school students, upon their completion of general education, must achieve level 3 of English, which is relevant to level B1 of CEFR, and acquire approximately 2500 English words. To achieve the goals, MOET applied a systematic change in the general curriculum. English is taught from grade 3 to grade 12, accompanied by a new set of textbooks.

It follows the systematic and theme-based curriculum approved by the Minister of Education and Training (MOET, 2012). The aim of this set of textbooks is to develop students' communicative competence, therefore it leaves more room for speaking and listening skills than the old set published in 1992. Instead of offering only one volume for each grade as the old set, each grade of the new set consists of two volumes. There are 24 reading texts per level in the new set of textbooks, while the old English textbooks just offer only 16 reading texts for each grade.

In general, textbooks play an important role in the process of education because it is the main source of medium of instruction. Tollefson and Tsui (2018) intensified the importance of resources in language education and the necessity of state intervention in textbook design to support the ongoing programs for linguistic minority communities. They also put the choice of language of instruction in the central position amongst other pedagogical questions. In foreign language learning and teaching, textbooks also play a crucial part. In many instructional contexts, they constitute the syllabus teachers are inclined (or expected) to follow. Furthermore, exams are often based on textbook content (Harwood, 2010). In addition, in Vietnam, English textbooks used in the general education system are designed, evaluated and implemented homogeneously across the nation. Besides, Vietnamese teachers' traditional and linear conceptualization of literacy and language learning is shaped by the national ideologies of literacy teaching (Nguyen & Bui, 2016). These ideologies often convince teachers that teaching resources and strategies (in this case, for teaching English) may only be drawn from textbooks. Another guidance for teachers published in 2017 by MOET

also emphasized that teachers must follow textbooks' contents (MOET, 2017). Therefore, the linguistic resources provided by textbooks are especially important in the Vietnamese context. Notwithstanding its importance, there have been very few academic evaluations of the new set of textbooks after five years of implementation. Dang and Seals (2018) evaluated English textbooks in Vietnam from a sociolinguistic perspective, focusing on four main sociolinguistic aspects: teaching approach, bilingualism, language variations, and intercultural communication reflected in the primary English textbooks. However, they just examined English textbooks for primary schools. There have been no synthesis evaluations of the whole set, and an approach from a corpus linguistics perspective is still missing in the process.

2.3. Phrasal Verbs

Phrasal verb, like collocation or n-gram, is a type of formulaic language. It is a multi-word verb which consists of a verb and a particle and/or a preposition to form a single semantic unit. It is considered to be problematic because the meaning of this unit cannot be understood based on the meanings of the constituents. Instead, learners must take the whole unit to understand. Therefore, the meanings of PVs are quite unpredictable (Huddleston & Pullum, 2002, p. 273) and they have to be 'acquired, stored and retrieved from memory as a holistic unit' (Wray & Michael, 2000). Moreover, some phrasal verbs carry more than one meaning. Gardner and Davies (2007) found that each of the most frequent English PVs had 5.6 meaning senses on average. Phrasal Verbs are important to learners of English because they appear quite frequently in the English texts. The results from a corpus search of the British National Corpus (BNC) showed that learners will encounter one PV in every 150 words of

English they are exposed to (Gardner & Davies, 2007). Vilkaitė (2016) study investigated the frequency of occurrence of four categories of formulaic sequences: collocations, phrasal verbs, idiomatic phrases, and lexical bundles. Together the four categories made up about 41% of English, with lexical bundles being by far the most common, followed by collocations, idiomatic phrases, and phrasal verbs.

The complexity of formulaic language and the barriers it causes which prevent learners from achieving native-like level are well documented. Ellis, Simpson-Vlach, and Maynard (2008) investigated how the corpus-linguistics metrics of frequency and mutual information (MI) are represented implicitly in native and non-native speakers of English, and how it affects their accuracy and fluency of processing of the formulas of the Academic Formulas List (AFL). Durrant and Schmitt (2009) extracted adjacent English adjective-noun collocations from two learner corpora and two comparable corpora of native student writing and calculated the t-score and MI score in the British National Corpus (BNC) for each combination extracted. Hinkel (2002) showed that L2 writers' texts had fewer collocations than those from L1 writers. Verspoor and Smiskova (2012) provided a typology for chunk use in L2 language and show that the more L2 input learners receive, the more, and longer, chunks they use. Similarly, a study by Verspoor, Schmid, and Xu (2012) showed that more advanced learners will use more words with targets like collocations. As for phrasal verb itself, Schmitt and Redwood (2011) examined whether English-Language Learners' knowledge of phrasal verbs is related to the verbs' frequency in the BNC. The results revealed a significant positive correlation: on the whole, the more frequent the phrasal verb, the higher the performance of learners. Hundt and Mair (1999) explored

text frequencies of phrasal verbs with 'up'. The results turned out that in press writing, both the type and token frequency of phrasal verbs have increased between the 1960s and the 1990s. By contrast, in academic writing, type and token frequencies were rather stable or even decreasing.

The difficulties of phrasal verbs seem to be intensified to Vietnamese learners of English as they do not appear in this language. Therefore, to Vietnamese learners, there is a need to induce their attention to this crucial part of speech in the teaching process. Given the lack of a corpus-based evaluation of textbook in Viet Nam, the absence of phrasal verbs in Vietnamese, this study focuses on comparing the two sets of textbooks at the lexical level, and pay much attention to phrasal verbs to evaluate the differences as well as the improvement of the new textbooks at the word level. Therefore, the research question for this research is:

What are the differences regarding the lexical profile in the two sets of textbooks?

3. Methodology

3.1. Compiled Corpora

There are two compiled corpora, which comprise reading texts taken from the two sets of textbooks. Compared with the new version, the textbook for elementary school is absent in the old set, the junior textbook (from grade 6) is just an introduction to English with some simple dialogues. Regarding the high-school level (grade 10 to grade 12), both of them include four English skills. Therefore, the researcher only focused on high-school textbooks as they are more comparable. The old textbooks, which was published in 1991, are composed of 12744 tokens with 2661 types, while the new ones, which was

first introduced in 2014, have 16812 tokens altogether with 3273 types. The researcher did not include dialogues as they are spoken languages.

3.2. Method

As for the comparison of the wordlist in general, the two corpora were analysed by Antconc software (Anthony, 2019) to extract the wordlist, then the two wordlists are compared by Venny 2.1.0 (Oliveros, 2015) to see the similarities and differences. Next, the profiles of the two wordlists are compared with the New General Service List (NGSL), using lextutor.ca, to see the coverage of the vocabulary because 2800 words in the NGSL provides more than 92% coverage for learners to read most general texts of English (Browne, Culligan, & Phillips, 2013). The combination of NGSL and New Academic Word List (NAWL) also comes out with the same coverage (Browne, Culligan, & Phillips, 2013). In addition, research showed that high-frequent words should be given priority to teach first. (N. C. Ellis, Simpson-Vlach, Römer, O'Donnell, & Wulff, 2015; N. Ellis et al., 2008).

As the new English textbooks were designed so that upon completion of the general education programme, students can meet the B1 level of the Common European Framework of Reference (CEFR), the researcher also applied this framework to analyse the vocabulary profile. There are two bands in this corpus. The Waystage List is indeed the Key English Test (KET) Vocabulary List, which drew on vocabulary from the Council of Europe's Waystage (1990) specification. Its covers vocabulary appropriate to the A2 level on the Common European Framework of Reference (CEFR). The Threshold list is the Preliminary English Test (PET) Vocabulary List which covers

vocabulary relevant to the B1 level on the Common European Framework of Reference (CEFR), with reference to vocabulary from the Council of Europe's Threshold (1990) specification and other vocabulary which corpus evidence shows is high frequency.

As for phrasal verbs, the corpora were analysed by Sketchengine website with the code [tag="V.*+"] [] {0,4} [tag="RP"] to look for phrasal verbs in the compiled corpora. The extracted phrasal verbs were compared together to see the similarities and differences in terms of frequency and complexity. Regarding the frequency of PVs, the researcher referred to the PHaVe list (Garnier & Schmitt, 2014) which comprises 150 most frequent phrasal verbs and their most common meanings. These PVs cover more than 75% of the occurrences in the Corpus of Contemporary American English (COCA) so it is quite reliable to check the frequency of phrasal verbs. Concerning the complexity of the two lists, the researcher categorized them into 6 levels, ranging from A1 to C2 (CEFR) based on their classification in the English Vocabulary Profile (EVP) published by Cambridge University Press. The meaning of the Phrasal verbs varied between classes; therefore, the researcher had to look at the whole concordances to determine which level of proficiency they belong to.

4. Results

By using Venny 2.1.0, the quantitative results showed that the two sets of textbooks have 1435 mutual words, 1237 included exclusively in the old textbooks, and 1843 exclusive words of the new ones.

4.1. Word profiler

The lexical complexity of the two sets of textbooks were compared by the lexical profile

measures. When word lists were imported to *lextutor*, words were counted as *tokens*, an individual occurrence of a linguistic unit in speech or writing. Similar tokens were counted as one *type*, an abstract category, class, or category of linguistic item or unit. Therefore, the number of types analysed in *lextutor* was fewer than the number of tokens we had got from Venny programme. The mutual word list and exclusive word list were classified into frequency bands of the New General Service List (2800 words) and New Academic Word List, using lextutor.ca. On the system, the

NGSL consists of 3 bands, with 1000 highest frequent words for the first band (NGSL 1), 1000 less frequent words in the second bank (NGSL 2), and 800 lowest frequent words in the third band (NGSL 3). The results are presented in Table 1. Looking at the two sets in general, it seems that although the new set has more low-frequency words and academic words than the old one, this difference is not very remarkable. The percentages of tokens included in the NGSL and NAWL are almost the same for the old and new set (95.5% and 95.2% respectively).

Table 1: Lexical profile of the old textbooks and new textbooks

NGSL+NAWL	Old textbooks		New textbooks	
Frequency band	Types	Text coverage (Tokens)	Types	Text coverage (Tokens)
NGSL 1	1524	85.6%	1708	82.9%
NGSL 2	470	5.7%	610	7.7%
NGSL 3	243	3%	254	3.2%
NAWL	98	1.2%	154	1.4%
Total NGSL + NAWL		95.5%		95.2%
OFF-List	326	4.5%	547	4.84%

However, when analyzing mutual words and exclusive words separately, the figures are slightly different. Regarding the words that two sets share in common, they account for 68.2% in the first 1000 words, 25.8% in the next 1000 words, and 6.6% in the last 800 words. Looking at the exclusive wordlists, the old textbooks cover 56.1% in the NGSL, while the new one covers 57.7%. As for the academic wordlist, the new textbooks

have more academic vocabulary, and the proportion of academic words in the list is also higher than the old one (7.2% and 5.3% respectively). Therefore, it seems that the lexical sophistication in the new set is higher than the old one, although it is not considerable. Interestingly, the old textbook has fewer words, but the proportion of off-list words is higher than the new one (Table 2).

Table 2: Lexical similarities and differences of the old textbooks and new textbooks

Frequency band	Mutual Words		Old textbook		New textbook	
	Types	Text coverage (Tokens)	Types	Text coverage (Tokens)	Types	Text coverage (Tokens)
NGSL 1	973	68.2%	299	24.2%	463	25.2%
NGSL 2	226	25.8%	244	19.8%	418	22.7%
NGSL 3	94	6.6%	149	12.1%	180	9.8%
NAWL	32	2.2%	66	5.3%	133	7.2%
OFF-List	102	7.14%	477	38.59%	646	35.09%

Concerning their coverage in the CEFR list framework, the reading texts in the old English textbooks seem to have a higher proportion of vocabulary covered in the list, with 84.7% although they have fewer word counts. The new set offers longer texts in total, but the coverage is slightly lower (80.2%). More importantly, the new textbooks have a higher percentage of off-list words compared with the old ones (19.82% and 15.25% respectively) (Table 3). There might be a question about the complexity of the off-list

words. In other words, there is a chance for the off-list words in one set of textbooks to be more advanced than the other. Nevertheless, even when the researcher analysed the off-list vocabulary with reference to NGSL and NAWL, the complexity of the two unclassified wordlists are almost the same across levels (Table 4). Therefore, it can be said that the new set does not make significant progress in providing learners with appropriate vocabulary in response to CEFR benchmarks.

Table 3: Lexical profile of the old textbooks and new textbooks with reference to CEFR list

CEFR	Old textbooks		New textbooks	
Frequency band	Types	Text coverage (Tokens)	Types	Text coverage (Tokens)
List 1 (waystage)	1140	77.2%	1204	71.7%
List 2 (Threshold)	405	7.5%	492	8.5%
Total List 1 + List 2		84.7%		80.2%
OFF-List	1116	15.25%	1576	19.82%

Table 4: Lexical profile of the off-list words from CEFR framework in the two sets with reference to NGSL + NAWL list

OFF-list	Old textbooks		New textbooks	
Frequency band	Types	Text coverage (Tokens)	Types	Text coverage (Tokens)
NGSL 1	307	27.4%	428	27.1%
NGSL 2	249	22.2%	359	22.7%
NGSL 3	156	13.9%	196	12.5%
NAWL	74	6.6%	135	8.5%
Total NGSL + NAWL		70.1%		70.8%
OFF-List	334	29.82%	461	29.20%

4.2. Phrasal verbs

As a whole, there are 34 types of phrasal verbs (PVs) in the old set of textbooks, occurring 41 times in the entire texts because some of them occur more than once in the reading texts, such as: *pick up, go out, carry out, take up, open up, clean up, get up.* 19 of them are included in the PHaVE list (Garnier & Schmitt, 2014) as the most frequent phrasal verbs, the remaining ones are classified as

off-list PVs (Table 5). In most cases, the PVs reserve the consistent meanings when they reoccur, except for the verb 'open up'. This verb is presented with two different meanings: It means '*open*' in the concordance '… During a maths lesson, she raised both arms and **opened up** her fingers one by one until all ten stood up…'; while it carries the meaning '*Make STH become available or possible*' in the concordance '…They know a new world is **opening up** for them…'

Looking at the sophistication of the PVs list in the old English textbooks, 25 out of 34 of the total PVs are classified by EVP. Specifically, there is 1 PV in A1 level (*go out*), 4 PVs in A2 level (*pick up, find out, get off, turn on*), 7 PVs in B1 level (*carry out, take up, put down, throw away, take out, go down*), 9 PVs in B2 level (*go on, set up, make up, stand up, open up, go off, cut down, live on, get up*), 2 PVs in C1 level (*build up, come up*) and 2 PVs in C2 level (*wipe out, lead up*). There are 9 unclassified PVs, which are: *jump up, lift off, jot down, wash away, speed up, carry along, run off, clean up, lay down.* In general, from the result, it seems that most PVs fall in the intermediate level (B1 and B2 levels).

Table 5: List of PVs in the old English textbooks

On-list PVs				Off-list PVs	
Order of frequency	PVs (occurrence times)	Order of frequency	PVs (occurrence times)		
1	go on	36	carry out (x2)	get up (x2)	throw away
2	pick up (x2)	41	take up (x2)	turn on	jump up
4	come up	48	open up (x2)	speed up	lift off
6	find out	56	get off	live on	jot down
8	go out (x2)	58	put down	carry along	wipe out
11	set up	60	go off	run off	lead up
17	make up	65	clean up (x2)	take away	wash away
24	take out	84	build up	cut down	
26	go down	109	lay down		
30	stand up				

In the new set of English textbooks, there are also 34 types of phrasal verbs, but they occur 44 times in the texts. 5 out of 7 of the re-occurred PVs are included in the PHaVE list, which means they are high frequent verbs and should be paid attention to (Table 6). However, less than half (16/34) of the PVs in the new textbooks appear in the PHaVE list. In other words, most of them are infrequent PVs. Regarding their classification in the CEFR levels, there is one PV in A1 level (wake up), 1 PV in A2 level (grow up), 10 PVs in B1 level (give up, look up, set out, fill up, set up, move out, carry out, bring up, hand out, go up), 10 PVs in B2 level (make up, try out, help out, read out, cut down, slow down, pay off, heat up, come up (with), keep up (with)), 2 PVs in C1 level (move on, build up), and no PV in C2 level. There are up to 10 unclassified PVs, which are go along, sweep out, start up, drop out, save up, wash away, dress up, emerge out, move around, get out. In short, the PVs introduced in the new textbook also focus on intermediate levels (B1 and B2 levels), but their distribution between levels is not as equitable as the old list.

Table 6: List of PVs in the new English textbooks

On-list PVs				Off-list PVs	
Order of frequency	PVs (occurrence times)	Order of frequency	PVs (occurrence times)		
4	come up (with)	35	wake up	try out	wash away
10	grow up (x2)	36	carry out (x2)	sweep out	set out (start)
11	set up (x3)	45	bring up	start up (x2)	fill up

13	get out	50	move on	help out	drop out (x2)
16	give up	68	slow down (x4)	save up	dress up
17	make up	78	pay off	read out	cut down
20	look up	131	move out (x2)	emerge out	heat up
33	go up	144	go along	hand out	move around
				keep up (with)	build up

Comparing the frequency of the two lists, there are only 14/34 PVs in the new textbooks appearing in the top 100 most frequent PVs, which account for 51.4% of all PV occurrences in BNC corpus (Gardner & Davies, 2007), whilst this number in the old one is 18/34 (Table 5 and Table 6). Therefore, it can be said that in terms of level of frequency, the quality of the Phrasal verbs in the old textbooks outweighs the new ones, although the new set has a greater number of PV occurrences.

When it comes to mutual Phrasal Verbs, the two lists have 7 mutual PVs altogether, however, 3 of them are infrequent PVs. Most mutual PVs share similar meanings in two sets, except the verb 'come up'. In the old textbook, it carries a C1 level's meaning, while in the new set, it comes with a B2 level's meaning. The phrasal verbs and their relevant concordances are presented in Table 7 as follows:

Table 7: Meanings of mutual PVs in two sets of English textbooks

PVs	Old textbooks	New textbooks
come up	**to happen, usually unexpectedly (C1 level)** […We often share our feelings, and whenever problems **come up**, we discuss them frankly and find solutions quickly…]	**(+*with*) Bring forth or produce (B2 level)** […They may work somewhere abroad, or speak to career advisers who can help them **come up** with a plan…]
carry out	**Put into execution** […One of the most important measures to be taken to promote the development of a country is to constantly **carry out** economic reforms…] […Vietnam **carried out** an intensive programme for its athletes…]	**Put into execution** […they design and **carry out** project aim to reduce fossil fuel consumption, find renewable fuel for public transport, and promote other clean air efforts…] […Many sea turtle natural reserves have been set up in Terengganu, Pahang, Sabah and other places to **carry out** different projects to save the species]
set up	**Establish or create STH** […Many organizations have been **set up** and funds have been raised…]	**Establish or create STH** [It was **set up** in 1961, and had its operations in areas such as the preservation of biological diversity, sustainable use of natural resources, the reduction of pollution, and climate change…] […After graduation, he **set up** his own medical practice…] […Many sea turtle natural reserves **have been set up** in Terengganu, Pahang, Sabah and other places to carry out different projects to save the species.]

make up	**Form the whole of an amount or entity** […It is **made up** of the following subjects…]	**Form the whole of an amount or entity** […women **make up** 47 per cent of the British workforce]
cut down	**Use a sharp tool such as a knife to break the surface of something** […They are changing weather conditions by **cutting down** trees in the forests…]	**Use a sharp tool such as a knife to break the surface of something** […Another reason for the temperature rise **is the cutting down** of forests for wood, paper or farming…]
build up	Increase or cause STH to increase […A great deal of excitement still **builds up** well before Tet…]	**Increase or cause STH to increase** […it brings individuals knowledge, enhances their existing skills, stimulates learning experiences, and **builds up** systems of values…]
wash away	**Remove or carry STH away** […the Indonesian Red Cross Headquarter in Banda Aceh was **washed away**, but a temporary office was in place within hours…]	**Remove or carry STH away** […The self-cleaning glass window and the fabric used to make umbrellas are both inspired by the smooth leaves of a lotus plant, with their ability to **wash away** dirt in the rain…]

5. Discussion

The study analysed, compared and contrasted the lexical resources of the two sets of textbooks in Viet Nam. In general, although the quantity features of the new set are greater than the old one, it does not guarantee a remarkable improvement in the lexical resources. With regards to lexical complexity, it does not considerably surpass the lexical resources provided by the old set published more than 20 years ago. As for formulaic language, or Phrasal verbs in particular, the new set has not paid much attention to their complexity and diversity. One possibility is that the new set focuses more on communicative competences in the light of communicative language teaching, rather than grammar-translation methods. Thus, its content has more improvement in speaking and listening activities, which results in the maintained, or even worse, quality of lexical resources.

With reference to the CEFR list framework, the new set of English textbooks fails to improve the opportunities for learners to learn target words and phrases which they may have to encounter in examinations at A2 or B1 levels. This is a considerable shortcoming as research showed that the more words being exposed in the learning process, the more likely they are acquired (Cobb & Boulton, 2015).

6. Conclusion

In conclusion, this study provides an evaluation of the current high school English textbooks in Vietnam from Corpus linguistics perspective, an uncommon approach to Language Planning Policy, and also a especially new approach in the Vietnamese context. It can shed light on the improvement and considerations in regard to lexical resources included in the reading texts. The results of this study are just limited to the comparison of the two sets at word level. Nonetheless, to some extent, it still provides an insight into the alliance of quantity and quality of the newly implemented textbooks. The results from this study can serve as evidence for the shortage of target words and a necessity to update the textbooks with more advanced and appropriate

vocabulary, which can help enhance learners' success in English proficiency tests. However, the efficacy of learning material is a combination of different factors. In this regard, there is an urgent need to conduct a more academically comprehensive evaluation of the new set in all levels of education as well as from different linguistic perspectives, so as to improve it punctually.

References

Vietnamese

Bộ Giáo dục và Đào tạo. (2012). *Quyết định về việc ban hành chương trình giáo dục phổ thông môn tiếng Anh thí điểm cấp Trung học Phổ thông* (No. 5209/ QĐ-BGDĐT). Hà Nội.

Bộ Giáo dục và Đào tạo. (2017). *Công văn về việc hướng dẫn thực hiện chương trình giáo dục phổ thông hiện hành theo định hướng phát triển năng lực và phẩm chất học sinh từ năm học 2017-2018* (No. 4612/ BGDĐT-GDTrH). Hà Nội.

Thủ tướng Chính phủ. (2008). *Quyết định về việc phê duyệt Đề án "Dạy và học ngoại ngữ trong hệ thống giáo dục quốc dân giai đoạn 2008 - 2020"* (No. 1400/QĐ-TTg). Hà Nội: Văn phòng Chính phủ.

Thủ tướng Chính phủ. (2017). *Quyết định về việc phê duyệt điều chỉnh, bổ sung Đề án dạy và học ngoại ngữ trong hệ thống giáo dục quốc dân giai đoạn 2017 - 2025* (No. 2080/QĐ-TTg). Hà Nội: Văn phòng Chính phủ.

English

Anthony, L. (2019). AntConc [computer software]. Retrieved from https://www.laurenceanthony.net/software

Baker, P. (2009). *Contemporary corpus linguistics*. New York: Continuum.

Baker, P. (2010). *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.

Browne, C., Culligan, B., & Phillips, J. (2013). The New General Service List. Retrieved from http://www.newgeneralservicelist.org

Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 478-497). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139764377.027

Dang, T. C. T., & Seals, C. (2018). An evaluation of primary English textbooks in Vietnam: A sociolinguistic perspective. *TESOL Journal, 9*(1), 93-113. https://doi.org/10.1002/tesj.309

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations. *International Review of Applied Linguistics, 47*(2), 157-177.

Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*, 375-396.

Ellis, N. C., Simpson-Vlach, R., Römer, U., O'Donnell, M. B., & Wulff, S. (2015). Learner corpora and formulaic language in second language acquisition research. In F. Meunier, G. Gilquin, & S. Granger (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 357-378). doi: 10.1017/CBO9781139649414.016.

Fitzsimmons-Doolan, S. (2015). Applying Corpus Linguistics to Language Policy. In F. M. Hult & D. C. Johnson (Eds), *Research Methods in Language Policy and Planning: A practical guide* (pp. 107-117). Oxford, UK: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118340349.ch10

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly, 41*, 339-359.

Garnier, M., & Schmitt, N. (2014). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research, 19*(6), 645-666. https://doi.org/10.1177/1362168814559798

Harmer, J. (2007). *The practice of English Language Teaching*. Edinburgh: Pearson Education Ltd.

Harwood, N. (2010). *English language teaching materials: Theory and practice*. Cambridge: Cambridge University Press.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features*. Mahwah: Lawrence Erlbaum.

Hornberger, N. H. (2006). Frameworks and models in language policy and planning. In T. Ricento (Ed.), *An Introduction to Language Policy: Theory and Method* (pp. 24-41). Oxford: Blackwell Publishing.

Huddleston, R. D., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. New York: Cambridge University Press.

Hundt, M., & Mair, C. (1999). "Agile" and "uptight" genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics, 4*(2), 221-242.

Hutchinson, T., & Torres, E. (1994). The textbook as agent of change. *ELT Journal, 48*(4), 315-328.

Kornellie, L. R. (2014). Inclusion of non-native English literatures in English textbooks toward critical and multicultural literacy. *Asian EFL Journal Professional Teaching Articles, 79*, 4-23.

Litz, D. (2005). Textbook evaluation and ELT management: A South Korean case study. *Asian EFL Journal, 48*(1), 1-53.

Ngo, T. H. T., & Luu, Q. K. (2018). A review on

grammar teaching of English textbook entitled "Skillful: Listening and speaking, student's book pack 2". *Social Science and Humanities Journal*, *2*(12), 764-775.

Nguyen, H. T. M., & Bui, T. (2016). Teachers' agency and the enactment of educational reform in Vietnam. *Current Issues in Language Planning, 17*(1), 88-105. https://doi.org/10.1080/14664208.2016.1125664

Nguyen, L. V. (2013). *Integrating computer-mediated communication into foreign language education: A Vietnamese sociocultural context of higher education*. Munich: Lincom Academic Publishers.

Nguyen, L. V., & Ngo, Q. M. H. (2018). English as a medium of instruction: A case study at a gifted high school in Vietnam. *The Journal of Asia TEFL, 15*(4), 1083-1102.

Oliveros, J. C. (2015). Venny (Version 2.1.0) [Software]. Available from http://bioinfogp.cnb.csic.es/tools/venny/index.html

Quero, B. (2017). A corpus comparison approach for estimating the vocabulary load of medical textbooks using the GSL, AWL and EAP science lists. *TESOL International Journal*, *12*(1), 177-192.

Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In F. Meunier,

S. De Cock, G. Gilquin & M. Paquot (Eds.), *A taste for corpora* (pp. 173-208). Amsterdam: John Benjamins.

Tollefson, J. W., & Tsui, A. B. M. (2018). Medium of instruction policy. In J. W. Tollefson & M. Pérez-Milans (Eds.), *The Oxford Handbook of Language Policy and Planning* (pp. 257-279). New York: Oxford University Press.

Verspoor, M. H., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, *21*(3), 239-263. https://doi.org/(...)6/j.jslw.2012.03.007

Verspoor, M. H., & Smiskova, H. (2012). Foreign Language writing development from a dynamic usage based perspective. In R. Manchón (Ed.), *L2 writing Development: Multiple perspectives* (pp. 17-46). Berlin: De Gruyter Mouton.

Vilkaitė, L. (2016). Formulaic language is not all the same: Comparing the frequency of idiomatic phrases, collocations, lexical bundles, and phrasal verbs. *Taikomoji Kalbotyra*, (8), 28-54.

Wray, A., & Michael, R. P. (2000). The function of formulaic language: An integrated model. *Language and Communication*, *20*(1), 1-28.

ỨNG DỤNG NGÔN NGỮ HỌC KHỐI LIỆU VÀO VIỆC ĐÁNH GIÁ SÁCH GIÁO KHOA TIẾNG ANH Ở VIỆT NAM

Huỳnh Thị Thu Nguyệt[1], Nguyễn Văn Long[2]

1. Đại học Sư phạm Quốc gia Đài Loan
162 Hoà Bình Đông, Đại An, Đài Bắc, Đài Loan
2. Trường Đại học Ngoại ngữ - Đại học Đà Nẵng
131 Lương Nhữ Học, Cẩm Lệ, Đà Nẵng, Việt Nam

Tóm tắt: Nghiên cứu sử dụng ngôn ngữ học khối liệu vào việc so sánh hai bộ sách giáo khoa tiếng Anh trung học phổ thông ở Việt Nam. Khác với những nghiên cứu đánh giá sách giáo khoa truyền thống vốn tập trung chủ yếu vào ngữ pháp, nghiên cứu này tập trung đánh giá mức độ hiệu quả của bộ sách mới ở cấp độ từ vựng, cụ thể là từ và cụm động từ đặc ngữ (phrasal verb). Kho ngữ liệu lấy từ bài đọc của hai bộ sách được xử lý bằng phần mềm Antcont để tạo danh sách từ vựng, sau đó hai danh sách này được so sánh bằng phần mềm Venny 2.1.0. Kết quả phân tích dữ liệu có thể dùng để đánh giá chất lượng từ vựng của hai bộ sách, cụ thể là những từ vựng giống và khác nhau, cũng như phân tích so sánh độ khó của từ. Có thể xem đây là một trong những thử nghiệm đầu tiên trong việc sử dụng ngôn ngữ học khối liệu vào việc đánh giá sách giáo khoa tiếng Anh tại Việt Nam. Kết quả nghiên cứu hy vọng góp phần vào việc cải thiện chất lượng sách giáo khoa tiếng Anh, cũng như cung cấp thêm dẫn chứng khoa học cho nghiên cứu trong lĩnh vực thiết kế giáo trình.

Từ khóa: ngôn ngữ học khối liệu, đánh giá sách giáo khoa, từ vựng, cụm động từ đặc ngữ, độ khó của từ.