Opportunities to learn high-frequency and newly-introduced words in Japanese and Taiwanese Senior High-school EFL textbooks: A comparative study

Tomohisa Hirano

A thesis submitted for the degree of Doctorate of Philosophy in Applied Linguistics Department of Language and Linguistics University of Essex

Department/School of Language & Linguistics
University of Essex

January 2024

Abstract

This study compares and contrasts senior high school EFL textbooks made in Japan and Taiwan in terms of the vocabulary learning opportunities they provide, focusing on high-frequency and newly-introduced lemmas. This examination is warranted because of the importance of vocabulary development in the process of learning English and the fact that textbooks serve as the primary source of language input for EFL students. Taiwan has received a significantly higher ranking than Japan in international English proficiency examinations (e.g., TOEFL), despite the fact that English is a foreign language in both countries and that the Japanese and Chinese languages are both historically distant from the English language. The findings may potentially make a valuable contribution to EFL textbook development in Japan.

The present study addresses two sets of research questions. The first set of research questions examines how many high-frequency lemmas (a lemma being a headword and its most common inflections) are included in the target textbooks and which proportion of the running words in the textbooks they can account for. The second set of research questions investigates the frequency levels of newly-introduced lemmas, their repetition rate, and the extent to which they appear across the units of each textbook.

The analysis was conducted using a corpus constructed from the target textbooks as well as two frequency-based lemmatized wordlists, the new-GSL (Brezina and Gablasova, 2015) and the BNC wordlist (Kilgarriff, 2006). Findings

revealed that, in comparison to the Japanese textbooks, the Taiwanese textbooks offered their students a greater number of opportunities to learn both high-frequency and newly-introduced lemmas. The differences in the quantity of English input and the selection of target vocabulary are deliberated upon as possible factors contributing to the aforementioned differences between the Japanese and Taiwanese textbooks.

Acknowledgements

First of all, my deepest appreciation goes to my supervisor, Sophia Skoufaki. She provided invaluable assistance and comprehensive support in every aspect of my Ph.D. dissertation by using her professional experience and knowledge of the academic research and vocabulary fields. Without her meticulous and steadfast help, the completion of this research project would not have been possible.

Moreover, I would like to express my gratitude to Yi-Chia Huang. She provided Taiwanese English textbooks used in local senior high schools. Due to the impact of COVID-19, I could not purchase the textbooks locally. Thus, without the textbooks she provided, this research could not focus on differences in vocabulary learning opportunities between the Japanese and Taiwanese senior high school English textbooks.

Finally, I am deeply grateful to my parents. Because the onset of the COVID-19 pandemic coincided with the commencement of my Ph.D., I decided to conduct my Ph.D. research in my home country, Japan. Consequently, I spent almost the entire Ph.D. duration at my parents' home in Japan. Their unwavering support in every aspect during this period is undeniably the foremost reason I have been able to persist in my Ph.D. journey until today. Without their understanding and heartfelt support, completing this research would have been an insurmountable task.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Research aims and significance of the present study

This study compares English as a Foreign Language (EFL) textbooks used in Japanese and Taiwanese state senior high schools in terms of the learning opportunities they provide for high-frequency and newly-introduced lemmas. Considered from a broad perspective, this study is warranted because of the significance of vocabulary learning as an aspect of language learning (see, for example, Milton, 2009; Nation, 2013; Schmitt, 2000). Moreover, textbooks can be a primary source of language input for EFL students (Milton, 2009). More specifically, the present research is motivated by gaps in vocabulary-focused research on EFL textbooks. First, most vocabulary-focused textbook research has been conducted on international English textbooks (e.g., Eldridge and Neufeld, 2009; Hsu, 2009; Matsuoka and Hirsh, 2010). However, research on textbooks created locally has been conducted on a relatively small scale. Second, in terms of which words are focused on in textbook studies, previous research has focused on text coverage by high-frequency words because it indicates how many words in the text students need to know to comprehend it adequately – adequate comprehension is operationalised by the researchers in each study. To the best of my knowledge, few studies have examined how newly-introduced vocabulary is presented in target textbooks. However, learning new words and developing one's knowledge of words can be fostered through textbooks. Therefore, it would be beneficial to determine whether

EFL textbooks introduce words to students in ways that foster vocabulary learning, such as through sufficient repetition. Third, few studies have examined differences in the vocabulary included in locally produced EFL textbooks across nations. However, by identifying differences between locally produced EFL textbooks used in countries with similar English learning environments, it will be possible to identify possible reasons for English proficiency gaps among such countries. For example, analysing the learning opportunities for high-frequency vocabulary in EFL textbooks can indicate whether the English language learners in different EFL countries can sufficiently learn them through textbooks. Because of the significance of high-frequency words in English learning, this variation between countries may be substantial. In order to clarify potential reasons for the differences in the English proficiency levels of the countries involved, it will be beneficial to compare textbooks across EFL countries.

The EFL countries of Japan and Taiwan are compared in this study regarding the textbooks utilized by senior high schools in each locality. They share some similarities in their language environment, language distance from English and first language (L1), and English instruction systems. First, since both Taiwan and Japan are EFL environments, the amount of English input in students' daily lives may be comparable, and EFL textbooks can serve as students' primary English input source. Second, both Japanese and Chinese, which are the official languages of Japan and Taiwan, respectively, share a similar language distance from English. According to Xinxin, et al. (2022), language distance is a strong predictor of how easily a learner

will learn a new language. Generally speaking, the smaller the language distance, the easier it is to learn the target language. Third, Taiwan and Japan have similar educational systems. In both countries, there are six years in elementary school and three years in junior-high and senior high school. Additionally, English instruction begins in the third grade of elementary school in both countries. Examining the words in textbooks and those newly introduced in textbooks can help form a rounded picture of the opportunities for vocabulary learning that Japanese and Taiwanese EFL textbooks offer to learners and to discern similarities and differences between them. Eventually, the findings of this study will be applicable to the development of EFL teaching materials.

1.2 Research background

Second language (L2) learners must be exposed to a substantial amount of language input (Webb and Nation, 2017) to develop their vocabulary knowledge. However, it is often claimed that EFL students find it challenging to receive sufficient language input because the amount of English input in their daily lives or outside the classroom is likely to be relatively low. EFL students are more likely to learn the majority of their vocabulary in the EFL classroom and from textbooks (Alsaif and Milton, 2012; Bouhlal, Horst and Martini, 2018; Jordan and Gray, 2019). Therefore, the vocabulary that EFL students encounter in their textbooks can have a direct impact on the vocabulary they learn.

Consequently, one of the requirements of English textbooks, particularly EFL textbooks, is to provide students with ample opportunities for vocabulary learning. Such richness of vocabulary learning opportunities in this study will be identified by whether EFL textbooks offer EFL learners many opportunities to learn many of high-frequency words and newly-introduced lemmas. Milton (2009) argues that one's vocabulary size, or the number of words one knows, has a strong correlation with their English performance. According to him, on the CEFR level, English learners with a larger vocabulary size tend to demonstrate a higher level of English use. For EFL learners to demonstrate A2 to B1 level on CEFR, Milton (2009) states that they need to know the most frequent 2500 or more and the most frequent 3750 words and more needed to move onto CEFR level B2 to C1. Due to the importance of textbooks in EFL environments as a source of vocabulary input, textbooks must include a variety of useful vocabulary so that students can increase their vocabulary through incidental or intentional classroom learning. In addition, according to Webb and Nation (2017), it is feasible for learners to develop their vocabulary knowledge if specific conditions are fulfilled. These conditions include a sufficient repetition rate and adequate distances between occurrences of a word in their input. In order for students to effectively develop their vocabulary knowledge, it is crucial to determine whether textbooks offer rich opportunities to learn many high-frequency words and whether they introduce new words under conditions likely to promote vocabulary learning.

Additionally, it should be important to check the national curriculum

guidelines of the country where the target textbooks are used so as to ascertain whether English textbooks provide their students with useful and sufficient vocabulary input. National curriculum guidelines can play a significant role in vocabulary learning, particularly in the classroom (Milton, 2009), because they highlight specific language elements that students are expected to learn, such as vocabulary and grammar, or the teaching approach the government recommends (e.g., Communicative approach). Thus, the vocabulary that students can learn from textbooks can be influenced by national curriculum guidelines or teaching-material developers' decisions (O'Dell, 1997). Consequently, it is essential to comprehend what the Japanese English education system requires Japanese students to learn and to investigate whether this requirement can be met with the vocabulary offered in Japanese textbooks.

The English curriculum guidelines in Japan are revised each decade. It is essential to determine the degree to which Japanese English textbooks and the vocabulary students can learn from these textbooks align with the updated recommendations in the guidelines. The most recent curriculum guidelines published by The Ministry of Education, Culture, Sports, Science and Technology of Japan (henceforth, MEXT) propose three learning objectives for senior high school level English classes (MEXT, 2018):

1) "To deepen students' understanding of English sounds, vocabulary, expressions, grammar and functions, and acquire the skill of using this knowledge appropriately in actual communication through listening, reading, speaking and

writing in accordance with the purposes, scenes and situations."

2) "To foster the ability to accurately understand the overview, main points, details, speakers' or writers' intention, etc., about everyday and social topics in English, and appropriately convey and exchange information and ideas of these topics in accordance with the purposes, scenes and situations in which communication takes place."
3) "To cultivate a willingness to communicate proactively and autonomously, deepening understanding of the culture behind English and considering the listener, reader, speaker or writer."

According to these objectives, students are required to comprehend English accurately and develop the ability to communicate effectively in general English. Vocabulary learning is not the only factor relevant to achieving these objectives, but it should be an important factor, as vocabulary knowledge is required for comprehension and communication in English.

However, the Japanese national curriculum does not specify the vocabulary that high school seniors should learn from their textbooks (JACET, 2016). Although the national curriculum of Japan specifies that the English class in the first year should introduce 400-600 new words, and 700-950 in each of the second and third years, no wordlists or criteria for the selection of target vocabulary for textbooks are provided. Therefore, the answer to the question of which vocabulary students can learn for effective English communication or accurate English comprehension depends heavily on the textbook publisher. Thus, it would be beneficial to investigate

which vocabulary textbooks offer their students and which vocabulary aligns with the objectives of the Japanese national English curriculum.

In addition, the years of senior high school are crucial for students who wish to pursue higher education. The vocabulary they acquire during their senior years of high school is crucial because the vocabulary they can learn prior to entering university can serve as a foundational English skill at university. Schmitt (2008) claims that learning the most frequent words facilitates the transition to English for Academic Purposes (EAP) or English for Specific Purposes (ESP) learning. The purpose of this study is to determine whether Japanese English textbooks provide opportunities for students to acquire useful vocabulary during their senior years of high school.

1.3 Research objectives

This research project seeks to identify differences in the learning opportunities that Japanese and Taiwanese senior high school level EFL textbooks provide for the learning of high-frequency and newly-introduced lemmas. A reasonable learning objective for EFL students is to learn the vocabulary needed to comprehend the majority of the words used in the texts they encounter (Nation, 2013). Several studies suggest that learners need to know 95 or 98% of the words in a general-English (i.e., not technical) text they are reading to comprehend it adequately (e.g., Hirsh and Nation, 1992; Laufer, 1989; Nation, 2006, Schmitt et al,

2011). Thanks to the analysis of large English L1 corpora (i.e., digitized compilations of written or transcribed language), one can identify the vocabulary that English L1 speakers use most frequently in daily life. Words that are frequent in English L1 contexts are considered useful learning targets for EFL learners[1] because they occur frequently in authentic English texts. For instance, high-frequency words can account for more than 80 percent of all words in authentic English contexts (Nation, 2006). Thus, learning high-frequency words is fundamental to English learning. Thus, EFL textbooks are required to expose students to many high-frequency words. At this point, this study will reveal how many word tokens high-frequency words account for in the textbooks (word token analysis) and how many of high-frequency words are in the target textbooks (word lemma analysis).

In addition to the aforementioned analyses, by analysing the frequency levels of newly-introduced lemmas using wordlists compiled from the frequency order of English words, this study will reveal which word frequency levels in these wordlists receive the most attention in Japanese and Taiwanese textbooks. Even though high-frequency vocabulary plays an important role in English learning, they do not provide the level of coverage required for satisfactory understanding of general English texts. Therefore, this study aims to identify whether the textbooks provide their students with useful new lemmas that can help students approach the level of vocabulary coverage required for satisfactory comprehension of general

[1] They will also need to learn English words that express concepts that are frequent in their EFL contexts but not in English L1 contexts (e.g., Cheung, Chung and Skoufaki, 2010).

English text.

To examine how textbooks introduce newly-introduced lemmas, this study focuses on the repetition and recycling of them. In other words, this study examines how many times students can encounter and reencounter the target lemmas throughout each textbook's units. First, the repetition of vocabulary has been extensively discussed by vocabulary researchers, but the exact number necessary to guarantee vocabulary learning has not yet been determined. Nevertheless, researchers agree that the more opportunities students have to encounter a word, the more likely they are to learn it (e.g., Teng, 2016; Waring and Takaki, 2003; Webb, 2007). Second, to consolidate their vocabulary knowledge, students must periodically revisit the target words. Even though newly-introduced lemmas in the Japanese and Taiwanese textbooks can be repeated many times in a single unit of a textbook, students' knowledge of the lemmas they have not encountered in a long time can deteriorate (Webb and Nation, 2017). To retain and/or develop their vocabulary knowledge, they must encounter the target lemmas again in other units of a textbook. This study will examine whether the textbooks provide beneficial opportunities for learning newly-introduced lemmas by examining the repetition rate with which students encounter these lemmas and the distance between the repetitions of vocabulary throughout a textbook's units. Moreover, from the results of the vocabulary repetition and its distance among a textbook, the present study attempts to identify how many of newly-introduced lemmas are theoretically learnable for students in the textbooks.

In sum, the objectives of this study are to compare Japanese and Taiwanese textbooks in terms of the following two sets of research questions. The first set relates to analysing the occurrence of high-frequency lemmas in the textbooks. The second set relates to the introduction of new lemmas in textbooks.

The first set of research questions:

1. How much do high-frequency lemmas account for the lexical coverage in the Japanese and Taiwanese textbook series examined in this study?
2. How many of high-frequency lemmas are in the Japanese and Taiwanese textbook series examined in this study?

The second set of research questions:

3. Which frequency bands in the BNC-6318 do newly-introduced lemmas in the Japanese and Taiwanese textbooks come from?
4. How many times do the newly-introduced lemmas occur in the Japanese and Taiwanese textbooks?
5. How many newly-introduced lemmas recur across units of the Japanese and Taiwanese textbooks?
6. How many of the newly-introduced lemmas are, in theory, learnable through the Japanese and Taiwanese textbooks?

1.4 Methodology

This study uses corpus linguistics methods to examine vocabulary in EFL

textbooks. Corpus linguistics is a relatively new method, but it enables us to observe and analyse the target language in an entirely novel way (e.g., Anthony, 2017; Bibler et al, 1998; Hunston, 2002; Sinclair, 2005; Tognini-Bonelli, 2001). According to Szudarski (2018), the utilization of corpus linguistics in language pedagogy research can be classified into two distinct categories: direct and indirect application. The former is used for data-driven learning, which is defined as learners' engaging "in their own linguistic analysis in which they focus on specific linguistic features" (p.104) (Szudarski, 2018). That is, learners themselves use some corpora to analyse language features and learn them. In indirect application, he claims that "corpora are used to inform the design and development of syllabuses, tests and teaching materials" (p.96). For example, typical corpus linguistics research in this way involves the creation of frequency-based wordlists. For example, numerous frequency-based wordlists have been created from corpora such as the British National Corpus (BNC) wordlist (Nation, 2006), the British National corpus/Corpus of Contemporary American (BNC/COCA) wordlist (Nation, 2012), and the New General Service List (new-GSL) (Brezina and Gablasova, 2015). These frequency wordlists were compiled by analysing a vast amount of vocabulary used in real-world contexts such as television, radio, magazines, newspapers, etc. These wordlists serve as a starting point for learning vocabulary.

Moreover, corpus research to determine how many words learners need to comprehend authentic English texts (e.g., Harish and Nation 1992; Laufer, 1998; Schmitt et al 2011) indicates that 95-98% vocabulary coverage is required for

adequate comprehension of English texts, and the number of words learners need to know is also roughly identified (See more details in 2.2.4). Therefore, corpus research has a significant impact not only on the research field, but also on the classroom. The present study will analyse the target textbooks with the indirect application of Corpus linguistics in language pedagogy research because its aims are to examine the opportunities to learn high-frequency and newly-introduced lemmas in EFL textbooks of Japan and Taiwan.

In order to examine the vocabulary used in the Japanese and Taiwanese textbooks in the present study, a total 312,877 word token corpus was created from three Japanese textbooks and six Taiwanese textbooks. Based on the corpora, two kinds of wordlists were created (lemmatized wordlists and newly-introduced lemmas wordlists). By analysing the corpora data with some frequency-based lemmatized wordlists, the present study will answer each research questions to identify differences between the target textbooks.

1.5 The thesis outline

In addition to the present chapter, this thesis includes four chapters: Literature review, Methodology, Results and Discussions, and Conclusion. In the Literature review, Section 1 discusses the units of analysis in vocabulary research, how many words students must learn and how well they should learn them, and how students can develop their vocabulary knowledge. Section 2 reviews factors that

affect vocabulary learning. Section 3 discusses research on vocabulary in English textbooks. Section 4 explains the rationale and research questions of the present study.

Chapter 3, Methodology, provides the rationale behind the selection of Taiwan as Japan's counterpart in the comparison of textbooks and the selection of the target textbooks in this study. It then summarises how the textbook corpus and the wordlists extracted from it were created. Finally, it summarises how this study examines its research questions.

The results of this study are presented and discussed in Chapter 4, Results and Discussion. The thesis concludes with Chapter 5, where a summary of this study and recommendations for improving Japanese textbooks are provided. This chapter concludes with a discussion of the study’s limitations and avenues for future research.

Chapter 2 Literature Review

2.1 Introduction

This chapter discusses research on vocabulary learning and teaching related to the topic of this dissertation. Section 2.2 discusses definitions of ‘word’ and the nature of vocabulary knowledge and learning. Several word units, which have been often used in vocabulary research are discussed to introduce their roles in vocabulary research at Section 2.2.1, namely “token”, “word form”, “word family”, “flemma” and “lemma”. Research on the vocabulary knowledge learners need to know is discussed in Sections 2.2.2 to 2.2.6. These sections introduce receptive and productive vocabulary knowledge, the depth and size of vocabulary knowledge learners need, the relationship between the depth and size, and the roles of word frequency in English learning. Section 2.2.7 and 2.2.8 discuss research on how vocabulary can be developed and how vocabulary should be taught. In Section 2.3, the focus moves to factors that affect vocabulary development, such as the vocabulary repetition and quality of attention. Furthermore, Section 2.4 reports textbook research conducted in terms of vocabulary. This section contains the analysis of English textbooks regarding to the selection of vocabulary and vocabulary knowledge presented in textbooks. Section 2.5 explains the rationale for the present study and introduces its research questions.

2.2 Vocabulary research

2.2.1. What is a word?

In vocabulary research, "word" has been defined as “token”, “word form”, “word family”, “lemma”, and “flemma”. Depending on their research objectives, vocabulary researchers select one of these five conceptualizations of "word" as their unit of analysis (e.g., Milton, 2009; Reed, 2000).

According to Nation (2013), if words are defined as "tokens," the researcher simply counts every word form in a text. By contrast, if words are defined as "word form," then the same word is counted only once. For instance, there are eight tokens and seven word forms in the sentence "It is not easy to say it correctly”. Adopting the token as unit of analysis can be helpful when analysing the actual number of word forms in a text or the reading speed of students. Adopting the “word form” as unit of analysis can be useful when researchers examine how many words learners need to read a general-English book. However, when a study counts the number of word forms to determine how much vocabulary English language learners need to read general texts, these units are ineffective because some words do not need to be counted separately. Some irregular forms must be learnt individually, but for some vocabulary, learners can apply their knowledge of inflectional morphology and suffixes and prefixes to derive related words. Thus, the "word family", "lemma", and "flemma" are more useful units than “word form” in research for counting words in English texts.

A "word family" is comprised of a headword, its inflections (e.g., "-ed", "-ing", "-s"), and its derivations ("-ment", "-able", "-tion", "un-", etc.). Consequently, the word family "add" includes "add", "adds", "adding", "added", "addition", "additional", "additionally", etc., depending on which prefixes and suffixes a researcher assumes the learners in their study already know. "Lemma" includes a stem and its most common inflections, such as the verb "add" and "adds", "adding", and "added"." "flemma" is similar to "lemma" in that it also consists of a headword and its most common inflections. The distinction between "lemma" and "flemma" is that a "flemma" can contain multiple POS. For example, "use" is both a noun and a verb. The "lemma" for the noun “use” is different from the lemma for the verb “use” whereas both the noun and verb “use’ form a single “flemma”.

2.2.2. Vocabulary knowledge: receptive and productive

Before discussing the depth and size of vocabulary knowledge, it is important to note that vocabulary knowledge has been divided into two types of word knowledge: receptive/passive and productive/active (e.g., Laufer, 1998; Laufer and Paribakht, 1998; Henriksen, 1999; Nation, 2013; Read, 2000; Schmitt, 2014). Receptive knowledge is necessary in order to recognize the form and meaning of a word through listening or reading (e.g., Nation, 1990). Waring (1997) claims that in order to demonstrate receptive vocabulary knowledge, students must be able to provide a specific L1 translation for a L2 word so that they can demonstrate their

ability to recognize the target word and recall its meaning. Moreover, according to Webb and Nation (2017), receptive vocabulary knowledge is the ability to comprehend words in listening and reading. Nation (2013) also mentions that receptive knowledge requires learners to know what grammatical pattern the word appears in, what collocation it is most likely to be used with, how frequently the word appears in texts, and the word's association. Therefore, receptive vocabulary knowledge is required for text comprehension during reading and listening (Nation, 2020). By contrast, productive knowledge includes the ability to use a word in speaking and writing (e.g., Nation, 2013). Webb and Nation (2017) define productive vocabulary knowledge as the knowledge necessary to use a word. According to Webb (2005), productive vocabulary knowledge is the knowledge required for learners to generate a word to express their opinions or feelings to others. Thus, productive knowledge requires various kinds of knowledge, such as knowing how to pronounce or spell a word and its grammatical characteristics and collocations (e.g., Nation, 1990; Webb and Nation, 2017).

Both receptive and productive vocabulary knowledge are highly important, but research suggests that receptive knowledge is more likely to be learnt than productive knowledge (e.g., Fan, 2000; Laufer, 1998; Nation, 2013; Schmitt, 2000; Webb, 2008). The majority of vocabulary appears to be gained either through reading and listening (e.g., Nagy, Anderson and Herman, 1987). Moreover, when teachers teach words in the classroom, the methods are more likely to be receptive than productive because receptive learning tasks are easier to design, assess, and

practice (Webb, 2005). Moreover, Webb (2005) states that when students learn a word receptively, they develop receptive knowledge, whereas productive learning results in the acquisition of productive knowledge. Thus, in an environment where receptive learning is more prevalent, it seems natural that receptive knowledge can be learnt faster than productive knowledge.

Similarly, Nation (2013) also argues that it should be more challenging for learners to gain productive knowledge than receptive knowledge because of two significant differences between receptive and productive knowledge. First, learners need to know more aspects of a word for productive purposes than for receptive purposes. Second, because the majority of vocabulary is learnt through receptive learning in the classroom, there are more opportunities for students to learn receptive knowledge of a word during their learning. Therefore, while not an absolute rule, receptive knowledge is often acquired before productive knowledge.

Numerous researchers have hypothesized that more words are known receptively than productively (e.g., Fan, 2000; Laufer, 1998; Nation, 2013; Webb, 2008). Webb (2008) investigates the relationship between receptive and productive knowledge of word meaning and form. This study used tests of both receptive and productive translation to determine the difference between the two types of vocabulary knowledge. In the receptive translation test, the Japanese university student participants were required to provide an appropriate L1 translation for the target L2 word. By contrast, the productive translation test required them to write a L2 word that corresponds to the intended L1 meaning. 180 words were selected

from three distinct frequency bands in the COBUILD dictionary. Results indicate that the receptive translation test scores are significantly higher than the productive knowledge scores across all three frequency bands. Thus, in terms of meaning and form, receptive knowledge is likely to be greater than productive knowledge regardless of the frequency levels of a word. This study also demonstrates that the more receptive knowledge students possess, the more productive knowledge they can demonstrate. Therefore, Webb (2008) concludes that students are likely to have more receptive knowledge than productive knowledge, at least in terms of meaning and form. In addition to this, he suggests that receptive knowledge offers some indication of students' productive knowledge. Webb (2008) indicates that learners' receptive knowledge can be larger than their productive knowledge. However, to truly know a word, learners eventually need to possess many kinds of knowledge both receptively and productively.

2.2.3 Vocabulary depth: How well learners need to know a word

Vocabulary depth is an important aspect of vocabulary development. It refers to the extent to which students have gained vocabulary knowledge (e.g., Marzban and Hadipour, 2012; Read,1993; Yanagisawa and Webb, 2020; Qian, 1989). If learners know the form and meaning of a word, does it mean they have mastered the word? Schmitt (2000) notes that the link between form and meaning appears to be the most crucial aspect of vocabulary knowledge. Nonetheless, Nation (2013)

suggests that knowing a word involves form, meaning, and use at the most fundamental level. The framework of vocabulary knowledge proposed by Nation below, one of the most well-known frameworks, discusses which aspects of word knowledge must be learnt in order to fully know a word.

Table 2.1: What is involved in knowing a word (Nation, 2013)

Form	Spoken	R: What does the words sound like? P: How is the word pronounced?
	Written	R: What does the words look like? P: How is the word written and spelled?
	Word parts	R: What parts are recognizable in this word? P: What words parts are needed to express meaning?
Meaning	Form and meaning	R: What meaning does this word form signal? P: What word form can be used to express this meaning?
	Concept and referents	R: What is included in the concept? P: What item can the concept refer to?
	Association	R: What others words does this word make us think of? P: What other words could we use instead of this one?
Use	Grammatical functions	R: In what pattern does the words occur? P: In what patterns must we use this word?
	Collocations	R: What words of types of word occur with this one? P: What words or types or words must we use with this one?
	Constraints on use	R: Where, when and how often would we meet this word? P: Where, when and how often can we use this word?

*R=receptive, P=productive

Table 2.1 shows that to know a word, one needs to know many kinds of information

about the target words, such as their pronunciation and spelling, their stem, prefix, and suffix, their meaning, definition, and alternative words, the grammatical functions, collocations and their register and frequency. The more vocabulary knowledge learners have for a word, the deeper their vocabulary depth becomes. Thus, to know a word means that learners know these nine aspects of vocabulary knowledge both receptively and productively.

2.2.4 Vocabulary size: How much vocabulary learners need to know

Vocabulary size and reading performance are strongly connected to each other (e.g., Laufer, 1992). The vocabulary size is the number of words students know (e.g., Anderson and Freebody, 1981; Schmitt, 2014). According to Schmitt et al. (2011), readers will inevitably encounter unknown words, which can impair their reading comprehension. Therefore, the crucial question is how much vocabulary in general English texts L2 learners should know in order to comprehend these texts successfully.

Knowing 95% and 98% of the word tokens in a general English text have been proposed as optimal text coverage. Laufer (1989) examines how much vocabulary university students at the University of Haifa need to pass their reading comprehension test. The passing grade for this test is 55%. In the initial phase of this study, students were required to demonstrate reading comprehension by responding to multiple-choice and open-ended questions about the reading texts. In

addition, they were required to underline unknown words in the texts and calculate the number of unknown words and their test score. In the second stage, students were required to translate the target 40 words to determine how well they knew the words. She concluded, based on the outcomes of the two tests, that at least 95% of vocabulary coverage enabled students to pass the University of Haifa's reading comprehension test. By contrast, Hirsh and Nation (1992) suggest that 98% of vocabulary coverage is required to read non-simplified texts for pleasure. Their research aims to determine how much of a non-simplified text can be covered by the 2000 most common words and how much vocabulary is required to achieve a 98% vocabulary coverage rate. This study suggests that the fewer unfamiliar words there are in English texts, the more pleasurable the reading becomes. They illustrate the difference between the 95 and 98% vocabulary coverage per 100 tokens. While the 95% coverage contains one unknown word every two lines, the 98% coverage contains one unknown word every five lines. They recognize that this difference can significantly impact recreational reading. Thus, they consider 98% vocabulary coverage optimal.

Hu and Nation (2000) examine the differences in reading comprehension among four vocabulary coverages: 80%, 90%, 95%, and 100%. For each vocabulary coverage, a multiple-choice test and a cued written recall test are administered to assess students' reading comprehension. The multiple-choice exam consists of 14 questions with four answer options per question and assesses students' comprehension of both explicit and implicit textual information. In addition,

a cued written recall test is administered to assess students' reading comprehension. This test consists of 27 story-related questions that provide learners with opportunities to demonstrate their comprehension of the texts. For instance, some of the questions are "Who is narrating this story?", "What occurred at the very beginning of the story?" and "What followed?". In this study, they use a non-simplified fiction text with a strong chronological plot, so they note that the text is highly readable. Despite the text's high readability, Hu and Nation (2000) conclude that learners need to know 98% of the text's running words. This study defines adequate comprehension as at least 12 correct responses out of 14 questions on the multiple-choice test and approximately 70 out of 124 possible points on the cued written recall test. Even between 90 and 95% of vocabulary coverage, only a minority of learners demonstrate adequate or near-adequate comprehension by guessing some unknown words. However, none of the readers who covered 80% of the vocabulary adequately comprehended the text. These findings lead Hu and Nation (2000) to the conclusion that 98% of vocabulary coverage is required for adequate reading comprehension when reading for pleasure. This 98% vocabulary coverage has also been supported by other studies (e.g., Nation, 2006; Schmitt et al., 2011).

Regarding vocabulary size, which means how much vocabulary L2 learners need to possess to achieve the thresholds of 95 and 98% of vocabulary coverage, there have been several suggestions. Laufer (1992) investigates the amount of vocabulary that students need to know to achieve a vocabulary coverage of 95%. Two reading tests and two vocabulary size tests were administered to 92 university

students to assess their reading abilities and vocabulary size. To examine the relationship between vocabulary size and reading comprehension, subjects were divided into five groups based on their vocabulary size (below 2000, 2000, 3000, 4000, and 5000 words families). Results indicate that the group knowing 3000 words families had a reading comprehension score of 56%. Knowing 4000 word families resulted in the average score of 63%, while knowing 5000 word families resulted in the average score of 70%. Thus, this study indicates, based on these findings, that an increase of 1,000-word families correlates to a 7% improvement on a reading comprehension test. Therefore, Laufer (1992) concludes that 3000 to 5000-word families are required to achieve a vocabulary coverage of 95%.

Many researchers have made estimates of the vocabulary size necessary to achieve 98% coverage of general texts. Hirsh and Nation (1992) examine 2600, 5000, and 7000 words to demonstrate how much text these three vocabulary sizes can cover in the unabridged novels "The Pearl," "Alice in Wonderland," and "The Haunting." The results suggest that 5000 words are required in order to achieve a vocabulary coverage of 98% when reading short and non-simplified novels for pleasure. However, Nation (2006) conducted research on the vocabulary size necessary to achieve 98% vocabulary coverage for unassisted comprehension of written and spoken English, such as reading novels or newspapers and watching films. Nation (2006) suggests that if 98% vocabulary coverage is the ideal goal, then learners need to know 8000–9000 word families in order to read authentic texts without a dictionary. In addition, high-frequency words are more frequently used in

spoken language than in written language. Thus, 6000–7000 word families appeared necessary for 98% speech coverage. In sum, learners need to know 8000–9000 word families for written texts and 6000–7000 word families for spoken language to adequately comprehend it (Nation, 2006). Nonetheless, 95% of vocabulary coverage also seems reasonable if learners can consult a dictionary or a teacher (Nation, 2006).

Nevertheless, the truly optimal vocabulary coverage and size for English learning remain unknown. Laufer (1996) claims that the difference between 95 and 98% can be made up by attempting to guess a few unknown words. Schmitt et al. (2011) conclude that 98% vocabulary coverage seems to be a more reasonable goal than 95%, but they also report that there are no specific vocabulary coverages at which reading comprehension improves dramatically. Rather, they suggest that reading comprehension increases gradually as the number of learners' vocabulary coverage increases. In addition, they argue that even if learners know every word in a text, this does not guarantee that they will fully comprehend it because lexical depth and reading skills also influence reading comprehension. Therefore, attaining a certain vocabulary size to cover the ideal vocabulary coverages cannot guarantee that learners will attain a particular level of reading comprehension (Webb and Nation, 2017). Thus, learners are more likely to achieve a higher level of reading comprehension if they can cover more word tokens in a reading passage with their vocabulary size. However, it is impossible to specify the truly optimal vocabulary coverage and size targets until the relationship between vocabulary coverage and

size and reading comprehension is better understood (Schmitt et al., 2011).

In addition, recent research indicates that meaning recall is a better predictor of students' reading comprehension than meaning recognition (e.g., McLean, Stewart, and Batty, 2020; Zhang and Zhang, 2020). That is, learners must be able to recall the meaning of words to achieve adequate reading comprehension. Therefore, being able to recognise the meaning of the most frequent 3000 word families does not guarantee adequate reading comprehension. Instead, adequate reading comprehension is predicted better when students can recall the meaning of the vocabulary in their reading. Consequently, a certain vocabulary size can be useful as a vocabulary learning objectives for L2 learners, but developing recall knowledge of this vocabulary is also necessary.

2.2.5 Relationship between vocabulary depth and size in reading and listening comprehension

The connection between vocabulary size and depth remains unclear (Schmitt, 2014). This appears to be due, at least in part, to the various conceptualizations of vocabulary size and depth. While ‘vocabulary size’ is defined as the number of words a learner knows (e.g., Schmitt, 2014) and the typical target knowledge is the form-meaning connection, ‘vocabulary depth’ refers to how well a learner knows a word (e.g., Qian, 1989). The target knowledge in this case is diverse because the vocabulary depth consists of many different aspects of knowledge (See

Table 2.1). Therefore, the relationship between the two concepts can vary based on how a study defines and operationalises vocabulary depth. Schmitt (2014) states that although there are numerous definitions of vocabulary depth, including word knowledge components, lexical organization, receptive or productive mastery, and fluency, they overlap and are interrelated. It is, therefore, nearly impossible to assess vocabulary depth with a single test and vocabulary depth test scores will vary depending on which aspects of vocabulary depth are tested and how they are tested.

Nevertheless, studies have attempted to determine how the relationship between vocabulary size and depth can contribute to L2 reading or listening comprehension (e.g., Marzban and Hadipour, 2012; Mehrpour et al., 2011; Rahman and Iqbal, 2019; Teng, 2014). All of these studies indicate that both vocabulary size and depth can contribute to learners' reading and listening comprehension, but vocabulary depth appears to be a better predictor of learners' comprehension. The majority of them employed comparable techniques, such as the Vocabulary Levels Test by Nation (1990) and Schmitt et al. (2011) and the Word Association Test created by Read (1993, 1998). This may account for the similarity of the studies' outcomes. Assuming a study employs the same methodology, this may suggest that vocabulary depth is a more significant factor in learners' comprehension.

2.2.6. Word frequency

Words differ in terms of how valuable they are to learners (Webb and Nation,

2017). The value of each word is typically considered to be indicated by its frequency level because the more frequent a word is, the more likely a learner is to encounter it and need to use it. Word frequency can be a guide for learners to determine which vocabulary they need to learn first (Vilkaite-Lozdiene and Schmitt, 2020). According to Nation (2006), learners are required to learn the most frequent 8000–9000 word families to achieve 98% lexical coverage in reading and 6000–7000 word families in listening. Nation (2006) elicits these numbers of word families from the list of the 14 most frequent groups of 1000 word families in the British National Corpus wordlist (BNC). This fact illustrates the importance of learning vocabulary based on the frequency level. Words in general English have been categorized as high-frequency, mid-frequency and low-frequency. Learning high-frequency vocabulary is encouraged for L2 learners, especially at the early stages of learning (Schmitt, 2000) because learners are more likely to encounter high-frequency words than mid- and low-frequency words. Nation (2006) reports that these words can account for approximately 80% or more in both written and spoken English texts. Furthermore, Schmitt (2000) claims that learning high-frequency words allows learners to increase their comprehensible language input. Therefore, many vocabulary researchers suggest that high-frequency words deserve explicit teaching because they are the most useful words in English (e.g., Nation, 2013; Schmitt, 2000; Read, 2004; Vilkaite-Losdiene and Schmitt, 2020).

Because of the importance of high-frequency words for language learning, several high-frequency wordlists have been created to indicate the most useful

words for learners. The General Service List (GSL) created by West (1953) is the first high-frequency wordlist. The GSL contains approximately 2000 word families. These 2000 words were not necessarily the most common in English, although frequency was one of the factors in making the GSL. Since the purpose of the GSL by West (1953) was to make a wordlist of the most useful words for teachers and learners, various criteria were applied, such as ease of learning or necessity in English use (Webb and Nation, 2017).

Many studies have shown that the 2000 most frequent word families can cover a large amount of English text in many situations (e.g., Brezina and Gablasova, 2015; Hyland and Tse, 2007; Nation, 2004). However, words which are the most frequent can change with the passage of time. In fact, the GSL includes several words that used to be frequent but are not anymore, such as “telegraph” (Webb and Nation, 2017). Therefore, more recent high-frequency wordlists have been created. Three notable high-frequency wordlists are the British National Corpus wordlist (BNC) created by Nation (2006), the British National Corpus/Corpus of Contemporary American English (BNC/COCA) wordlist by Nation (2012), and the New-General Service List (New-GSL) created by Brezina and Gablasova (2015). The BNC wordlist is created from 14th 1000 word-families of the BNC to identify how many word families are needed to comprehend authentically written and spoken English using a 100 million-token corpus data that consists of 90% of written English text and 10% of spoken text in the UK. The BNC/COCA wordlist resembles the BNC wordlist, but it consists of representatives of 1-25th 1000 word families from both

spoken and written English in the UK and USA. The New-GSL was developed to replace the GSL. It contains the most frequent 2494 lemmas from 12 billion running words from four different corpora (*LOB, BNC, BE06, and EnTenTen12*[2]).

Dang and Webb (2016) compare the most frequent 2000 headwords in the GSL, BNC, COCA, and New-GSL with 18 spoken and written corpora (nine spoken corpora and nine written corpora) to identify which corpus is the best for learners. The 18 corpora represented various varieties of English used all over the world such as American-English, British-English, Canadian-English, Hong Kong-English, Indian-English and so on. The 18 corpora were anticipated to provide a comprehensive picture of the vocabulary that is crucial for L2 learners. Results indicate that the BNC/COCA wordlist shows the greatest average coverage among the four major wordlists in the corpora used in this study. The BNC/COCA list shows higher scores than the BNC wordlist in five out of the nine written corpora, despite the fact that the BNC wordlist has much larger running words in written texts than the BNC/COCA. Moreover, it shows higher scores than the New-GSL in all 18 corpora. However, since the BNC/COCA contains some words that are general but not considered high-frequency words (e.g., *hello, mama, or darling*), it is possible that the performance of the BNC/COCA is not as high as the New-GSL. In the

---

[2]LOB stands for The Lancaster-Oslo-Bergen Corpus, a 1 million words corpus made in 1961.
BNC stands for The British National Corpus, a 100 million made in 1990s.
BE06 stands for The BE06 Corpus of British English, a 1 million words corpus made in 2005-2007.
EnTenTen12: The English Web Corpus (12 billion words, from 2012) Available at https://www.sketchengine.eu/ententen-english-corpus/

comparison, focusing on only the most frequent 1000, 1500, and 1996 headwords in 18 corpora, the New-GSL shows the highest coverage among the four wordlists. Therefore, Dang and Webb (2016) conclude that the BNC/COCA wordlist tends to show the highest average coverage among the major four wordlists in spoken and written discourses, but if the coverage provided only by the most frequent words is a target, the New-GSL is the best list for learning high-frequency vocabulary. More recently, Dang et. al. (2022) conducted further research for their previous research. This study compares 78 teachers' perceptions of word usefulness and 135 Vietnamese EFL learners' vocabulary knowledge of 973 non-overlapping words between the BNC/COCA wordlist (545 words) and the New-GSL (428 words). Since these 973 words are unique to each wordlist, this study attempts to identify which wordlists contain more words that teachers consider useful and are well-known by EFL learners. This study reveals that the teachers perceive that the BNC/COCA wordlist contains more useful words for the teachers, and the words belonging to the BNC/COCA wordlist are better known by the learners. Thus, Dang and Webb (2022) conclude that the BNC/COCA wordlist is more suitable than the New-GSL for EFL learners. However, due to the limitations of this study, such as targeting only one background of teachers and learners or the different units of counting (word family for the BNC/COCA wordlist and lemma for the new GSL), they suggest that further research is needed.

Researchers have offered various operationalizations of 'high-frequency words'. Nation (2013) suggested that the most frequent 2000 word families be

labeled as high-frequent vocabulary, and the figure was widely cited in many vocabulary articles and teachers' guidebooks (e.g., Read, 2004; Schmitt, 2000; Thornbury, 2002). Dang and Webb (2016) suggest that the most frequent 1000-word families be a more suitable learning goal during initial learning. They note that beyond the 1000 most frequent word families, the number of function words (e.g., prepositions, articles, or conjunctions) considerably decreases. By contrast, the decrease in content words (nouns, verbs, adjectives, or adverbs) is more gradual. Thus, they suggest that it be more realistic for both teachers and learners to set their initial learning goal at the most frequent 1000 word families or 800 lemmas first and move onto the 2000 to 3000 word families as a learning goal in English learning. However, the most frequent 3000 word families appear to be more widely accepted as high-frequency vocabulary because the most frequent 3000 word families can often offer enough text coverage and can allow learners to communicate in a range of situations (e.g., Schmitt and Schmitt, 2014; Waring and Nation, 1997).

However, knowing high-frequency words is not enough to achieve vocabulary coverage of 98% when reading authentic general English texts. Since high-frequency words can only cover 80% of English texts in many cases, mid-frequency words are needed to achieve 98% lexical coverage. Schmitt and Schmitt (2014) suggest that the most frequent 3000 word families form the high frequency level, the next 4000 to 9000 word families form the mid-frequency level, and the rest of the word families form the low-frequency level. Schmitt and Schmitt (2014) claim that achieving 98% lexical coverage in authentic English text largely depends on

learning mid-frequency-level words. Similarly, according to Webb and Nation (2017), mid-frequency words can be essential for text coverage. Besides, according to Zipf's law (1936,1949), a tiny percentage of English words occur very frequently. Consequently, the more frequent a word is, the more likely it is to cover a significant portion of running words in texts. Mid-frequency vocabulary is not as frequent as high-frequency vocabulary, but it is still essential for comprehension and fluent expression. Thus, learning mid-frequency level vocabulary as well as the high-frequency level can help learners read authentic English texts with sufficient comprehension.

Technical vocabulary (i.e., words specific to disciplines or groups of disciplines) and general academic words (i.e., words that are not specific to certain disciplines but are used more frequently in academic than in non-academic speech and writing) are also important to learners who aim to deepen their English language skills for specific purposes. However, it might be possible for learners to more easily learn technical and academic words if they already know high- and mid-frequency words because then they can guess the meaning of technical or academic words from context. Besides, many academic words have both everyday and academic meanings (Skoufaki and Petrić, 2021). Therefore, teaching and learning high- and mid-frequency words can be an effective way to enhance learners' vocabulary development.

Schmitt and Schmitt (2014) suggest that the words beyond the most frequent 9000 word families be categorized as low-frequency words. Some low-frequency

words are infrequent in general language but are used frequently in some specialized topic areas (technical words). However, low-frequency words are very limited and rarely affect the vocabulary coverage in a text (Vilkaite-Lozdiene and Schmitt, 2020). Therefore, it would not be useful to spend a lot of classroom time on learning low-frequency words. Nation (2013) suggests that it is more effective for learners to learn word learning strategies such as guessing the meaning, drawing on word part knowledge, or using dictionaries to deal with low-frequency words than teaching them individually.

In sum, by learning high- and mid-frequency words, learners stand a high chance of having vocabulary coverage between 95% and 98% when reading general English. This foundation in vocabulary knowledge can enable learners to learn low-frequency words through vocabulary learning strategies such as guessing or using a dictionary.

2.2.7. How is vocabulary learnt?

One of the ways vocabulary learning takes place is incidentally. Webb (2020) mentions that the definition of incidental learning varies, but within Applied Linguistics, it is defined as learning as the by-product of a meaning-focused task (Chen and Truscott, 2010; Ellis, 1997, cited in Webb, 2020). Most vocabulary in the first language (L1) can be acquired incidentally thanks to the rich spoken and written language input L1 learners can daily receive (Nagy et al., 1985). In English as a

second language (ESL) settings, learners have relatively more opportunities to receive a large amount of language input to incidentally develop their vocabulary knowledge than in EFL environments. In EFL environments, however, it is very difficult for learners to receive sufficient language input through their daily lives because they do not use English outside the classroom in most of the cases. Thus, it tends to be more difficult for EFL learners to develop their vocabulary knowledge only through incidental learning.

In incidental learning, in general, the more language input learners have, the more likely they are to learn vocabulary. Therefore, researchers agree that native speakers learn the vast majority of their L1 words incidentally and faster (e.g., Nagy, Herman, and Anderson, 1985; Nation, 2013; Webb and Nation, 2017; Schmitt, 2000). Nagy, Herman, and Anderson (1985, p. 252) conclude that “the number of words to be learnt is too enormous to rely on word-by-word instruction.” In other words, it is almost impossible for teachers to teach every word learners need. Moreover, L1 learners of English can develop their vocabulary knowledge faster and acquire more vocabulary than L2 learners. Young L1 English learners can learn approximately 1,000 new word families per year (Biemiller and Slonim, 2001; Goulden, Nation, and Read, 1990). At university level, most L1 learners may know around 15,000–20,000 word families (D’Anna, Zechmeister, and Hall, 1991; Goulden, Nation, and Read, 1990). By contrast, EFL learners tend to learn fewer words per year. For example, L1 Greek EFL learners can develop 400 to 500 new lemmas per year (Milton, 2006). Primary school children in the Hungarian EFL environment learn around 300 to 400

new lemmas per year (Orosz, 2009). Webb and Chang (2012) conducted a longitudinal study to investigate Taiwanese learners' yearly vocabulary gain. Findings indicate that Taiwanese learners can learn 18 to 430 new word families in a year. Moreover, after nine years of instruction, they conclude that 47% of the learners can master the most frequent 1000 words and 16% can master the most frequent 2000 words in the Vocabulary Level Test (VLT) (Nation, 1983, 1990; Schmitt, Schmitt, and Claphan, 2001). Thus, there is a very large gap in vocabulary learning speed and attainment caused by the difference in the amount of language input received by L1 learners and EFL learners.

However, some studies indicate that incidental vocabulary learning can take place in EFL environments when learners focus on reading or listening to English. Therefore, incidental learning can be likely to take place during extensive reading and listening. During extensive reading and listening, learners will encounter words many times again and again to acquire them.

Webb and Chang (2015) examine incidental learning through reading and listening with 10th grade readers. Moreover, this study also investigates the relationship between vocabulary gain and the frequency and distribution of encounters with 100 target words used by the graded readers. 61 Taiwanese English learners were divided into two groups: the incidental learning group and the traditional approach group. A pretest, post-test, and delayed post-test were conducted to ensure the differences between the two groups. As results, the incidental learning group showed higher learning gains on the post-test and delayed

post-test. Word frequency and the distribution of encounters did not lead to significant differences among conditions.

Liu and Zhang (2018) conduct a meta-analysis about the effect of extensive reading on English vocabulary learning. They investigate: a) whether incidental learning can enhance learners' vocabulary development; b) the best length of the extensive reading programs, c) what the most effective reading material is and d) what educational treatment in the extensive reading is most effective to promote learners' vocabulary learning. 21 original studies in terms of the effect of extensive reading were selected and were divided into seven categories so as to see the overall and specific effects of different valuables (participants, length of instruction, reading materials, treatment, control group, test use and test reliability). In each category, the original studies were separated into further conditions. For example, in the case of participants, there were five groups for their ages such as junior high school, senior high school, university, adults and children. By comparing the findings in the 21 original studies, this study addresses three research questions about the effect of extensive reading on English vocabulary acquisition of EFL learners, the best length of treatment in extensive reading, and the most effective reading materials and pedagogical treatment of extensive reading on English instruction. Results indicate that the learners who had an extensive reading program showed significantly higher outcomes in vocabulary learning. Moreover, this study reveals that the most appropriate duration for extensive reading is less than three months. Furthermore, it indicates that using graded readers as reading material leads the

learners to better results than other reading materials and that comprehension questions, dictionary usage, and vocabulary exercises such as sentence-making and vocabulary worksheets can enhance learners' vocabulary learning.

More recently, Webb, Uchihara, and Yanagisawa (2023) conducted a meta-analysis study that revealed the effectiveness of incidental vocabulary learning in second language settings. This study examined the effectiveness of incidental vocabulary learning through three types of meaning-focused input, reading, listening, and reading while listening. All three types of meaning-focused input show similar incidental vocabulary learning gains. However, this study also reveals that the amount of learning gains through incidental learning can be varied by moderator variables such as learners' characteristics (e.g., L2 proficiency level or institutional levels), materials (e.g., text type and audience), learning activities (e.g., spacing, mode of input), and methodological features (e.g., approaches to controlling prior vocabulary knowledge).

Even though incidental vocabulary learning can take place in EFL settings, it is very difficult for EFL learners to gain language input sufficient for successful incidental learning (Webb and Nation, 2017). Moreover, a lot of research claims that the learning gains from incidental learning tend to be relatively small in EFL environments (e.g., Horst et al., 1998; Warning and Takaki, 2003; Brown, Warning, and Donkaewbua, 2008; Pellicher-Sanchez and Schmitt, 2010). Therefore, many researchers claim that intentional learning is responsible for most vocabulary growth in EFL environments (e.g., Cobb, 2007; Laufer, 2003; Webb, 2008b). Intentional

learning can occur when learners' attention is directed at the target vocabulary. For example, intentional learning may occur when a teacher teaches a word to students in the classroom using flash cards or some tasks focusing on the target word, or when students search for the meaning of words with a dictionary.

Webb (2002) investigates the effects of incidental learning from reading and intentional learning with L1-nonsence word pairs on the learning gains of vocabulary knowledge. Findings indicate that learning gains from incidental learning depend on the amount of word repetition in the text. However, findings also indicate that learning gains from intentional learning with word pairs are much greater in all aspects of vocabulary knowledge than those from incidental learning. Laufer (2003) also suggests that L2 learners can gain more vocabulary knowledge through word-focused tasks than through reading. In this study, she attempts to examine the validity of some basic assumptions in vocabulary learning through reading in an L2 environment. The assumptions are the noticing assumption, the guessing ability assumption, the guessing-retention link assumption, and the cumulative gains assumption. Moreover, she compares the learning gains between reading and word-focused tasks to ascertain their effectiveness for L2 learners. Results indicate that the four assumptions in vocabulary learning through reading are not appropriate for L2 learners because vocabulary gains through reading by L2 learners are very small. Conversely, the tasks focusing on the target words allow the learners to make more vocabulary gains. Thus, this study concludes that intentional learning through word-focused tasks can be more effective in L2 vocabulary learning.

However, intentional learning also has its limits. First, teachers cannot teach all words in the classroom. As said before, Nation (2006) suggests that 8000–9000 word families are required to read authentic English texts without any help. It seems impossible for teachers to teach such many words in the classroom. Second, developing vocabulary depth is also an important aspect of vocabulary learning. However, it is difficult for teachers or learners to intentionally develop various kinds of vocabulary knowledge, such as collocations, derivations, and associations, for every word (Webb and Nation, 2017). Learning the amount of vocabulary knowledge requires learners to encounter the target vocabulary repeatedly (e.g., Webb and Nation, 2017; Pigada and Schmitt, 2006; Webb, 2007). Third, retention is also a significant aspect of vocabulary learning. Intentional learning allows learners to have more vocabulary gain (Laufer, 2003; Webb, 2002), but if learners cannot re-encounter the target words after the intentional learning in the classroom for a long time, the vocabulary knowledge will decay. Thus, time is also an essential factor in vocabulary learning (Webb and Nation, 2017). In fact, Webb, Yanagisawa, and Uchihara (2020) report that the learning gains with some word-focused activities can be relatively larger on the immediate post-test, but they decrease on the delayed post-test. Therefore, it seems difficult to maintain the vocabulary knowledge learners learnt only through intentional learning. In other words, learners need to encounter the target words again and again through reading or listening to retain word knowledge for longer.

Thus, it is important to promote both intentional learning and incidental

vocabulary learning in EFL settings (Hunt and Beglar, 2005; Krami and Bowles, 2019; Laufer, 2003; Teng, 2015). EFL learners cannot rely only on incidental learning because of the insufficient language input outside the classroom, but intentional learning cannot lead to the long-term retention of many words. Therefore, some words can be learnt incidentally and others intentionally while students also need to have opportunities to encounter the target words repeatedly to retain/develop them in the long term. In other words, intentional and incidental learning should complement each other.

2.2.8 Nation's (2007) four strands of vocabulary learning

Nation (2007) suggests that effective language learning can take place if learners are given appropriate language learning opportunities. Nation (2007) considers meaning-focused input, meaning-focused output, language-focused learning and fluency development as the four strands of vocabulary learning. Among the four strands, language-focused learning can be conducted through intentional learning and the other three strands can take place through incidental learning.

In meaning-focused input, learners' main focus or interest should be on understanding, gaining new knowledge or enjoying the target language through what they read or listen to in the target language (Nation, 2007). However, language input from reading or listening should be such that they can surely understand the texts they are reading or listening to. With the input hypothesis, Krashen (1981) proposes

that learners should be provided with comprehensible input (i+1 level) as much as possible. Accordingly, when Lightbown and Spada (2013) explain Krashen's input hypothesis, they use the term, "comprehensible input hypothesis". According to them, the hypothesis is that language acquisition and learning are likely to occur when learners are exposed to sufficiently comprehensible language input. In Krashen's notation, "i" represents the learner's current level of language or the knowledge they have already learnt before and "+1" means that the language input they are facing needs to be slightly beyond their current level of language or include some new knowledge such as unknown words, grammatical forms or aspects of pronunciations (Lightbown and Spada, 2013).

In the meaning-focused input strand, Nation (2007) suggests that learners should know or at least be already familiar with most of the language in meaning-focused activities. Moreover, the reading text and activities should also include only a small proportion of unknown language features or vocabulary (i.e., i+1 level) so that they do not interfere with learners' comprehension. Providing learners with comprehensible input also helps to lower their stress levels. According to Krashen's affective filter hypothesis (1981), stress can prevent learners from fully opening their mind to the language input. Thus, ideally learners need to know 95-98% of word tokens in textbook activities.

By contrast, meaning-focused output involves speaking and writing. This suggestion came from the output hypothesis proposed by Swain (1995). That is, since it appears that merely receiving a great deal of language input is insufficient

for language development, students must also practice applying the knowledge they gain from language input. Nation (2007) explains that learners should focus on conveying their thinking or opinion to someone else. Thus, they should not be afraid of their errors or mistakes in the grammar they use. Nation (2007) also mentions that learners can use some communication strategies, dictionaries or previous input to supplement the gaps between the receptive knowledge that learners already have for a word and the productive knowledge they want to develop. Through speaking and writing meaning-focused activities, learners can notice gaps in their knowledge (Nation, 2007; Swain, 1985). Teachers' corrective feedback about learners' output can also lead learners to significant learning (Ellis, 1989; Nation, 2007; Swain, 1995).

Language-focused learning involves providing some target language features, knowledge or vocabulary through intentional learning. That is, when teachers lead their students' attention toward one target language feature through grammar explanation or teaching vocabulary with using word cards, language-focused learning can occur in the classroom. Webb and Nation (2017) mention that many L2 and EFL courses depended on language-focused learning too much and in most of cases, it was not effective or efficient way. They state that learners are more likely to learn only half or less receptively and fewer productively. As Webb, Yanagisawa and Uchihara (2020) have reported, relying solely on intentional learning makes it challenging for learners to retain their learnt knowledge over an extended period. Therefore, in the language-focused learning for vocabulary, rather than the direct vocabulary teaching, teachers should develop students' vocabulary

learning strategies (Webb and Nation, 2017). In other words, teachers should train and encourage students' autonomous deliberate vocabulary learning.

Fluency development has a close relationship with the meaning-focused input and output. Its purpose is to enhance learners' already known knowledge, or what they learnt before. That is, fluency development is not the stage aiming at learning some new language knowledge, but at enhancing the knowledge learners already have or learnt before. Thus, unknown words or grammar should not be included in the activity. The ideal level for this strand should be "i-1 level", which means the level slightly under the learner's current language competency. In other words, the fluency development strand is required to provide the learners with materials they can fully comprehend. Nation (2007) suggests four ways to check whether a class or activity is suitable for fluency development: (a) the activities in the fluency development should consist of only the knowledge learners are already familiar with, (b) the purpose of the activity focuses on receiving or conveying meaning, (c) the activity requires learners to react faster than usual and (d) there are many input and output opportunities in the activities.

The four strands Nation (2007) proposes are useful to establish a well-balanced language course or vocabulary learning program. They can provide many opportunities to learn new knowledge and enhance the knowledge learners have already learnt. However, as Webb and Nation (2017) claim, many language courses tend to focus on only language-focused learning in the classroom. The time in the classroom seems to be very limited. Therefore, it might be difficult for teachers to

incorporate the four strands into one class. Yet, well-balanced teaching and learning are required to offer sufficient learning opportunities for learners. To make learning in the classroom more effective, applying the four strands seems worthwhile.

2.3 Factors affecting vocabulary learning

Webb and Nation (2017, p.60) observe that “vocabulary learning occurs because certain conditions are established which facilitate learning.” These conditions are divided into two key factors, vocabulary repetition and the learner’s quality of attention. In general, the more encounters with target words learners have, the more likely they are to learn them. Similarly, the deeper the quality of attention to target words, the faster and better learners can learn them.

2.3.1 Vocabulary repetition

Vocabulary repetition has been considered as one of the most important factors in vocabulary learning. There has been much research to investigate the effect of vocabulary repetition needed for incidental learning and intentional vocabulary learning (e.g., Brown et al., 2008; Horst et al, 1998; Laufer and Razovski-Roitblat, 2011; Pellicer-Sanchez, 2017; Peters, 2014; Peters and Webb, 2018; Rott, 1998; Teng and Xu, 2022; Uchihara et al, 2019; Webb, 2007; Waring and Takaki, 2003; Vidal, 2011; Zhang, 2022). The one widely shared suggestion among many studies focusing on repetition in vocabulary learning is that the more encounters with

the target words learners can have, the more likely they are to learn the words (e.g., Teng, 2015; Waring and Takaki, 2003; Webb, 2007). However, the exact number of repetitions required for vocabulary learning has been unclear so far. The reason for this might be that the required repetition rate can differ depending on the target words or the methods in the study such as learning intentionally or incidentally. Since the amount of required repetition can differ between incidental and intentional learning, it is necessary to discuss the effect of repetition on vocabulary learning in both learning contexts.

2.3.2 Repetition in incidental learning

Incidental learning, such as vocabulary learning that takes place through extensive reading or listening, requires learners to have a large amount of input to encounter the target words repeatedly. Webb and Chang (2015) suggest that the amount of incidental learning depends on how much input learners can have. That is, the more input learners can have through incidental learning, the more opportunities learners can gain to repeatedly encounter and learn words. Therefore, incidental learning and vocabulary repetition are deeply connected.

Various studies have examined the number of repetitions needed for vocabulary learning in incidental learning contexts. While some studies claim that a few encounters can enhance vocabulary learning in incidental learning, others suggest that relatively more encounters are required.

Rott (1999) claims that a few encounters with a word can show significant learning gains. This study examines whether intermediate level learners can incidentally acquire and retain the target unknown words through reading. This study involved 12 target words (TWs) and divided them into two sets (TWs1 and TWs2), ensuring each set consisted of three nouns and three verbs so as to compare acquisition and retention between the two sets for greater generalization of the results. In terms of the subjects, 95 university-level learners participated in the study and divided into two groups, one groups for TWs1 and the other for TWs2. Moreover, the subjects in each group were further separated into three different repetition groups (two, four and six) to examine the effect of vocabulary repetition. The subjects' receptive and productive learning gains through reading were measured by recognition test and supply-definition task through a post-test (after reading) and two delayed tests (after one week and two weeks). Results indicate that the subjects who encountered the target words two to six times during their reading showed significantly higher knowledge of the target words on every post-test (immediately after reading, one week and one month later) than the other subjects who did not encounter the target words in reading. Moreover, regarding the retention of knowledge, most of the participants retained the receptive target word knowledge even after four weeks post-test, but both six-exposure group and two-exposure group decrease their productive knowledge significantly after four weeks. However, this study reports that the amount of receptive and productive knowledge of target words the subjects can encounter in reading can still be significantly larger than that

of the subjects who could not encounter the words in reading. Furthermore, any significant differences are not found between two and four encounters in the learning gains of receptive and productive knowledge, but six encounters result in significantly higher learning gains for both receptive and productive knowledge. Thus, Rott (1999) concludes that incidental reading can be effective for vocabulary growth and although two or four encounters with words can affect learners' learning gains in vocabulary learning, six encounters show the highest effect on both receptive and productive vocabulary knowledge.

Vidal (2011) also examined how many repetitions are sufficient for incidental vocabulary learning in reading. Vidal (2011) compares the effect of several vocabulary learning factors such as vocabulary repetition, type of word, elaboration and predictability of word form and part through incidental reading and listening. 230 university students were separated into listening condition, reading condition and controlled condition. 112 subjects in the listening condition took three 14-15 mins lectures and 80 subjects in the reading condition was required to read three reading texts. Since the lectures and reading texts were developed from the same spoken and written authentic sources, they both contained the same content and target words. 38 subjects in the controlled condition did not have any listening and reading input in this study. Each lecture/text had 12 target words, so there were 36 target words in total. These target words were classified as three types of words: technical, academic and low-frequency. In terms of vocabulary repetition, this study set six frequency bands, spanning from one to six encounters. In addition to these, by

investigating the relationship between vocabulary retention and two other factors (word elaboration and predictability of unknown words from word part and form), Vidal (2011) examines how four factors influence vocabulary learning gains. This study conducted a pretest, an immediate post-test and a delayed post-test (one month later). Results indicate that both listening and reading can contribute to the acquisition of target words, but the reading group shows higher vocabulary gains on both post-tests and longer retention than the listening group, especially among low-proficiency learners. The finding of the repetition rate reports that two or three repetitions allow the learners to show their learning gains and retention in reading. Moreover, the learning gains are enhanced gradually between three and five repetitions and there are no significant differences from five to six repetitions in learning gains. By contrast, in the listening group learning gains occur after five or six encounters. Thus, this study concludes that learners can increase their vocabulary knowledge incidentally through both listening and reading, but more efficiently through reading. In terms of the comparison of learning conditions such as vocabulary repetition and the predictability, types and elaboration of words, this study suggests that the vocabulary repetition is the best predictor of vocabulary acquisition in reading among the four conditions. However, in listening, the predictability of words becomes the best predictor for learning gains. Therefore, this study reveals that the contribution of the learning conditions can be different depending on the type of activity learners take for vocabulary acquisition.

Moving on to studies which suggest relatively many repetitions are needed

in incidental learning, Webb (2007) investigates the effects of repetition on the learning of orthography, association, grammatical function, syntax, and meaning and form for both receptive and productive level. The aims of this study are to identify how many encounters with target words are needed by learners to develop receptive and productive knowledge of the target words. The repetition rates in this study were one, three, seven and ten occurrences. 121 Japanese English learners participated in this study. Results indicate that even at just one encounter, sizeable learning gains can be found in the knowledge of orthography, association, grammatical function, syntax, and meaning and form at receptive level. In terms of productive knowledge, the learners demonstrate large gains in orthography, or spelling with one encounter. The subjects in this study show the high scores on most of the aspects of productive knowledge, but the meaning and form were relatively lower than others. Thus, Webb (2007) suggests that the learning of orthography, association, grammatical function and syntax can be faster than that of meaning at the productive level. After three encounters, receptive knowledge of orthography, grammatical function and syntax and productive knowledge of association become significantly larger than after a single encounter. Webb (2007) suggests that these kinds of knowledge might develop earlier than meaning. Moreover, since relatively high scores on the productive tests began to be found at this point, this might indicate that three encounters are needed to gain enough word knowledge to use the word correctly. After seven encounters, there was relatively little difference from the mean scores after three encounters on the receptive level. However, larger gains are found for all

the target vocabulary knowledge at productive level. After 10 encounters, the learning gains become significantly higher for all aspects of word knowledge in both receptive and productive level than 7 encounters on four of the 10 tests. The test-score of association and syntax are especially high. Therefore, he concludes that by 10 encounters, learners probably learn to recognize spelling, meaning, association and grammatical function. Besides, Webb (2007) concludes that learners can possibly use words with grammatical accuracy, syntax, spelling, association and meaning after 10 encounters.

Pellicer-Sanchez and Schmitt (2010) also report that more than 10 encounters with target words might be needed for substantial vocabulary learning. In their study, subjects were 20 Spanish learners of English. This study aims at identifying 1) the effects of reading an authentic novel on developing their vocabulary knowledge of spelling, word class and meaning, 2) how many encounters are needed in order to learn the aforementioned three kinds of vocabulary knowledge and 3) the relationships between learners' attitudes and vocabulary acquisition. This study examined these questions by a combination of multiple-choice test and semi-structured one-to-one interviews to ensure the degree of target knowledge learners gained. The tests were a recognition test of spelling, a recall test of word class, and both recognition and recall test for word meaning. Results indicate that reading a single authentic novel allows for a measurable effect on the incidental learning of all three kinds of target vocabulary knowledge. However, the degrees of vocabulary knowledge mastery are not equal among spelling, word class and meaning. Meaning

recognition yields the greatest learning gains. Spelling is the second highest learning gain on this study. Moreover, the vocabulary knowledge of word class is following spelling and the smallest learning is meaning recall of the target words. Thus, this study confirms learners are more likely to develop first meaning recognition, then spelling recognition, then a recall of word class and, finally, meaning recall through incidental learning. In terms of the repetition for learning the target knowledge, among five different frequency bands (once, 2-4, 5-8 10-17 and 28 or more), this study finds while there is no significant difference between once and 2-4 bands, the 5-8 band shows a noticeable increase. However, the 10-17 band shows much higher increase on all kinds of vocabulary knowledge. Moreover, the 28+ band also shows further increase after the 10-17 bands, but the increase is not as large as that between the 5-8 and 10-17 bands. Thus, this study concludes that the words learners encountered more than 10 times can be learnt better on every target vocabulary knowledge than words with lower repetition rate. Furthermore, this study suggests that learners' positive attitude toward reading can also be a facilitative learning factor.

Pellicer-Sanchez (2014) conducts an eye-tracking study that investigates the vocabulary repetition needed for vocabulary learning during reading. Data was collected from 23 L2 subjects and 25 L1 subjects of the university level. The target words in this study were all nonwords, which look English words, but do not exist. Six nonwords were selected ensuring they are equally guessable from the contexts and repeated eight times. In terms of the procedure, this study recorded subjects'

eye movement while they were reading a story to examine their reading time and the number of fixation (i.e., Eye-tracking). After reading, they answered the true-false comprehension questions. Then, the subjects were required to complete vocabulary tests: a form recognition test and a meaning recognition test individually in paper format. Moreover, the subjects were asked to answer meaning of target words through a personal interview. Results indicate that both L1 and L2 subjects learn a considerable number of six target words after being exposed eight encounters in reading and show positive outcomes in the vocabulary tests (i.e., a form recognition, a meaning recognition and meaning recall). Moreover, the eye-tracking reveals that between three to four encounters, there was a significant decrease in the reading time and fixation for the L2 subjects, whereas the L1 subjects need only the first encounter for the decrease. Moreover, after eight encounters, both L1 and L2 subject can read the unknown words in a similar way as known words. Therefore, Pellicer-Sanchez (2014) concludes that eight repetitions might be needed for incidental vocabulary learning in reading.

Teng (2015) examines the effect of repetition in reading a Graded reader. In this study, 30 university students took part in a reading program. Their learning gain through incidental learning and retention of the receptive and productive vocabulary knowledge (meaning, collocation and lexical progression) were measured with the Word Associates Test developed by Read (1993; 2004), which contains a multiple-choice test to examine the depth of receptive and productive vocabulary knowledge with targeting collocations and synonymous relationship between a target word and

eight options. Moreover, in order to examine the effect of vocabulary repetition with Graded readers, 36 target pseudo-words were divided into six different frequency bands (1, 4-6, 8-10, 14-16, 18-20 and 21-22 occurrences). The procedure of this study had three sessions. In the first session, students read a book and take a test to produce a synonym or a collocation for each target words. Seven days later, for the second session, students read the book again and complete the multiple-choice test to identify the collocation, synonymous relationship between target words and eight options. At the third session, three months later, the tests were administrated again to investigate the students' retention of vocabulary knowledge they learnt in reading. Results indicate that the learning gain through incidental learning can occur largely depending on the repetitions in reading, but the number of the words the subjects can learn is very small. Teng (2015) concludes that at least 14 repetitions are needed for receptive knowledge and 18 repetitions for productive knowledge to develop through incidental learning. However, the tests conducted after three months show that the students in this study fail to retain the vocabulary knowledge they gained. Therefore, Teng (2015) argues that it is beneficial to emphasis the significance of integrating intentional and incidental learning in order to provide students with practical applications of the vocabulary knowledge they have learnt.

Vocabulary researchers have reached a consensus that vocabulary repetition in incidental vocabulary learning is a crucial variable for promoting L2 vocabulary acquisition and that there can be considerable variation in the number of repetitions required for learners to demonstrate learning gains. However, the exact

number of encounters learners require for incidental vocabulary acquisition is still unknown due to methodological variation across studies. Uchihara et al. (2019) note that the focus of vocabulary researchers in incidental vocabulary learning has shifted from identifying a threshold number of encounters to comprehending the complexity of the relationships between vocabulary repetition and other variables in incidental vocabulary acquisition.

Uchihara et al. (2019) examines the relationship between vocabulary repetition and incidental vocabulary learning by meta-analysing primary studies which report correlations between the number of encounters with vocabulary and vocabulary learning. The purposes of this study are to identify the 1) overall relationship between vocabulary repetition and incidental learning and 2) variables which can moderate the relationship between vocabulary repetition and incidental learning. The potential moderator variables in incidental vocabulary learning research are categorized into four categories: learner variables (e.g., age, gender, proficiency, motivation, working memory, background knowledge), word characteristics (e.g., imageability, concreteness, cognateness, the length of words or POS), text characteristics (e.g., genre, text length, richness of contextual clues, or significance of the word for text comprehension) and methodological or treatment variations (e.g., spaced or massed learning conditions, or use nonwords or real words). This study focused on 10 moderator variables from the learner variable and methodological or treatment variable categories. Findings indicate that incidental vocabulary learning seems to have a medium effect on vocabulary learning, but the

size of repetition effect is not as large as the authors expected. Moreover, this study confirms that the moderator variables can affect the relationship between vocabulary repetition and incidental learning. For instance, the age and the amount of a learner's vocabulary knowledge has a significant influence for the relationship between their learning gains and vocabulary repetition. This study reports that older learners tend to gain more beneficial effects from vocabulary repetition in incidental learning than younger learners. Additionally, learners who do not have rich vocabulary knowledge are more likely to develop their vocabulary knowledge through frequency of encounters. In other words, the vocabulary growth of learners who have a high proficiency level of English is less likely to be explained by the frequency of encounters in incidental learning. Moreover, massing seems to be more effective than spacing in incidental learning. Finally, incidental reading and listening may be more advantageous modes of input than reading while listening or viewing. Based on the findings in this study, Uchihara et al. (2019) concluded that the size of vocabulary repetition effect in incidental learning varies across studies because of differences across studies in terms of moderator variables.

2.3.3 Repetition in intentional learning

While a number of studies have examined the effect of vocabulary repetition in incidental learning, the role of repetition in intentional learning has not been examined as much as that in incidental learning. However, intentional learning

seems to be responsible for vocabulary development in EFL environments. In fact, many studies have shown that incidental learning is moderate and time-consuming (e.g., Horst, et al, 1998; Teng, 2015, Webb and Chang, 2016). Moreover, unless learners receive a large amount of input, it is difficult for them to consolidate their vocabulary knowledge through incidental learning. Therefore, the need for intentional learning becomes more salient in EFL environments, where time and input are very limited. Laufer (2006) suggests that intentional learning is indispensable for EFL learners to develop their lexical competence, which she defines as “a combination of different aspects of vocabulary knowledge, together with vocabulary use, speed of access, and strategic competence” (p.162). Therefore, the topic of repetition rate required for intentional vocabulary learning is worthy of investigation, especially for learners in EFL environments.

Research suggests that learning gain through intentional learning is more likely to occur with a smaller number of repetitions than in incidental learning. Webb and Nation (2017) refer to earlier studies which indicate that seven repetitions are needed to learn most words (Crothers and Suppers, 1967; Lado, Baldwin and Labo,1967; Tinkham, 1993,1997; Waring, 1997 cited in Webb and Nation, 2017). Laufer and Rozovski-Roitblat (2011) investigate the effect of the task types and repetition in vocabulary learning and retention. The tasks used in this study are Focus on Form (FonF: reading a text while using dictionaries or teachers’ help) and Focus on Forms (FonFs: reading a text with word focused exercises). The aforementioned FonF task, on one hand, is likely to induce incidental learning and

defined as “drawing students’ attention to linguistics elements as they arise incidentally in lessons whose overriding focus is on meaning or communication” (p.45-46) (Long, 1991, cited in Laufer and Rozovski-Roitblar, 2011). On the other hand, the FonFs task is designed to teach distinct linguistic structures in distinct lessons in a sequence predetermined by syllabus writers. It is probable that this will promote intentional learning (Laufer and Rozovski-Roitblat, 2011). In order to examine the difference between the FonF (incidental) and FonFs (intentional) tasks, 30 target words appeared in each task. Additionally, the 30 target words were separated into three repetition bands, 2 to 3, 4 to 5 and 6 to 7 times. In other words, this study examines six different learning conditions (two tasks with each three different repetition band) and the learners encounter totally 60 new words (10 words in each learning conditions). Vocabulary learning gains were measured with unannounced passive recognition and recall tests for meaning. Results indicate that FonFs task shows higher score in the recall test than FonF task when learners encounter the target words more than 4 times. In terms of the recognition test, 6 to 7 encounters in the FonFs task show higher learning gain, while FonF did not show any significant increase in this study. Thus, this study reveals that an increase in repetition improves word retention only in FonFs conditions at least among once to 7 encounters. Laufer and Rozovski-Roitblat (2011) conclude that FonFs (intentional learning) can have a significant impact on vocabulary learning with less vocabulary repetition than FonF (incidental learning with dictionaries and teachers’ help).

Additionally, Peters (2014) examines how form recall is affected by the

effects of repetition, learning single words or collocations and the time of post-test administration. Subjects were introduced to 12 single words and 12 collocations through one, three and five repetitions in non-communicative decontextualized activities. Results show a large and positive impact of vocabulary repetition during intentional learning. This study reveals that the more repetition learners can have, the more likely they are to recall the target words. In other words, five encounters can lead to better recall scores than one encounter for both single words and collocations in intentional learning. Similarly, Zhang (2022) suggests that vocabulary learning gains are deeply connected with repetition, but at least seven times are needed for significant learning gains to occur in explicit aural vocabulary instruction.

However, Teng and Xu (2022) offer a different perspective on the effect of vocabulary repetition in intentional learning. They investigate the effect of productive and receptive activities to promote receptive knowledge into productive and how much repetition is needed for this progress. The findings in this study indicate that, regardless of the amount of vocabulary repetition, productive activities, such as L1-to-L2 translation or L2 sentence writing, are more effective in turning receptive knowledge into productive than receptive activities, such as multiple-choice or gap-filling activities, in which learners choose a word from a list to complete a sentence. Moreover, in terms of repetition, Teng and Xu (2022) suggest that 2 to 3 times play a significant role not only to promote productive mastery, but also enhancing learners' performance in activities. However, they point it out that the effect of vocabulary repetition in explicit learning start decreasing after four or more encounters. Thus,

this study concludes that only the first several repetitions can lead to considerable learning gain.

All these studies suggest that fewer repetitions are necessary for intentional than for incidental vocabulary learning. While many researchers suggest that nearly 10 or more encounters are needed to learn new words through incidental learning, the required repetition rate tends to less than 10 encounters in intentional learning. Therefore, in environments where it is difficult for learners to have many encounters with target words through incidental learning, intentional learning seems to be a more effective and pragmatic learning method than incidental learning.

2.3.4 The quality of attention

The quality of attention relies on four essential learning conditions, namely, *"noticing", "retrieval", "varied encounter and use",* and *"elaboration*" (Webb and Nation, 2017). The first condition, "*noticing",* involves paying attention to the target words through guessing their meaning from context, using a dictionary or highlighting the target words in text or flashcards. Schmidt (1990) suggests that "noticing is the necessary and sufficient condition for converting input to intake […] Paying attention is probably facilitative, and may be necessary if adult learners are to acquire redundant grammatical features" (p.129). Moreover, "*noticing*" can occur when learners provide some information about the target words such as the meaning or form to other learners who do not know the words through group activities. According

to Newton (2013), the words 'negotiated' by learners are more likely to be learnt than non-negotiated vocabulary in context.

The second condition is "*retrieval"*. Receptive retrieval occurs when learners re-encounter a word form and recognize it. Productive retrieval occurs when learners express the meaning associate with a word form or use the appropriate word form in their writing or speaking. The few studies on vocabulary retrieval suggest that retrieval is a powerful learning condition in vocabulary learning (e.g., Atikah and Rezki, 2018; Barcroft, 2015).

Barcroft (2007) investigates the effect of retrieval in L2 vocabulary learning with 44 native English speakers who learn Spanish as their L2. The target words in this study were 24 words with each related picture. The subjects learnt the 24 words through both a retrieval-oriented (12 words) and a control condition (12 words). In the control condition, 12 word-pictures were shown for 12 seconds for two trials. In the retrieval-oriented condition, each picture was shown alone without its corresponding word for 6 seconds and after that, each corresponding word was shown for 6 seconds for two trials. This process was done twice. Thus, the retrieval-oriented condition allows the subjects to retrieve the target words in their learning stage. Results indicate that the vocabulary in retrieval-oriented condition is more likely to be learnt by the subjects than the control condition.

The third and fourth conditions – *"varied encounters/use"* and *"elaboration"* – are related to each other. According to Webb and Nation (2017), "*elaboration*" means the enrichment of knowledge of a target word by encountering and using its

different word forms and meanings. They suggest that elaboration is a form of varied encounters and use. The more varied encounters and use learners have, the more elaboration they can experience.

Examples of learning activities that utilize the learning conditions established by Webb and Nation in accordance with the quality of attention are presented in Table 2.2. (2017).

Table 2.2: Examples of activities that promotion vocabulary learning conditions (adapted from Webb and Nation (2017, p.63))

Quality of attention	Incidental	Intentional
Noticing	1. Guessing form context 2. Noticing a gap when speaking and writing	Highlighting a word in a text Focusing on the form or meaning of a word on a flashcard Using a dictionary or glossary Being taught words
Retrieval	1. Seeing a previously encountered word while listening or reading, and recalling its meaning 2. Recalling and using a recently encountered word as part of conversation or writing	Remembering words on flashcards Doing cloze exercises after reading a text Playing a game that involve remembering the names of objects (e.g., Kim's Game) Recalling a list of words
Varied encounters	1. Seeing a previously encountered word in a new form or context while listening or reading, and recalling its meaning (e.g., linked skills) 2. Extensive reading	Looking at different examples of the words used in context Doing an exercise consisting of true/false sentences
Varied use	1. Recalling and using a recently encountered word in a new way in conversation or writing (e.g., linked skills)	Doing cloze exercises Doing topic-based continuous writing Giving a presentation
Elaboration	1. Encountering and using a word to communicate 2. Describing pictures 3. Reading interactively	Using memory techniques to link L1 and L2 words (the keyword technique) Creating a chart or map of related words (semantic mapping)

	(reading and discussing in a group)	Analysing word parts Looking at the different senses of a word to determine its core meaning

Webb and Nation (2017) explain that these four learning conditions interact with each other. For example, when learners retrieve a word, they may also notice something new about the word or they recognize the different use of the word. When a word is used with a form that learners have not encountered before, they can have opportunities for *"noticing", "varied use*" or *"elaboration"*. Some of these significant learning conditions are reflected in Technique Feature Analysis by Nation and Webb (2011), a framework for evaluating the effectiveness of vocabulary learning activities.

In terms of the quality of attention, Laufer and Hulstijn's (2001) Involvement Load Hypothesis (ILH) is a first attempt for researchers and teachers to promote the operationalization of the general labels of 'attention' and 'elaboration' into concrete task-specific constructs. Laufer and Hulstijn (2001) define 'involvement load' as the combination of any of those involvement factors. Retention of unfamiliar words in vocabulary learning requires three conditional factors in a task: *"Need," "Search"* and *"Evaluation"*. The combination of these three factors is called 'involvement'. "*Need*" is the motivational, non-cognitive dimension of involvement and the primary responsibility is to determine whether students independently select a new word or whether teachers assign it to them. If it is teachers' selection, "*Need*" becomes 'moderate'. *"Need*" is 'strong' when students choose a word because they need to learn it or out of interest. "*Search*" and *"Evaluation*" are the two information processing dimensions of involvement. "*Search*" means that learners need to find a

meaning of an unknown word using a dictionary or asking their teachers. *"Evaluation*" requires learners to judge or assess whether a word fits its context in an activity or task. When learners are required to make an original sentence with the unknown word, "*Evaluation*" will be 'strong'. However, if, for example, learners need to recognize a difference between words for some activity such as gap-fill tasks, *"Evaluation"* will be 'moderate'. These three involvement factors can be induced by a teacher-designed task or a real-life communicative situation. The ILH suggests that the higher involvement a task includes, the better retention can be achieved. Moreover, tasks with a higher involvement load will be more effective for learners to retain their new vocabulary than tasks with a lower involvement load. According to Laufer and Hulstijn (2001), it might be possible to state that the higher involvement toward a word induced by a task will lead to better retention. However, the authors suggested that the involvement load could vary depending on a teacher-designed task even though the tasks are the same. For example, in reading comprehension questions and writing tasks, if the unknown word is shown somewhere on the task, students do not have to do 'Search', but without any information about the unknown word, students need to do this involvement component. Moreover, whether there is a relevant glossary for the task or not can affect the 'Need'. Finally, the weight of the three involvement factors might be not equal. The authors stated that the 'Search' might be less important than other two factors.

Because few studies compare the effect of the repetition and quality of attention on vocabulary learning, it might be difficult to suggest which is more

effective. However, one study has suggested that the quality of attention has the strongest effect on vocabulary learning. Laufer and Rozovski-Roitblat (2015) are one of the few experiments that compare the effects of repetition and quality of attention on vocabulary learning. In their study, 185 EFL learners were allocated to three different tasks, Reading only (R), Reading with Focus on Form (F) and Reading with one encounter in text and Focus on Forms (1+Fs). The repetition rates were differently set for each task type group. The R group was divided into three “number of encounters” conditions, 1) six or nine times, 2) 12 or 15 times, and 3) 18 or 21 times through reading different texts. The F group could use a dictionary during their reading and encountered the target words among “2 to 3”, “4 to 5” and “6 to 7” times in the reading texts. The 1+Fs group had the same number of encounter conditions as the F group, but only the first encounter with the target words in this group occurred in the reading texts and other encounters occurred in some intentional activities. That is, if a target word appears three times in reading, the F group can encounter the word three times during their reading, while the 1+Fs group could see the first encounter with target words in the reading texts and two encounters through word-focused exercises after their reading. The word-focused exercises for the 1+Fs group were filling in sentence, matching the target words with their definition or synonyms, and supplying L2 translation for the L1 equivalents of the target words in isolation and sentence. Therefore, the subjects in the 1+Fs group could learn the target words with deeper attention than other two groups. The leaning gains were measured through pretest and post-test. The pretest contained passive recall and

recognition and the post-test contained active and passive recall and recognition tests. Results indicate that the smallest number of encounters in the 1+Fs group (two to three) lead to higher scores than the maximum encounters in the R (18 and 21) and F groups (six to seven). Laufer and Rozovski-Roitblat (2015) conclude that task type, which makes learners pay their attention to the target words is more important than the repetition, or how many times learners encounter the target words. As reasons for these results in this study, they mention that the word-focused instructions or activities are likely to include some of important learning conditions such as noticing, elaboration and the involvement suggested by Laufer and Hulstijn (2001).

2.3.5 Other factors contributing to vocabulary learning

While repetition and the quality of attention are significant factors in vocabulary learning, other factors also affect vocabulary learning. They are word frequency, word length and cognateness.

Word frequency has been defined as the number of occurrences of the target words in a large reference corpus such as the BNC or BNC/COCA (e.g., Horst, Cobb and Meara, 1998; Reynolds et al., 2015) or the frequency of occurrences in learners' classroom input or instruction materials (Demetriou, 2017; Vidal, 2003). As discussed before, learners need to know high-frequency words because they will need to recognize and use them frequently.

Additionally, some researchers suggest that the word length is also an important vocabulary learning condition (e.g., Alsaif and Milton, 2012). Alsaif and Milton (2012) examines the impact of some learning difficulties such as concreteness, word length and repetition on vocabulary learning with 22 EFL textbooks in Saudi Arabia. Results indicate that three learning difficulties can predict learners' outcome in vocabulary learning. Among the three, word length can predict the learnability of the target words the most. According to this study, the more syllables a word has, the more difficult the learning becomes. Therefore, Alsaif and Milton (2012) conclude that the learnability of a word can be predicted by a combination of the three learning difficulties in vocabulary learning. However, Laufer (1996) suggests that if a long word has some familiar components such as inflections or deviation knowledge, learners can learn it more easily an unknown short word that cannot be broken down into known word parts. Thus, she indicates that the number of letters/syllables in a word does not always determine the word difficulty and studies about the word length need to consider the relationships between the effect of word length and other lexical knowledge on vocabulary learning.

Cognates can promote vocabulary learning (Rogers, Webb and Nakata, 2015). Daulton (2008, cited in Urdaniz and Skoufaki, 2022) mentions that cognates have received mainly two definitions, one diachronic and the other synchronic. The diachronic definition suggests that cognates are the words, which are etymologically related in the form and meaning between L1 and L2. By contrast, the synchronic definition suggests that cognates are the words, which are similar in the form and

meaning regardless of whether they are etymologically related or not. According to Daulton (2008), the synchronic definition is more important in language learning research because learners' perceived cross-linguistic similarities affect their vocabulary learning.

Cognates can reduce the learning burden of a word (Rogers, Webb and Nakata, 2015). Webb and Nation (2017) explain that the learning burden is the difficulty of learning a word or how much effort learners need to learn the target words. Because cognates share language similarities between L1 and L2, learners are more likely to learn cognates than non-cognates in vocabulary learning (Harris, 2019).

Urdaniz and Skoufaki (2022) examine the effect of cognateness, word frequency and word length on learners' recognition knowledge of English academic vocabulary. 38 Spanish L1 university students who had learnt English for at least four years answered three vocabulary tests: the online vocabulary size test, LexTALE (Lemhofer and Broersma, 2012), English Yes/No vocabulary test, and a Spanish Yes/No vocabulary test. LexTALE was used to estimate the learners' English proficiency level. The main data collection instrument is the English Yes/No vocabulary test which consisted of 52 lexical items selected from the most frequent 1000 lemmas in Academic Vocabulary List (AVL, Gardner and Davies, 2014) and 35 pseudowords. 26 of the target English words have Spanish cognates and 26 do not. Through the Spanish Yes/No vocabulary test, this study ascertained whether the learners in this study knew the corresponding L1 words of the target English cognates. Results from a multiple regression analysis indicate that word frequency

is the best predictor of learners' L2 English recognition knowledge of academic vocabulary while Cognateness is the second most important predictor. Moreover, an interaction between word frequency and cognateness shows that the effect of word frequency becomes stronger when the target words do not have cognates in Spanish. From these findings, Urdaniz and Skoufaki (2019) conclude that word frequency and cognateness can predict Spanish EFL learners' recognition knowledge of academic vocabulary.

2.4 Vocabulary in EFL textbooks

2.4.1 Research on vocabulary in EFL textbooks

The textbook students use influences the vocabulary they learn and their performance on classroom activities (e.g., Ayu and Inderawati, 2018; Bergstorm, Norberg and Nordlund, 2023; Criado and Sanchez, 2009; Matsuoka and Hirsh, 2010; Cao, 2018; Sun and Dang, 2020). Milton (2009) also argues that the textbook is an indispensable resource for L2 students, particularly in EFL settings. To optimize learning through textbooks, words should be selected for inclusion in textbooks so that learners can understand most textbook content and learn useful words as well. The question then is how the effectiveness of EFL textbooks can be evaluated. There appear to be three major approaches to evaluate EFL textbooks in terms of examining the vocabulary they provide (e.g., Sun and Dang, 2020; Yang and Coxhead, 2022).

The first approach is to estimate vocabulary load to reach the thresholds of lexical coverages between 95 and 98% in the vocabulary used in textbooks. The vocabulary load is defined as “the lexical difficulty involved in understanding the words in spoken and written text” (Webb and Nation, 2017 p.285). Additionally, the lexical coverage is defined as the number of words in the text that are known by readers (e.g., Nation, 2006, see more detail in 1.3). To sufficiently comprehend the contents of English texts, students need to possess 95-98% of vocabulary used in the texts (e.g., Hirsh and Nation, 1992; Laufer, 1989; Nation, 2006; Schmit et al., 2011; van Zeeland and Schmitt, 2013). Therefore, when it comes to examining the vocabulary in textbooks, it is important to estimate how many words student need to know and to ensure whether the vocabulary used in the textbooks mostly consists of vocabulary students know. Thus, for example, if the vocabulary load in a textbook requires students to have many low-frequency words, it can mean that the textbook seems to be too difficult for the students to read. It should be highly important for textbook publishers to provide students with appropriate levels of vocabulary in textbooks so that students can comprehend and have opportunities to develop their vocabulary knowledge. Therefore, the question that how many words or word families students need to know to cover 95-98% of vocabulary in textbooks appear to be an important question on the evaluation of textbooks in terms of the vocabulary learning.

Research on the vocabulary load of textbooks has been conducted several times so far. Browne (1996) examines the readability and lexical difficulty of three

EFL textbooks used in Japanese high schools with VocabProfile and traditional readability formulas such as the Flesh-Kinkaid. According to the readability formulas, the content of the textbooks is simple, but the lexical difficulty is extremely high. Thus, the vocabulary in the textbooks is above the level of high school students. According to this study, 30 to 40% of the vocabulary that Japanese learners encounter in context is unknown. Rather than memorizing numerous low-frequency words in senior high school, he suggests that they can build a solid foundation by learning high-frequency words and University Word List (Nation, 1990) Another study conducted by Sun and Dang (2020) examines the vocabulary load of some Chinese senior high school English textbooks requires students to cover 95-98% of vocabulary. This study reports that the students need a vocabulary size of 3000 word families to reach 95% of lexical coverage and 8000-11000 word families for 98%. However, as a result of the Updated Vocabulary Levels Test (Webb, et al. 2017), which was conducted to reveal the amount of vocabulary knowledge the students have, Sun and Dang (2020) also report that only five out of the 265 students mastered the most frequent 3000 word families in this study. Thus, Sun and Dang (2020) concluded that since most of the students have insufficient vocabulary knowledge of the most frequent 3000 word families, textbooks which require them to have 3000 word families to achieve 95% text coverage and 8000-11000 word families to achieve 98% text coverage are clearly too demanding.

Moreover, Le and Dinh (2022) investigate the vocabulary load of Vietnamese senior high school English textbooks. This study analysed 41,137 token word corpus

of the target textbooks, named *Tieng Anh 10* by BNC/COCA 1-25K in the Vocabprofiler created by Cobb. Results indicate that the Vietnamese students in this study need to possess a vocabulary size of 3000-5000 word families to cover 95-98% of vocabulary used in the textbooks. However, this requirement seems to be too challenging for the students because their receptive vocabulary knowledge is only around 2000 word families. Therefore, Le and Dinh (2022) also conclude that the vocabulary load of the Vietnamese textbooks is too challenging and it is needed to promote students' effective vocabulary learning and general comprehension of materials by focusing on the most frequent 2-3000 word families.

According to Shin et al. (2011), the South Korean elementary and secondary school textbooks contain vocabulary which is too difficult for students to learn. This study utilized five wordlists to examine the frequency level of vocabulary used in South Korea's 140 textbooks authorized by the government: the General Service List (including 2000 word families), the Academic Word List (AWL: 570 word families), the British National Corpus (BNC; including 14,000 high-frequency word families), the Freiburg-Brown Corpus of American English (FROWN) and the Freiburg-LOB Corpus of British English (FROWN) (FLOB). According to the results, 67.78% of the token words in elementary and secondary school EFL textbooks does not exist in the GSL or AWL word lists. Moreover, the comparison with the BNC also indicates that the vocabulary size in secondary school surpasses the maximum vocabulary level of the 3000 words of the senior high school base word lists (BWL) by the National Curriculum. Thus, Shin et al. (2011) conclude that the vocabulary size in

the target textbooks is not governed by the BWL guidelines established by the National Curriculum, and they recommend the development of more appropriate BWL so that students can acquire more learning-appropriate vocabulary. If the vocabulary load in textbooks becomes too much for the target students, the content cannot be comprehended sufficiently and learning can be more difficult or impossible. Therefore, it should be essential to provide students with comprehensible texts by using useful words so that they can cover the ideal lexical coverages in the whole textbooks. By analysing the vocabulary load, it is possible to ascertain whether the textbooks provide students with appropriate level texts so that they can deepen their knowledge through the textbooks.

In terms of international English textbooks, Matsuoka and Hirsh (2010) report that in *New Headway Student's Book Upper-Intermediate textbook* (Soars and Soars, 2005) 95.5% coverage was achieved by knowing the first and second 1000 word families in West's (1953) GSL and pre-teaching academic words and a few other words that students should be familiar with. This study analysed the target textbooks with Range (Heatley, Nation and Coxhead, 2002). Findings indicate that the target textbooks can provide opportunities to deepen upper-intermediate level students' knowledge of the second 1000 most frequent words and words the students need to learn from context. It is plausible to hypothesize that students who utilized the textbooks examined in Matsuoka and Hirsh's (2010) study were more capable of comprehending the material than those whose textbooks predominantly comprised low-frequency or mid-frequency words with limited high-frequency words.

The second approach to analyse the effectiveness of English textbooks is to count the number of high-frequency words in EFL textbooks. This is essential because learning high-frequency words is typically the most important aspect of English learning. The more high-frequency words in EFL textbooks, the greater the likelihood that students will have the opportunity to learn essential words. This method enables teachers and students to determine whether EFL students can encounter the essential words in their textbooks. In order to determine whether a textbook provides useful vocabulary for EFL students, it is crucial to count the number of high-frequency words.

Nakayama (2021) analyses the vocabulary in nine Japanese EFL senior high school textbooks. This study aims to investigate two aspects of high-frequency words found in the target textbooks. The first is to examine the population of high-frequency words in the target textbooks, while the second is to assess their diversity. This study utilized the New General Service List (NGSL: Browne, Culligan, and Philips, 2013), which is comprised of the 2,801 most common flemmas, to determine these aspects. Results indicate that the Japanese textbooks in question contain a high proportion of high-frequency flemmas in the NGSL (around 92%). However, the textbooks do not provide many of high-frequency flemmas (only 38% of NGSL words at the most among the textbooks). In other words, the variety of high-frequency words in Japanese textbooks is extremely limited. Nakayama (2021) concludes that, since textbooks employ a limited variety of high-frequency words, teachers must expose students to other high-frequency words.

Nakayama (2022) analyses the vocabulary in a textbook series used in Japanese junior-high schools. As in Nakayama (2021), the focus is on the population and diversity of high-frequency words in the three textbooks from a single series. In addition, this study also investigates which high-frequency words the textbook series emphasizes in its content. The NGSL wordlist was used for this study as well (Browne, et al, 2013). As in Nakayama (2021), most of the word tokens in the textbooks are covered by NGSL words, but the variety of the HF words in the textbooks is very limited. More specifically, 95% of the word tokens in the target textbooks are covered by the NGSL but only 37% of NGSL lemmas appear in the textbooks. A high percentage (80%) of lemmas from the first frequency band in the NGSL (lemmas 1 to 560) appear in the textbooks but lemmas from other frequency bands are rarely used in the textbooks. The NGSL2 (lemmas 561 to 1120) is 6.2%, NGSL3 (lemmas 1121 to1680) is 3.9%, NGSL4 (lemmas 1681 to 2240) is 1.8% and NGSL5 (lemmas 2241 to 2801) is 1.1%. A Therefore, Nakayama (2022) concludes that these Japanese EFL textbooks lack the essential vocabulary learning opportunities for students to comprehend English texts. These studies indicate that Japanese students in junior- and senior high schools are not given the opportunity to learn high-frequency words from their textbooks.

Additionally, Eldridge and Neufeld (2009) report that only 1400 of the most frequent 2000 word families are presented in the *Success* coursebook series (Longman). In addition, O'Loughlin (2012) examines three distinct levels of EFL textbooks from Oxford University Press's New English File using Vocabprofile (Cobb,

2009). The results indicate that only 1435 of the 2000 most frequent word families were presented in textbooks. Although there are some differences in the population of high-frequency words in textbooks, these studies indicate that EFL textbooks were unlikely to include the vocabulary used frequently by native speakers. It is difficult to create meaningful textbook content using only high-frequency words (Milton, 2009). However, ignoring high-frequency words results in insufficient comprehension of textbook material and an increased responsibility to independently acquire frequent vocabulary. Therefore, it appears crucial that textbooks strike a balance between high-frequency words and less frequent words. In conclusion, there is widespread agreement regarding the importance of learning high-frequency words, but at least according to these studies, it may be difficult for students to learn useful words from textbooks alone.

The third common approach in textbooks vocabulary analysis is to examine vocabulary repetition in textbooks. Vocabulary repetition is one of the most important factors in vocabulary learning (Webb and Nation, 2017 see more detail in 2.3). Learners must encounter target words multiple times to learn and retain them. Thus, the number of repetitions can be a factor in evaluating the effectiveness of vocabulary learning in an EFL textbook.

There have been numerous investigations into the efficacy of word repetition in vocabulary acquisition and learning (see more details in 2.3). However, it is unknown precisely how many exposures students require to master a word. Despite the importance of word repetition in vocabulary learning, very few such

studies have been conducted to examine the vocabulary repetitions of high-frequency words on ELT textbooks (Matsuoka and Hirsh, 2010). Matsuoka and Hirsh (2010) examine the repetition of high-frequency words in *NEW Headway Student's Book Upper-intermediate* (Soars and Soars, 2005) to reveal the vocabulary learning opportunities for the second 1000 most frequent words in West's GSL and for the words in the Academic Word List (Coxhead, 2000). This study created four different repetition bands (only once, five, seven and 10 occurrences) to ensure how many times students can encounter the target words through the contents. Moreover, they randomly selected six words from those which appeared more than five occurrences in the textbook to examine their which aspects of vocabulary knowledge were shown in the repetitions. Findings indicate that 33.3% of the most frequent 2000 word families appeared only once, 31.3% at five times, 21.2% at seven times, and 12.3% at 10 times. In addition, 42.8% of AWL word families occurred only once, 16.2% at five times, 8.9% at seven times, and 5.5 % at 10 times. However, the case study for the selected six words can show that the repetition of words in textbooks can be useful for deepening vocabulary knowledge of the target words. Since each selected word family can appear with new information such as different POS, inflections and derivations and collocations in every repetition, this study suggests that the textbooks can provide students with meaningful opportunities to deepen vocabulary knowledge of the most frequent 2000 word families in the GSL. However, since the textbooks has a limit to repeat many of the target words, they conclude that the opportunities for enhancing students' learning vocabulary through textbooks are

limited and suggest the need for an extensive reading program or other forms of language input so that students can encounter the words they learn in their textbooks in other contexts.

2.4.2 Vocabulary selection for textbook inclusion

The selection of vocabulary in language textbooks appears to vary in quality between authors or curriculum developers. According to Alcaraz (2009), the selection criteria for vocabulary in textbooks remain obscure. Vocabulary learning relies heavily on the acquisition of high-frequency words, but some research indicates that material developers and textbook publishers do not appear to select vocabulary based solely on word frequency. As Nakayama (2022) points out, textbooks are not reading materials but instructional materials. Thus, focusing solely on high-frequency words probably prevents students from engaging in a variety of classroom activities. Milton (2009) concurs that teaching materials should be based on a curriculum and the usefulness of words within activities and remarks that textbooks need to have vocabulary that spans various frequency bands because textbook chapters cover different topics. Therefore, regardless of how ideal it is to teach students the most frequent vocabulary at the beginner level, it is necessary to introduce mid- or low-frequency level words such as technical words or words related to chapter themes, textbook exercises and classroom activities.

Several researchers propose a variety of criteria for the selection of

vocabulary introduced in textbooks. These include word frequency, coverage, opportunism, learnability and teachability, and genres of interest (e.g., Nation and Newton, 1997; Thornbury, 2002; White, 1988, as cited in O'Dell, 1996).

Regarding the word frequency criterion, there is broad consensus regarding its importance in vocabulary instruction and acquisition (e.g., Nation,1990; Nation and Newton, 1997; Schmitt, 2000). High-frequency words can play an important role in comprehension and use of English. Therefore, the most frequent 2000 word families should be taught and learnt first, particularly in the early stages of language acquisition (e.g., Nation and Newton, 1997; Schmitt, 2000). Moreover, Thornbury (2002) also suggests the frequency level of words as one of the selection criteria for textbooks vocabulary. Furthermore, according to O'Dell (1997), the frequency level is the most evident criterion for vocabulary selection. However, the frequency level of words does not appear to be sufficient for learners, as there are numerous other words that should be known but are not frequently used. Consequently, the selection of vocabulary must now incorporate additional criteria.

Coverage is the second criterion (White, 1988, cited in O'Dell, 1997). This means that words with broader meanings should be taught before those with more specific meanings. For example, "go" should be taught and learned before "travel" and "walk." In this sense, "go" can have a broader sense, whereas "travel" and "walk" are more specific. Similarly, Thornbury (2002) suggests that the selection of the target vocabulary in the textbook is influenced by the core vocabulary. The definition of the core vocabulary is a word that can combine with other words, define other

words, or replace other words. For instance, "giggle" is defined as a type of laughter. Conversely, the term is not used to define "laugh". This indicates that "laugh" is a more useful or essential word than "giggle". Similarly, words that collocate with many other words are more useful than those that do not (e.g., “bright” > “radiant”). In addition, superordinate words are an excellent addition to a core vocabulary (Thornbury, 2002). For instance, "flower" is a superordinate word of "rose" and other words for flower names. Collectively, selecting words that students can use more frequently and broadly is the concept of Coverage underlying the selection of textbook vocabulary.

The third prevalent criterion is opportunism (White, 1988, cited in O'Dell, 1997) or usefulness (Thornbury, 2002). These indicate that a textbook or curriculum should include words that students immediately need to use in the classroom activities, regardless of their frequency. For instance, the classroom-related terms "pen", "teacher", and "blackboard" are relevant to this criterion. In order to deepen students' comprehension in the classroom, Zimmerman (2014) suggests paying close attention to the words that stand out in the reading material of a textbook. Similarly, the usefulness proposed by Thornbury (2002) implies that there is a situation, environment, or opportunity to employ the vocabulary. In other words, the most useful words can be sometimes the ones that students will need shortly after learning them. Thus, they both assert that students are more likely to learn the words that are crucial for their classroom success. Consequently, these should be chosen regardless of their frequency level.

The fourth criterion is word learnability and teachability, which means how easily students can learn the word or how easily teachers can teach the word to their students. Therefore, the learning burden of words is an essential factor in these criteria. The learning burden can be defined as the amount of effort students need to learn the target word (Webb and Nation, 2017). Thus, the greater the learning burden, the more difficult it is for learners to learn. On one hand, the words with the lowest learning burden are those with high learnability. High-teachability vocabulary, on the other hand, is vocabulary that teachers can teach more easily through gestures, cards, pictures, or actual objects. For example, in the classroom, it is easier for teachers to teach words associated with the classroom, such as "blackboard," than abstract or grammatical words, such as "though", even though "though" is more frequent and useful than "blackboard". Similarly, Thornbury (2002) also suggests that a word's similarity to the learners' L1 (i.e., cognates) seems to be beneficial criterion for learners, because that can be a good indicator of how learnable a word is for learners. In the case of some cognates, English words can be some knowledge that learners have already known in their L1. This can be more helpful for learners to learn the English words.

The final common criterion is the genres of interest, which means what vocabulary students want to learn (White, 1988, cited in O'Dell, 1997; Nation and Newton, 1997). According to O'Dell (1997), "lexical sets which are likely to interest students in any particular group should be included" (p.269). for example, students who aim at becoming medical doctors may want to learn some vocabulary related

with the field of medicine, while other students may want to learn math or economic related vocabulary. Nation and Newton (1997) suggest that once students have mastered the most frequent 2000 words, the selection of vocabulary should be based on how the students use English. For instance, if students wish to read academic papers or newspapers, they need to learn at least some of the academic words listed in the Academic Word List (AWL: Coxhead, 2000) or the Academic Vocabulary list (AVL: Gardner and Davies, 2014) or to learn additional meaning senses for words which are polysemous and have both well-known everyday senses and more specialised or even technical senses in academia (see Skoufaki and Petrić, 2021). In addition to high-frequency vocabulary, learning vocabulary listed in the AWL (Coxhead, 2000) or AVL (Gardner and Davies, 2014) can be beneficial for English learners, especially those preparing to enter university or at university. According to Coxhead (2000), on one hand, the AWL's 570 word families can account for 10% of academic text on average. On the other hand, Gardner and Davies (2014) claim that the AVL's 3015 academic lemmas can account for approximately 14% of running words in academic reading. If, on the other hand, they wish to read novels or magazines, they should focus on learning vocabulary frequently used in such kinds of English texts. According to O'Dell (1997), for more advanced levels, this criterion will become more important than other criteria. In other words, students are required to develop their vocabulary alongside their personal interests in order to comprehend more specialized text.

2.4.3 Vocabulary knowledge in textbooks

Examining how much vocabulary knowledge is introduced to students can be one of the most important aspects of evaluating vocabulary instruction in textbooks. Table 2.1 (p.20 in Section 2.2.3) summarises Nation's framework of vocabulary knowledge. The framework indicates that "knowing a word" requires familiarity with its form, meaning, and use (Nation, 2013). In addition, the framework identifies 18 kinds of receptive and productive vocabulary knowledge. As little research has been conducted on the vocabulary knowledge introduced in English textbook, the extent to which such essential vocabulary knowledge is presented in textbooks should be examined and findings should feed into recommendations for the kinds of vocabulary activities that need to be included in textbooks to help students develop their depth of vocabulary knowledge.

Brown (2011) examines vocabulary activities in nine General English textbooks of three different proficiency levels (beginner, pre-intermediate and intermediate levels) to identify what kinds of vocabulary knowledge are focused. Findings indicate that the target textbooks focus only on the form and meaning, grammatical functions and spoken form out of the nine essential aspects of vocabulary knowledge in Nation’s framework. Other vocabulary knowledge appears to be neglected or given less emphasis in vocabulary learning activities in these textbooks. According to this study, there appear to be variations in the levels of emphasis placed on the three aspects of vocabulary knowledge by the textbooks' vocabulary activities. For example, beginner-level textbooks tend to pay roughly

equal great attention to form and meaning and grammatical functions. Moreover, the Spoken form also obtain some attentions at this level. At the pre-intermediate level textbooks, the form and meaning has a great attention in activities, though the grammatical functions is also prominent. Besides, the Spoken form, association and collocations also have some attention to some extent, but their prominences largely vary depending on textbooks. Intermediate-level textbooks devote the greatest attention to form and meaning. Although the grammatical functions receive some attentions, the extent is less than other levels textbooks. Moreover, the other seven aspects of vocabulary knowledge (i.e., spoken form, written form, word parts, concept and referents, associations, collocations, and constraints of use) receive little attention in textbooks. Thus, the focused vocabulary knowledge can be varied a little depending on the proficiency level of textbooks, but the form and meaning seem to have robust attention in textbooks over the other aspects of vocabulary knowledge. However, Brown (2011) suggests that all nine aspects of vocabulary knowledge should be presented in textbooks. He argues that textbooks need a system that allows students to acquire new vocabulary knowledge whenever they encounter the target words again. For example, when students encounter a word for the first time, they can learn its form and meaning. At the second encounter, the spoken form is visible. Students may be able to learn their target words in vocabulary activities in textbooks more thoroughly if they are exposed to new vocabulary on each occasion. Thus, Brown (2010) suggests that textbook authors and teachers in the classroom are required to provide such opportunities for students to acquire all

nine aspects of vocabulary knowledge in the context of English language learning.

Alshumrani and Al-Ahmadi's (2022) research on the vocabulary knowledge in textbooks is another example. This research is conducted on twelve elementary and middle school EFL textbooks used in Saudi Arabia. This study analyses vocabulary learning activities in Saudi Arabian textbooks by employing Nation's framework of vocabulary knowledge. Like Brown (2011), this study reports that there are many vocabulary knowledge aspects which the target textbooks do not introduce students to. This study reveals, for instance, that the spoken forms receive some attention in primary and middle school activities, but the written forms do not, particularly in middle school textbooks. In addition, this study's textbooks do not contain any activities for learning collocations or constraints of use. The findings indicate, however, that the primary school textbooks devote the most attention to spoken forms, with form and meaning and grammatical functions receiving second and third priority, respectively, in the activities. Word meaning, including form and meaning, concept and referents, and associations, receives the most attention in middle school, followed by grammatical functions and spoken forms. Therefore, Alshumrani and Al-Ahmadi (2022) conclude that the level of emphasis on vocabulary knowledge aspects in textbook activities can vary according to students' age and proficiency level.

2.5. The present study

This study focuses on the learning opportunities that senior high school English textbooks used in Japan and Taiwan offer for high-frequency lemmas and newly-introduced lemmas. By comparing textbooks used in the two countries, this study also aims at identifying differences in the use and introduction of vocabulary in the two textbook series.

Taiwan is one of the most suitable countries to compare English instruction with Japan since it fulfils some essential conditions. First, Japan and Taiwan share significant factors that affect language learning such as the language environment or the language distance between their L1 and English. In terms of the language environment factor, English learners in ESL countries tend to achieve higher scores in international English proficiency tests than those in EFL countries (ETS, 2018, 2019, 2020, 2021). Moreover, people with L1s which are close to English can learn English faster than ones whose L1 are distant from English (e.g., Nishikawa-Van Eester, 2016; XinXin, Xiaolan and Ahmed, 2022). Thus, although Spain or France are also EFL countries, Spanish and French are more similar to English than Japanese or Chinese. In this case, it becomes more difficult to judge whether the reason for their higher English proficiency depends on their English textbooks or their language distance to English. Therefore, it seems to be more meaningful to compare the textbooks used in Taiwan and Japan, which share similar language environment and distance to English, but where English language attainment, as indicated by scores in international English proficiency tests, is higher in one country

(Taiwan) than in another (Japan). This comparison can make the present study to focus on the learning in the class or textbooks, which are main resources of EFL students' English learning.

Moreover, the textbooks students use in school can be a key material to improve their English. As Milton (2009) mentions, the textbook EFL students use can be a principle source of English input. Therefore, vocabulary contained in EFL textbooks can directly affect the vocabulary students can learn. To date, few studies have focused on newly-introduced lemmas in textbooks so far. Thus, it will be helpful to investigate the vocabulary used and introduced in Japanese and Taiwanese textbooks because it can show which and how much vocabulary students can potentially learn. Thus, this study examines the vocabulary in Japanese and Taiwanese textbooks as well as the differences between them in terms of how much vocabulary is repeated.

To analyse the vocabulary learning opportunities offered by senior high-school English textbooks used in Japan and Taiwan, this study makes two sets of research questions. One set of research questions is related to analysing the inclusion of high-frequency words in these textbooks. Nakayama (2021, 2022) reports that both junior and senior high school textbooks used in Japan do not contain various high-frequency words. Because the textbook can be a main source of English input for EFL students, it is useful to check a) the proportion of word tokens of the target textbooks covered by high-frequency lemmas and b) how many of high-frequency lemmas are in the textbooks. These topics are examined in the first set of

research questions in this thesis (see Table 3). In the comparison between the Japanese and Taiwanese textbooks in question, the present study can reveal differences in the learning opportunities for high-frequency lemmas in the textbooks.

The second set of research questions is concerned with scrutinizing newly-introduced lemmas in the Japanese and Taiwanese textbooks. First, the present study will investigate the word frequency levels of newly-introduced lemmas the Japanese and Taiwanese textbooks focus on during senior high school. High-frequency vocabulary knowledge is definitely essential, but mid-frequency vocabulary knowledge can also contribute to 95-98% of lexical coverages. By examining differences in the frequency levels between the Japanese and Taiwanese textbooks, it will be able to observe which textbook series can promote their students' vocabulary development. Second, vocabulary repetition is something unignorable in vocabulary learning (Uchihara et al, 2019). The number of the repetition does not guarantee that students can successfully learn target words, but the more encounters with target words students can have, the more opportunities they can have to learn them (e.g., Webb and Nation, 2017). Thus, it seems to be significant to examine how many times students can encounter the newly-introduced lemmas in the textbooks. Third, vocabulary repetition is significant, but it should be spaced. Webb and Nation (2017) argue that vocabulary students cannot reencounter for a long time easily deteriorates. Therefore, the present study will show how many of newly-introduced lemmas are recycled across units of each textbook. At the last question, the present study attempts to reveal how many of newly-introduced

lemmas in the Japanese and Taiwanese textbooks are theoretically learnable for student from results of vocabulary repetition and its distances. Webb (2007) suggests that 10 or more encounter are needed to learn both receptive and productive vocabulary knowledge. In addition to this, Webb and Nation (2017) mention that space of repetitions can be one of the significant vocabulary learning factors. Therefore, the present study conducts an investigation to theoretically determine the number of potentially learnable newly-introduced lemmas through the textbooks, considering criteria such as lemmas being used 10 times and more and appearing more than three units of the textbook.

In sum, the research questions in this study are following two set of research questions. The first set of research questions 1 and 2 focuses on the use and learning opportunities of high-frequency words in the target textbooks. The second set of research questions 3 to 6 is to examine the newly-introduced lemmas and their learning opportunities in the target textbooks. Table 2.3 summarizes the research questions in this study.

Table 2.3: The summary of research questions

The first set of research questions
1. How much do high-frequency lemmas account for the lexical coverage in the Japanese and Taiwanese textbook series examined in this study? 2. How many of high-frequency lemmas are in the Japanese and Taiwanese textbook series examined in this study?
The second set of research questions
3. Which frequency bands in the BNC-6318 do newly-introduced lemmas in the Japanese and Taiwanese textbooks come from? 4. How many times do the newly-introduced lemmas occur in the Japanese and Taiwanese textbooks? 5. How many newly-introduced lemmas recur across units of the Japanese and Taiwanese textbooks? 6. How many of the newly-introduced lemmas are, in theory, learnable through the Japanese and Taiwanese textbooks?

Chapter 3 Methodology

3.1 Introduction

Chapter 2 elaborated on vocabulary learning research and described the objectives of this investigation. Chapter 3 will provide the rationale for the methodological decisions made in this study as well as a summary of its approach. The reasons why Taiwan was selected as the comparative target of Japan are explained in Section 3.2. How textbooks were chosen as study targets will be described in Section 3.3. In Section 3.4, two frequency-based lemmatized wordlists, new-GSL by Brezina and Gablasova (2015) and BNC-6318 by Kilgarriff (2006), that were selected to search the textbooks for high- and mid-frequency words for this study will be described, along with an explanation of their selection process. The methodology employed to investigate each research question is summarised in Section 3.5 and 3.6.

3.2 Taiwan as a comparison target for Japan in terms of English language teaching and learning

The objectives of the present study are to examine the vocabulary in senior high school English textbooks used in Japan and Taiwan so as to identify differences in the vocabulary learning opportunities these textbooks provide. Other factors, such as teaching quality or students' motivation to learn English, can affect learners'

English language attainment and, subsequently, mean English proficiency test scores in each country. Yet, as Milton (2009) suggests, EFL textbooks can be the primary source of English for EFL students. Moreover, the vocabulary EFL students can learn through their textbooks can largely affect their vocabulary development. Hence, as an initial step towards investigating the causes behind the gap in mean English proficiency test scores between Japan and Taiwan, this study focuses on the EFL textbooks used in these countries, especially on the vocabulary in these textbooks. Through the findings of the comparison the Japanese and the Taiwanese textbooks, this study attempts to suggest some potential improvements for Japanese English textbooks based.

Japan and Taiwan appear to be one of the most appropriate pairs for comparing English language learning and teaching across nations. It is possible to compare the English language instruction in classrooms and textbooks in these two nations because they satisfy certain essential criteria. The criteria that needed to be met for two countries to be comparable in terms of their senior high-school textbooks in this study are the following: 1) EFL learners in one of the countries have, on average, higher English proficiency scores in international English tests such as TOEFL or IELTS than EFL learners in the other country, 2) the language environments of the two countries are comparable, and 3) the L1s of both countries have the same language distance to English. The rest of this section will elaborate on how these three criteria are met by comparing EFL secondary school textbooks in Japan and Taiwan.

In relation to the first criterion, the comparison target should be a country with higher average English proficiency than Japan. This is because gaps between the English proficiency of students in Japan and those in another country allow this study to suggest some potential improvements for the Japanese textbooks by identifying some differences in the vocabulary learning opportunities offered by EFL textbooks in each country. According to the summary of TOEFL iBT score data (ETS, 2018, 2019, 2020, 2021) for Reading, Writing, Listening, and Speaking, Japan has always ranked lower than China, South Korea and Taiwan. There are no statistically significant differences among the three Asian nations, but Japanese test-takers' scores are lower. Table 3.1 displays the differences in the total scores of four English skills on TOEFL iBT among the four Asian countries from 2018 to 2021.

Table 3.1: The total test scores of Reading, Writing, Listening and Speaking among China, South Korea, Taiwan and Japan on TOEFL iBT

	2018	2019	2020	2021
China	80	81	87	87
Republic of Korea	84	83	86	86
Taiwan	82	83	85	87
Japan	71	72	73	73

It is evident from Table 3.1 that Japanese test-takers score lower than test-takers from other Asian nations. Therefore, it is possible to consider that the English proficiency level of the Japanese appears to be lower than that of people in other three Asian countries. This fact aligns with the first criterion suggesting that there

should be a difference in English proficiency level between the nations compared in this study.

As the second criterion, this study acknowledges that the language environment is also an essential factor in the selection of target countries in comparative research in language learning. According to Krashen (1981), the environment is a crucial factor for language learners. Those who reside in the English environments can acquire English through means other than study, whereas people who live in EFL countries rely mainly on classroom teaching and textbooks for English learning. Moreover, those who live in a country with English as a second language (ESL) tend to show much higher English proficiency than people in EFL environments. For example, some ESL countries have higher total scores in TOEFL iBT than EFL countries. For example, Singapore's score is 98, India's is 96 and the Philippines' is 90 (ETS, 2021). Consequently, comparing ESL and EFL countries in terms of English teaching can make comparisons of vocabulary input from textbooks unclear or irrelevant because vocabulary input outside textbooks is very different between ESL and EFL countries. Since Japan and Taiwan are both EFL countries, their language environments are comparable.

The final criterion for selecting the comparison target in this study is the concept of "language distance," also known as "linguistic distance." According to Nerbonne and Hinrichs (2006), language distance can be defined as language similarity, which means the degree to which two languages share characteristics such as pronunciation, syntax, and semantics. In general, the greater the distance

between the target language and L1, the more difficult it is for learners to learn the target language (Nishikawa-Van Eester, 2016). For example, the Foreign Service Institute (https://www.state.gov/foreign-language-training/), which is part of the United States Department of State, estimates the difficulty of language acquisition for native English speakers. Japanese and Chinese, according to the website, are the most distant languages from English. This is due to the fact that every aspect of the two languages, including pronunciation, syntax, and semantics, is entirely distinct from English.

Based on these three criteria, China, South Korea, and Taiwan seem suitable as comparative countries for this study concerning Japan. However, due to challenges such as the impact of the COVID-19 pandemic, it became exceptionally challenging to identify and obtain senior high school textbooks used in China and South Korea. Fortunately, I managed to obtain an electronic copy of the senior high school Taiwanese textbook series from a teacher of English working in Taiwan. As a result, this study compared English senior high school textbooks in Japan and Taiwan. In the future, I aim to conduct similar research in order to identify the differences in the vocabulary learning opportunities provided by textbooks used in China, Korea and Japan.

3.3 Textbooks analysed in this study

In order to select appropriate textbooks, this study establishes one selection criterion, the penetration of textbooks in Japan and Taiwan. With respect to Japan,

the selection of textbooks is important factor because there are many textbooks available in Japan. The use of widely-used textbooks would make the findings of this study more applicable to Japanese English education. Therefore, the penetration of textbooks should play a significant role in selecting the appropriate ones for this study. But one issue with this condition is that textbook penetration appears to vary across Japan's prefectures, cities, and towns. Because the selection of textbooks used in Japanese senior high schools is determined by the Education Commission in each city, it is impossible to determine which textbook is the most widely used in Japan. Thus, this study concentrates solely on Tokyo, where the number of high schools is the greatest in Japan. In addition, the Tokyo Metropolitan Board of Education (2021) illustrates the adoption rates of senior high level English textbooks in Tokyo. Based on the data, the present study selected one textbook series widely used in Tokyo's senior high schools.

This study has selected one textbook series from each country, "English" by San min as the representative of Taiwan and "All aboard!" by TOKYO SHOSEKI as the representative of Japan. The Taiwanese Ministry of Education has approved San min's "English" as one of the country's most important textbooks. In Japan, the MEXT has authorized the "All aboard!" series. This series is one of the most popular and widely used textbook series in Tokyo, covering all three years of high school.

The "English" series consists of six textbooks for the three years of senior high school, with two textbooks per grade. In addition, the number of units in Taiwanese textbooks is 12 for the first and second years of senior high school and

10 for the third year. The All aboard! series of Japanese textbooks, on the other hand, consists of three textbooks, one for each grade.

3.4 The wordlists used in this study

Several wordlists are frequently employed in vocabulary research. Depending on the word unit employed, the contents of the wordlists may vary slightly. The BNC wordlist (Nation, 2006) is an example of wordlists that utilize "word family" as the word unit. In addition, the counting word unit of The New General Service List by Browne et al. (2013) is the "flemma." The new-GSL by Brezina and Gablasova (2015) uses lemma as its unit. Among these different word units, it is essential to select an appropriate one for this study to examine the vocabulary used in the target textbooks and provide students with more effective vocabulary learning through the textbooks.

This study uses "lemma" as the word unit because it contains only headwords and the most frequent inflections. The word units "word family" and "flemma" were dis-preferred because they tend to encompass too many related words. For instance, a word family includes all inflected and derived forms of a headword. Due to the lack of POS tagging in word family wordlists, words with the same spelling are also counted together despite the fact that their frequency and meaning may vary depending on POS. Brown, et al. (2022) claims that “most L2 learners can deal with lemmas but lack the morphological knowledge to make “word

family” a suitable unit” (p.5). The "word family" wordlists operate under the assumption that students possess the ability to deduce the meanings of unfamiliar words by utilizing their understanding of derivational affixes and inflectional suffixes (Coxhead, 2000). However, empirical evidence does not support this hypothesis (e.g., Ward & Chuenjundaeng, 2009). Furthermore, in word-family wordlists, the absence of POS tagging results in the aggregation of words with identical spellings, despite the fact that their frequency and semantic significance may vary based on POS. Likely, the word "flemma" consists of a headword, the most common inflections, and multiple POS. For example, “edit” can be both the noun and the verb with the same form. If we count “edit” according to flemmas, it would be only one flemma. However, it can be questionable whether a learner who know a word with one POS also knows the word with a different POS. According to Stoeckel, Ishii and Bennett (2020), only 56% of target words can be translated correctly between the different POS of words. In summary, knowing one member of the "word family" or "flemma" does not guarantee that learners can comprehend the other inflections, derivations, or POS meanings of the target word (e.g., Brown et al., 2020; McLean, 2018; Stoeckel, Ishii and Bennett, 2020). From the point, “lemma” can be the best word unit for the present study to analyse the vocabulary in target textbooks.

This study selects two lemmatized wordlists to identify the high- and mid-frequency vocabulary employed in Japanese and Taiwanese textbooks. One is the new-GSL developed by Brezina and Gablasova (2015), and the other is the lemmatized frequency wordlist developed by Kilgarriff based on the BNC (2006). The

former is a high-frequency wordlist containing 2494 of the most common lemmas. The latter includes both high- and mid-frequency lemmas, bringing the total number of lemmas on the list to 6318. In this study, the BNC-3000 was created by extracting the most frequent 3000 lemmas from the BNC-6318 so as to investigate the text coverage that high-frequency lemmas from two wordlists provide in the target textbooks. Using both the new-GSL and BNC-3000 wordlists allows for a more thorough investigation of the coverage that high-frequency lemmas provide the target textbooks. The overlap between the two high-frequency lemmatized wordlists is 91%.

The BNC-6318 can be useful during the examination of the vocabulary level of newly introduced lemmas in the textbooks. This study used the BNC-6318 wordlist because using only high-frequency wordlists would prevent this study from examining the use and introduction of mid-frequency lemmas in textbooks. Mid-frequency words are also essential for students to learn so that they can comprehend authentic English texts (Schmitt and Schmitt, 2014). Therefore, the ability to examine the use of mid-frequency lemmas in the textbooks can enhance the utility of the present study for EFL materials design.

3.5 Corpus compilation from Japanese and Taiwanese textbooks

The textbooks were pre-processed in three steps: 1) the PDFs of target textbooks were converted into text files, 2) some words were removed or converted,

and 3) TreeTagger was used to assign POS to each word. Initially, using AntFileConverter (Anthony, 2022), all textbook files were converted to .txt format. At the second step, the present study eliminated Chinese and Japanese words from the textbooks. These words are unnecessary for analysing the English vocabulary in the textbooks in question and can prevent this study from observing the actual amount of English language input each textbook series provides. Additionally, the Japanese and Taiwanese textbooks include proper nouns such as "Taji mahal" and personal names. The kind of proper nouns is typically counted as two words in general. However, in order to estimate the amount of language input or word tokens students can acquire as English input through the textbooks, these words were counted as a single word by converting “Taji mahal” into “Tajimahal”. Moreover, non-English letters and symbols, such as International Phonetic Association symbols, the POS of newly-introduced lemmas, U (i.e., "uncountable"), and C (i.e., "countable"), were also eliminated from the data. In addition, all abbreviated form were converted into their standard forms, such as "don't" and "do not" or "couldn't" and "could not," to facilitate the addition of POS tags in the next step.

POS tagging was done via Helmut Schmid's TreeTagger (Schmid, 1995). It is widely recognized as a potent tool within the field of natural language processing due to its exceptional capabilities in morphological analysis. The software's extensive linguistic support increases its usability in a variety of linguistic contexts. One of the primary benefits of Treetagger is its effective management of mid-frequency words. Finally, this software was preferred because it is free.

Following the aforementioned steps, the output of the POS tagging was input in an Excel sheet. Figure 3.1 is a screenshot of part of this Excel sheet, where the first column listed all tokens in the corpus, the second their POS, and the third the lemma each token belonged to. In order to have a lemmatised list of the words in the corpus, the 'Lemma' and 'POS' columns were merged through Excel's CONCAT function to form 'Lemma_POS' entries. In addition to performing this task for the words in the corpus, this task was performed for the new-GSL and BNC wordlists as well so as to later identify shared lemmas between the textbooks and these wordlists via searches in Excel. Figure 3.1 is the example of this process on Excel.

Figure 3.1: A screenshot of the Excel sheet that contained the word tokens in the corpus and their POS, lemmas and concatenated lemmas and POS

TOKEN	POS	LEMMA	LEMMA_POS
Pre-lesso	n	Pre-lesson	Pre-lesson_n
My	pron	my	my_pron
name	n	name	name_n
is	v	be	be_v
KimuraYu	n	KimuraYui	KimuraYui_n
Hi	UH	hi	hi_UH
My	pron	my	my_pron
name	n	name	name_n
is	v	be	be_v
KimuraYu	n	KimuraYui	KimuraYui_n
I	pron	I	I_pron
am	v	be	be_v
fifteen	x	fifteen	fifteen_x
years	n	year	year_n
old	adj	old	old_adj

A "Lemma_POS" list was made for each chapter of each Japanese and Taiwanese

textbook in order to search them for lemmas from the new-GSL and BNC-3000/6318 wordlists. However, while these lists display the simple POS for the lemmas, TreeTagger employs a more detailed POS. Therefore, this study simplified the TreeTagger's POS tags to match those of the new-GSL and BNC-3000/6318 wordlists. Table 3.2 shows the equivalences between TreeTagger's POS tags and the simpler ones that replaced them.

Table 3.2: TreeTagger POS tags and the simpler POS tags that replaced them

TreeTagger	Simpler version	TreeTagger	Simpler version
NN	n	VBP	V
NNS	n	VD	V
NP	n	VDD	V
NPS	n	VDG	V
VB	v	VDN	V
VBD	v	VDZ	V
VBG	v	VDP	V
VBN	v	VH	V
VBZ	v	VHD	V
VBP	v	VVN	V
VD	v	VVZ	V
VDD	v	VVP	V
VDG	v	MD	Mod
VDN	v	JJ	Adj
VDZ	v	JJR	Adj
VDP	v	JJS	Adj
VH	v	RB	Adv
VHD	v	RBR	Adv
VHG	v	RBS	Adv
VHN	v	WRB	Adv
VHZ	v	RP	Avp
VHP	v	IN	Con
VV	v	CC	Con
VVD	v	PP	Pron
VVG	v	PP$	Pron
WP	pron	DT	X

WP$	pron	CD	X
EX	e	PDT	X
TO	t	WDT	X

The definition of each tag and examples are available in the link below:
https://www.laurenceanthony.net/software/tagant/resources/treetagger_tagset.pdf

The POS of each lemma was checked so as to make sure that they are correct. For instance, the POS "t" can be either the infinitive "to" or the preposition "to." Therefore, they are modified the preposition “to” into “to_con” by checking the POS one by one. Moreover, some words could not be lemmatized correctly and some were shown as “unknown”. These words were also checked and modified correctly.

3.6 Data-processing for each research question

This study employs two sets of research questions to examine the learning opportunities for high-frequency lemmas and newly-introduced lemmas in textbooks. The first set of research questions consists of two inquiries: the proportion of high-frequency lemmas found in the word tokens and the variety of high-frequency lemmas found in textbooks. Brezina and Gablasova's new-GSL which includes the most frequent 2494 lemmatized words, and Kilgarriff's BNC-3000, which extracts the most frequent 3000 lemmas from the total 6318 lemmas were employed to investigate these questions.

In order to address these sets of questions the number of word tokens in the Japanese and Taiwanese textbooks was found in the corpus through the procedure summarised in section 3.5. In addition, by counting how many times

lemmas in the new-GSL and BNC-3000 are used in each word token of each unit using the COUNTIF function on Excel, it is possible to determine the word token coverage provided by high-frequency lemmas. For example, if a textbook unit has 4000 word tokens and high-frequency lemmas are used 2000 times in total, high-frequency lemmas provide 50% coverage in this unit. In terms of the lemma analysis, this study eliminated some words that appeared multiple times in order to count the exact number of lemmas in the textbooks. In the state of word token corpus, the inclusion of all lemmas makes it challenging to discern the actual variety of lemmas used in the textbooks. Thus, this process could be helpful to identify differences in the exact number of lemmas Japanese and Taiwanese students can encounter through their textbooks. Moreover, this study compared the corpus data of textbooks with the new-GSL and BNC-3000 wordlists on Excel. This way makes it possible to determine how many of high-frequency lemmas in the wordlists were found in the textbooks at least more than once.

The second set of research questions focuses on newly-introduced lemmas in the Japanese and Taiwanese textbooks. This study specifically examines how many and what levels of vocabulary are newly introduced in the target textbooks, how often students can encounter the target words in each textbook, and how many newly-introduced lemmas are recycled across the units of each textbook. In addition, this study attempts to determine, based on the repetition rates and distances between repetitions within textbook units, how many newly-introduced lemmas students could, in theory, learn from their textbooks.

This study created lists of newly-introduced lemmas in each textbook in order to investigate the number of newly-introduced lemmas and their frequency levels in BNC-6318 created by Kilgarriff (2006). Thus, a comparison is made between nine lists of new words (six for Taiwanese textbooks and three for Japanese textbooks) and the first 1000 (K1) to the last 318 lemmatized BNC wordlist entries. This study aims to determine how many new words are introduced and which frequency level bands they belong to through the first and second research question in the second set, respectively.

The third research question in the second set examined the repetition of newly-introduced lemmas in the target textbooks. In order to investigate this, the present study made five distinct repetition bands: "Once," "2 to 3," "4 to 6," "7 to 9," and "10 or more." By checking how many times newly-introduced lemmas in each new words list of each textbook appeared in the corpus data of Japanese and Taiwanese textbooks using the COUNTIF function in Excel, it is possible to determine the occurrence with which newly-introduced lemmas appear in the textbook's content. After identifying the repetition rate of each newly-introduced lemma in each textbook, these lemmas were divided into the five bands and the differences between the Japanese and Taiwanese textbooks were compared.

The fourth research question in the second set examined the distances between the repetitions of newly-introduced lemmas in each textbook's units. This was accomplished by color-coding each unit of the textbooks. For instance, unit 1 of a textbook is coloured red, unit 2 is coloured yellow, unit 3 is coloured red, and so

on. The newly-introduced lemmas are then extracted from the corpus data of each textbook using the new word lists. Then, they can be displayed with the unit colours. After that, by counting the number of different colours of the target word, it is possible to determine the textbook units in which the target words appeared. For instance, Figure 3.2 below is a part of this procedure. The left line is the word token date of all the units of the Japanese textbook 1. This textbook's newly-introduced lemmas are listed in the "NEW" column. The “CHECK” column examined whether the words in “TOKEN” and ones in “NEW” were shared. By extracting shared words from the “CHECK” line, the “New_textbook” column were generated with the colours of units where the words appeared. By placing the "New_textbook" and "NEW" lines in alphabetical order, it is easier to count different colours for each target words. For example, in Figure 3.2, "above_adv" has a single colour, and the blue colour indicates that the word appeared in Unit 7. Consequently, it is evident that the word is unique to Unit 7 and not reused in other units. On the other hand, “actress_n” has yellow and white colours on Figure 3.2. It means that the lemmas appeared in both Unit 2 or the last Unit of the textbook. Through this analysis, it is possible to determine how newly-introduced lemmas are reused within textbook units.

Figure 3.2: an illustration of the colour coding procedure used in this study

TOKEN	CHECK		NEW			New _text	NEW	result
Pre-lesso	#N/A		bathroom_n			above_adv	above_adv	1
my_pron	#N/A		bedroom_n			above_adv	account_r	1
name_n	#N/A		bookshelf_n			account_r	actress_n	2
be_v	#N/A		cheerful_adj			account_r	actually_a	1
KimuraYu	#N/A		cupboard_n			actress_n	admire_v	2
hi_UH	#N/A		dad_n			actress_n	adult_n	1
my_pron	#N/A		drama_n			actress_n	aggressive	1
name_n	#N/A		entrance_n			actually_a	agricultura	1
be_v	#N/A		fast-food_adj			actually_a	airplane_r	1
KimuraYu	#N/A		Iceland_n			admire_v	alive_adj	1
I_pron	#N/A		mathematics_n			admire_v	although_	1
be_v	#N/A		shy_adj			admire_v	amazed_a	1
fifteen_x	#N/A		talkative_adj			admire_v	amazing_a	3
year_n	#N/A		thick_adj			adult_n	anime_n	2
old_adj	#N/A		vase_n			adult_n	annual_ac	1

Chapter 4 Results and Discussion

4.1 Introduction

This chapter will report and discuss the results related to each of the two sets of research questions in this study. Sections 4.2 addresses the first set of research questions. Sections 4.3 is a reminder of the second set of research questions. Sections 4.4 to 4.7 each report the results related to each research question in the second set and discuss the findings of each research questions.

4.2 Results in relation to the first set of research questions

The first set of research questions investigates the component of high-frequency lemmas in the Japanese and Taiwanese textbooks through word token and lemma analyses. These research questions are:

1. How much do high-frequency lemmas account for the lexical coverage in the Japanese and Taiwanese textbook series examined in this study?
2. How many of high-frequency lemmas are in the Japanese and Taiwanese textbook series examined in this study?

This set of questions was first addressed by examining the coverage of word tokens by the new-GSL-2500 wordlist created by Brezina and Gablasova (2015) and the BNC-3000 wordlist, which was extracted from the BNC-6318 wordlist created by Kilgarriff (2006). Then the new-GSL-2500 and BNC-3000 lemmas were identified in the textbooks. In addition, the differences of the number of the word tokens and

lemmas in Japanese and Taiwanese textbooks are analysed. High-frequency lemmas were identified in these textbooks as a way of evaluating Japanese and Taiwanese high-school students' English vocabulary input from textbooks for two reasons. First, high-frequency lemmas are very important in English learning, because they can account for approximately 80% of word token in authentic English texts (e.g., Nation, 2006; Webb and Nation, 2017). Second, textbooks are not only one source for students, but they can be the principle source, especially in EFL countries (Milton, 2009). Therefore, if textbooks do not provide sufficient learning opportunities for high-frequency vocabulary, it might be difficult for students to even encounter these very useful words. Moreover, the percentages of high-frequency lemmas present in the content of the target textbooks can serve as a reliable indicator for determining whether the vocabulary employed is suitable for students (Sun and Dang, 2020).

Tables 4.1 and 4.2 show the tokens in each Taiwanese textbook ('TT1' to 'TT6') and in each Japanese textbook ('JT1' to 'JT3'), respectively, that correspond to new-GSL-2500 and BNC-3000 lemmas and the proportion of word tokens that these high-frequency lemma tokens occupy in total and per textbook unit.

Table 4.1: Percentages of the total number of word tokens and New-GSL-2500 lemma tokens per Japanese and Taiwanese textbook and textbook unit; the total new-GSL-2500 lemma tokens per unit appear within parentheses

	TOKEN	Unit1	Unit2	Unit3	Unit4	Unit5	Unit6	Unit7	Unit8	Unit9	Unit10	Unit11	Unit12	TOTAL	Average
TT1	45013	85%(2989)	85%(3036)	83%(3018)	87%(3125)	85%(2746)	83%(2920)	84%(2937)	82%(3031)	84%(3241)	88%(3795)	86%(3527)	86%(3683)	85%(38051)	85%(3170)
TT2	48079	84%(3227)	86%(3584)	83%(3343)	89%(3286)	87%(3114)	84%(3539)	86%(3418)	83%(3518)	83%(3228)	83%(3574)	83%(3562)	82%(3203)	84%(40506)	84%(3383)
TT3	55013	83%(3300)	85%(3789)	85%(4305)	85%(3774)	79%(3632)	84%(3688)	85%(4169)	80%(3434)	82%(3394)	83%(4001)	75%(4103)	81%(3639)	82%(45228)	82%(3773)
TT4	52400	82%(4014)	86%(4026)	79%(3368)	83%(3584)	82%(4244)	81%(3513)	80%(3329)	83%(3407)	83%(4021)	82%(3277)	83%(2964)	82%(3328)	82%(43075)	82%(3589)
TT5	41261	84%(4092)	80%(3205)	81%(3252)	82%(3255)	82%(3558)	81%(2799)	81%(3279)	84%(3451)	82%(3962)	83%(3040)	N/A	N/A	82%(33893)	82%(3389)
TT6	42129	83%(3277)	82%(3600)	82%(3385)	81%(3758)	81%(3223)	78%(2930)	84%(3216)	80%(3115)	82%(4250)	85%(3770)	N/A	N/A	82%(34524)	82%(3452)
TOTAL	283895														
JT1	8011	80%(604)	80%(708)	72%(361)	75%(339)	82%(1028)	76%(338)	78%(374)	80%(658)	81%(555)	78%(1367)	N/A	N/A	79%(6332)	78%(633)
JT2	10236	75%(388)	74%(502)	80%(647)	76%(624)	75%(1082)	76%(679)	80%(713)	81%(778)	77%(651)	77%(1843)	N/A	N/A	77%(7907)	77%(790)
JT3	10735	78%(432)	75%(519)	72%(413)	77%(688)	86%(1406)	80%(674)	73%(680)	83%(904)	78%(708)	78%(2048)	N/A	N/A	80%(8472)	78%(847)
TOTAL	28982														

Table 4.2: Percentages of the total number of word tokens and BNC-3000 lemma tokens per Japanese and Taiwanese textbook and textbook unit; the total BNC-3000 lemmas tokens per unit appear within parentheses

	TOKEN	Unit1	Unit2	Unit3	Unit4	Unit5	Unit6	Unit7	Unit8	Unit9	Unit10	Unit11	Unit12	Total	Average
TT1	45013	89%(3254)	89%(3191)	87%(3165)	90%(3248)	88%(2858)	87%(3068	87%(3051)	86%(3200)	88%(3404)	92%(3976)	90%(3685)	90%(3841)	89%(39941)	89%(3328)
TT2	48079	89%(3440)	89%(3745)	88%(3520)	92%(3395)	91%(3243)	89%(3725)	90%(3581)	89%(3745)	86%(3339)	87%(3748)	86%(3714)	88%(3434)	89%(42627)	89%(3552)
TT3	55013	88%(3486)	87%(3900)	89%(4499)	90%(3989)	84%(3830)	88%(3862)	88%(4315)	85%(3642)	88%(3641)	87%(4147)	77%(4247)	86%(3960)	86%(47518)	86%(3959)
TT4	52400	86%(4197)	89%(4200)	83%(3540)	87%(3745)	85%(4394)	85%(3700)	83%(3457)	84%(3517)	86%(4184)	86%(3445)	86%(3054)	86%(3514)	86%(44947)	86%(3745)
TT5	41261	88%(4277)	82%(3301)	86%(3457)	86%(3420)	85%(3685)	84%(2906)	85%(3443)	88%(3633)	86%(4209)	87%(3199)	N/A	N/A	86%(35530)	86%(3553)
TT6	42129	87%(3431)	86%(3772)	86%(3552)	86%(3961)	84%(3353)	82%(3096)	87%(3331)	86%(3310)	87%(4490)	87%(3875)	N/A	N/A	86%(36171)	86%(3617)
TOTAL	283895														
JT1	8011	82%(623)	83%(741)	73%(367)	81%(364)	86%(1085)	78%(350)	82%(392)	85%(699)	84%(576)	83%(1444)	N/A	N/A	83%(6641)	82%(664)
JT2	10236	78%(404)	77%(525)	81%(660)	82%(668)	81%(1165)	77%(689)	84%(751)	84%(805)	80%(670)	82%(1966)	N/A	N/A	81%(8303)	81%(830)
JT3	10735	81%(448)	76%(530)	79%(452)	81%(724)	89%(1467)	82%(688)	77%(717)	85%(930)	80%(727)	83%(2182)	N/A	N/A	83%(8865)	81%(886)
TOTAL	28982														

In Tables 4.1 and 4.2, the Column “Token” indicates that the Taiwanese textbooks have nearly ten times more word tokens (283,895 tokens in total) than the Japanese textbooks (28,982 tokens in total). This is the number of the whole English vocabulary in the target textbooks of each country. Therefore, this result indicates that the Taiwanese senior high school students can read ten times more English than the Japanese students. This difference in terms of word tokens between Japanese and Taiwanese textbooks exists across proficiency levels as well, since the mean token number across Taiwanese textbooks is 47,315.83 (SD = 4,944), while the mean token number across Japanese textbooks is 9,660.67 (SD = 1,184).

This high-frequency word token analysis also indicates another difference between the Japanese and Taiwanese textbooks. The “Average” columns in Tables 4.1 and 4.2 indicate that the Taiwanese textbooks use more high-frequency lemmas than the Japanese textbooks per unit across levels. For example, 82-85% of the word tokens are new-GSL-2500 high-frequency lemmas in each Taiwanese textbook in Table 4.1 and 86-89% are BNC-3000 in Table 4.2. On the other hand, the proportion of the high-frequency lemmas in the Japanese textbooks is less than 80% in Table 4.1 and around 80% in Table 4.2. This fact

means that the new-GSL-2500 and the BNC-3000 wordlists provide more coverage for the Taiwanese textbooks than for the Japanese textbooks across high-school years.

Tables 4.3 and 4.4 below show the total number of high-frequency lemmas used in the Japanese and Taiwanese textbooks and also show how many of them belong to the new-GSL-2500 and BNC-3000 wordlists. Moreover, Tables 4.3 and 4.4 also show how many high-frequency lemmas from the first 1000 (K1) to the last 500 (Table 4.3) and the third 1000 (K3 on Table 4.4) account for the total lemmas in the textbooks. In other words, Tables 4.3 and 4.4 show the differences in the amount of the high-frequency words students can encounter in the Japanese and Taiwanese textbooks.

Table 4.3: Percentages and the total number of high-frequency lemmas of the new-GSL-2500 per the Japanese and Taiwanese textbook; the actual number of high-frequency lemmas in each frequency band appears within parentheses

	LEMMA	New-GSL-K1	K2	the last 500	New-GSL-2500
TT1	2937	30%(867)	19%(559)	6%(190)	55%(1616)
TT2	3447	26%(902)	19%(658)	7%(225)	52%(1785)
TT3	4282	22%(958)	18%(772)	6%(275)	47%(2005)
TT4	4553	21%(951)	17%(778)	7%(296)	44%(2025)
TT5	4215	22%(948)	18%(773)	7%(277)	47%(1998)
TT6	4370	21%(936)	18%(782)	6%(262)	45%(1980)
TOTAL	9039	11%(981)	11%(971)	5%(443)	26%(2395)
JT1	1212	42%(506)	18%(217)	5%(64)	65%(787)
JT2	1620	35%(574)	16%(255)	6%(95)	57%(924)
JT3	1690	34%(648)	17%(288)	4%(75)	60%(1011)
TOTAL	2936	27%(784)	16%(483)	6%(166)	49%(1433)

Table 4.4: Percentages and the total number of high-frequency lemmas of the BNC-3000 per the Japanese and Taiwanese textbook; the actual number of high-frequency lemmas in each frequency band appears within parentheses

	LEMMA	BNC-K1	K2	K3	BNC-3000
TT1	2937	29%(839)	18%(520)	12%(356)	58%(1715)
TT2	3447	25%(871)	18%(634)	12%(424)	56%(1929)
TT3	4282	22%(922)	17%(744)	12%(513)	51%(2179)
TT4	4553	20%(921)	16%(742)	12%(567)	49%(2230)
TT5	4215	22%(919)	17%(737)	12%(512)	51%(2168)
TT6	4370	21%(914)	17%(722)	12%(538)	50%(2174)
TOTAL	9039	11%(963)	10%(937)	9%(853)	30%(2753)
JT1	1212	42%(503)	17%(203)	10%(125)	69%(831)
JT2	1620	35%(566)	14%(229)	11%(175)	60%(970)
JT3	1690	38%(636)	16%(277)	8%(139)	62%(1052)
TOTAL	2936	26%(766)	15%(455)	11%(313)	52%(1534)

According to Tables 4.3 and 4.4, the Taiwanese textbooks have 9,039 lemmas in total while the Japanese textbooks only have 2,936 lemmas. Thus, regarding the number of lemmas students can encounter through their textbooks, the Taiwanese textbooks have nearly three times more lemmas than the Japanese textbooks. This means that the Taiwanese textbooks can expose students to a wider variety of words than the Japanese ones. Moreover, out of the lemmas in the Taiwanese textbooks, the new-GSL-2500 and BNC-3000 lemmas account for 26% (2,395 words) and 30% (2,753 words) in total. Therefore, the remaining approximately 70% of the total lemmas should be mid-frequency or low-frequency lemmas. However, the Japanese textbooks contain 1,433 new-GSL lemmas (49%) and 1,534 BNC-3000 lemmas (52%) among the total lemmas. Accordingly, the ability of Taiwanese textbooks to display mid-frequency or low-frequency lemmas is approximately 70%, whereas the capability of Japanese textbooks to do so is approximately 50%. These results might be affected by the differences in the number of word tokens between Japanese and Taiwanese textbooks. Moreover, Tables 4.3 and 4.4 indicate that the Taiwanese textbooks contain 2,395 lemmas (96%) of the new-GSL-2500 wordlist and 2,753 lemmas (92%) of the BNC-3000 wordlist across all the Taiwanese textbooks. However,

the Japanese textbooks use 1,433 lemmas (57%) from the new-GSL-2500 wordlist and 1,534 lemmas (51%) from the BNC-3000 wordlist. That is, while the Taiwanese textbooks can provide students with over 90% of the new-GSL-2500 or BNC-3000 lemmas, the Japanese textbooks can only provide around 50% of these high-frequency lemmas. This comparison of the variety of high-frequency lemmas students can encounter through the contents of textbooks reveals that the Taiwanese textbooks can provide their students with more opportunities to learn or at least encounter essential words than the Japanese textbooks.

In sum, the findings in the research question 1 and 2 are similar with those in Nakayama (2021, 2022). In Nakayama (2021, 2022), it is concluded that high-frequency words account for a high proportion of texts in Japanese textbooks, but the variety of high-frequency flemmas in the New General Service List (NGSL)(Browne, Culligan, Phillips, 2013) used in the target textbooks was limited. Similarly, in the present study, both Japanese and Taiwanese textbooks showed the high proportion of high-frequency lemmas on Tables 4.1 and 4.2 (approximately 80% in the Japanese textbooks and around 85% in the Taiwanese textbooks). In terms of the variety of high-frequency lemmas, however, Tables 4.3 and 4.3 showed that the Japanese textbooks contained only around 50% of the

lemmas in the two high-frequency wordlists used in this study, whereas the Taiwanese textbooks covered over 90% of high-frequency lemmas in the wordlists. Since Nation (2006) suggests that high-frequency words can account for approximately 80% of word tokens in authentic English texts, the figures of the Japanese textbooks seem not to be low. However, the findings in the present study also indicate that the Japanese textbooks provide their students only with the limited variety of the essential lemmas of the textbooks.

4.2.1 Discussion of the findings in relation to the first set of research questions

This study revealed differences between the Japanese and Taiwanese textbooks in terms of a) the number of word tokens and lemmas, b) the textbook coverage that high-frequency lemmas provide and c) the variety of high-frequency lemmas. The findings of the first set of research questions are the following:

1) The word token analysis revealed that high-frequency lemmas provide higher coverage of the Taiwanese textbooks than of the Japanese textbooks.

2) The Taiwanese textbooks have approximately ten times more word tokens than

the Japanese textbooks.

3) The Taiwanese textbooks include over 90% of the lemmas belonging to the new-GSL-2500 and BNC-3000 wordlists while the Japanese textbooks use only around 50% of the lemmas in each wordlist.

4) The Taiwanese textbooks have nearly three times more lemmas than the Japanese textbooks.

According to these findings, the Taiwanese textbooks are superior to the Japanese textbooks because the former expose students to more vocabulary input in general and to more high-frequency vocabulary in particular than the latter. The following section will explore why the aforementioned differences in terms of word token numbers, lemma numbers and high-frequency word inclusion between the Japanese and Taiwanese textbooks have occurred.

4.2.2 Differences in the number of word tokens between the Japanese and Taiwanese textbooks

The first finding is that the Taiwanese textbooks had nearly ten times more word tokens than the Japanese textbooks. The simplest reason for this fact is be that the Taiwanese textbooks are written in English except for the

introduction of the meanings of newly-introduced lemmas. By contrast, the Japanese textbooks use a lot of Japanese to explain the target grammar in the unit or the introduction of the practice exercises. However, L1 use can be helpful for EFL learners in English learning if it is used appropriately. As Nation (2003) suggested, the use of the students' L1 in L2 learning can be helpful but its overuse should be avoided. Moreover, Atkinson (1987) suggested that using the students' L1 was helpful in eliciting language from students and checking their comprehension. Therefore, it is important to examine where the Japanese textbooks tend to use the L1 in the textbooks.

According to Nation (2003), there seem to be four possible occasions where the use of L1 is more likely to be beneficial for L2 learning. They are when giving instructions, explaining the meaning of words, complex ideas and difficult grammar rules. Actually, these suggestions about the use of the L1 are for teachers and students in the classroom, but it seems to be meaningful to apply them to the textbooks and check whether the Japanese textbooks use the L1 as Nation (2003) and Atkinson (1987) suggested.

The Japanese textbooks used the L1 in the following places:

1. The introduction of the main reading text of the unit and the reading points

which students should focus on in the text.

2. The explanation of the learning goal in the unit
3. The introduction of the meaning of words
4. The explanation of the target grammar
5. Instructions in target-grammar practice activities
6. The translation of examples of the grammar and sentences in target-grammar practice activities

Although it is not necessary to use the L1 in 1 and 2 above, the rest of these L1 uses can be useful. Points 3, 4 and 5 seem to be included in Nation's (2003) recommended L1 use in the class. In terms of point 6, the Japanese textbooks use many questions which require students to elicit English words from the Japanese translation to fill a blank in an English sentence. Thus, this one also seems to fit with Atkinson's (1987) suggestion. Therefore, it might be possible to say that the use of L1 in the Japanese textbooks seems to have certain meanings. Yet, even though the L1 use in the Japanese textbooks seems to have some rationales, the difference in the word tokens between the Japanese and Taiwanese textbooks is crucial. Therefore, the Japanese textbooks need to find some ways to compensate the difference in the word token between the

Japanese and Taiwanese textbooks Otherwise, it might be difficult for Japanese students to have opportunities to learn or encounter various words as the Taiwanese students can do. Finally, it's important to mention that empirical research is needed to test whether the L1 use in the Japanese textbooks can be really helpful for students or the L1 uses are truly valuable for students' learning.

4.2.3 The differences between the Japanese and Taiwanese textbooks in terms of high-frequency coverage and words necessary for 95-98% of vocabulary coverages

In terms of the proportions of high-frequency lemmas in the Japanese and Taiwanese textbooks, this study found that while the Taiwanese textbooks showed higher word token coverage by the new-GSL-2500 and BNC-3000 wordlists. There seem to be two possible reasons for the higher coverage provided by the new-GSL-2500 and BNC-3000 in the Taiwanese textbooks than in the Japanese ones. The first one is that the Japanese textbooks might use more words which are not in the new-GSL-2500 or BNC-3000 wordlist than the Taiwanese textbooks. In this case, since students are required to have mid-or low-frequency words, it should be more difficult to achieve 95-98% of vocabulary

coverage for their sufficient reading comprehension. The second possible reason is that the Japanese textbooks might not be able to recycle the high-frequency lemmas as much as the Taiwanese ones because of the smaller number of word tokens. To examine the extent to which each of these two possible hypothesises are correct, this research analysed the textbooks with the BNC-6318 wordlist (Kilgarriff, 2006). The wordlist can be suitable for this analysis because it contains not only high-frequency lemmas, but also some mid-frequency ones. By using the BNC-6318 wordlist (Kilgarriff, 2006), it is possible to ascertain whether the Japanese textbooks use a lot of mid-frequency lemmas or there are differences in the use of high-frequency lemmas between the Japanese and Taiwanese textbooks. Additionally, this analysis can also show how much vocabulary levels are needed to achieve 95-98% of vocabulary coverage to sufficiently comprehend the contents of the textbooks. The results are shown in Table 4.5 below. ‘K1’ stands for the most frequent 1000 lemmas in the BNC and the rest of the acronyms for the remaining 1000-lemma bands in descending order of frequency in the BNC. ‘K7’ stands for the last 318 lemmas.

Table 4.5: The percentages of textbook coverage by the BNC-6318 wordlist bands; the parentheses show the actual number of lemmas in each frequency band; the last column shows the coverage from lemmas outside the BNC-6318 wordlist

	BNC-K1	K2	K3	K4	K5	K5	K7	Others
TT1	77%(34524)	9%(3882)	3%(1535)	2%(954)	1%(566)	0.9%(416)	0.2%(79)	7%(3060)
TT2	77%(36353)	9%(4379)	4%(1893)	2%(886)	1%(681)	1%(459)	0.2%(120)	7%(3306)
TT3	74%(40726)	9%(4733)	4%(1893)	2%(1079)	1%(716)	0.9%(521)	0.2%(111)	8%(4557)
TT4	73%(38168)	9%(4755)	4%(2024)	2%(1050)	2%(804)	1%(712)	0.2%(130)	9%(4757)
TT5	73%(30041)	10%(3932)	4%(1557)	2%(1020)	2%(652)	1%(487)	0.3%(133)	8%(3439)
TT6	74%(31112)	9%(3592)	3%(1469)	2%(821)	2%(710)	1%(505)	0.3%(112)	9%(3808)
JT1	71%(5665	8%(644	4%(332)	2%(157)	1%(107)	1%(97)	0.3%(21)	12%(988)
JT2	69%(7078	8%(768	4%(457)	2%(229)	2%(188)	1%(123)	0.3%(34)	13%(1359)
JT3	70%(7554	9%(937	3%(374)	2%(233)	2%(176)	1%(122)	0.2%(22)	12%(1313)

Table 4.5 clearly indicated that “K1” column and the “Others” column seem to cause the differences in the word token coverages of the high-frequency lemmas between the Taiwanese and the Japanese textbooks. Seeing the “K1” columns on Table 4.5, the Taiwanese textbooks contain 21,0924 K1 words in total, which accounted for 74% coverage. The average of the K1 lemmas among the Taiwanese textbooks was 35,154 words (75% coverage). Conversely, the Japanese textbooks contain 20,297 K1 word tokens (70% coverage) in total. The average of the K1 lemmas among the Japanese textbooks was 6,765 words (70%

coverage). This finding means that the Taiwanese textbooks use the K1 level of the high-frequency lemmas more than the Japanese textbooks. In other words, it is possible to say that the Taiwanese textbooks recycle the K1 lemmas more than the Japanese textbooks.

Moreover, the “Others” column shows an important difference which affected the word token coverages between the Japanese and Taiwanese textbooks. The lemmas in the “Others” column are the words outside the BNC-6318 wordlist. Thus, they should be mid-frequency or low-frequency lemmas. The “Others” column of the Taiwanese textbooks showed 7% to 9% coverages in each textbook. On the other hand, the Japanese textbooks showed 12% to 13% coverages in each textbook. That is, Table 4.5 indicates that the Japanese textbooks tend to use mid-frequency or low-frequency lemmas more than the Taiwanese textbooks.

Both textbook series have similar coverage by the K2 to the last 318 words (K7). The differences in coverage between the Japanese and Taiwanese textbooks among K2 to K7 levels ranged only from 0.5% to 1%. Therefore, both textbook series use the lemmas in these BNC frequency levels to almost the same extent. That is, it can be concluded that the Japanese textbooks can use

the “K2” to the last 318 levels words without a significant difference from the Taiwanese textbooks.

From the analysis of Table 4.5, it is possible to say that the K1 and “Other” column are the causes to make differences in the proportions of high-frequency lemmas between the Japanese and Taiwanese textbooks. However, the K1 level lemma use can be easily increased if the reading texts in the Japanese textbooks become as longer as that in the Taiwanese textbooks. Because the K1 level words are the most frequent words in English, when the reading texts in the Japanese textbooks become longer, the percentage of the K1 words is more likely to be increased and as a result, the percentages of the words in “Others” column will be down-blended. In other words, the small amount of the word token in the Japanese textbooks can limit the use or recycling of the K1 words. From these findings, it can be concluded that the less high-frequency lemmas coverage in the Japanese textbooks might be caused by both the use of mid-frequency or low-frequency words and the small amount of the word token, which can limit the reuse of K1 words.

In terms of the vocabulary size students need to have to achieve 95-98% textbook coverage, Table 4.5 reveals that both Japanese and Taiwanese students

are required to know lemmas more than the most frequent 6,318 lemmas. The largest lexical coverage in the Taiwanese textbooks with the most frequent 6,318 lemmas was 94% of TT2. Regarding to other Taiwanese textbooks, the lexical coverages were over 90%, but did not reach at 95-98% of word tokens only with lemmas in the BNC-6318. In the Japanese textbooks, on the other hand, the lexical coverages were not beyond 90% of word tokens. That is, the required levels of vocabulary to cover 95-98% of word tokens in the Japanese and Taiwanese textbooks are over the most frequent 6,318 lemmas as well. This appears to be heavy for both Taiwanese and Japanese students. This finding is corroborated by previous research on the vocabulary load of EFL textbooks (e.g., Sun and Dang, 2020; Le and Dinh, 2022). According to previous research, textbooks tend to require students to have a large amount of vocabulary to achieve the thresholds of vocabulary coverage. In other words, it is likely to be challenging for students to have sufficient vocabulary knowledge to achieve 95-98% of vocabulary coverages in their textbooks because textbooks use vocabulary, which students have not learnt yet. As Nakayama (2022) suggests, textbooks are not reading materials, but teaching materials. Therefore, textbooks need to provide students with much vocabulary they do not know so that they can

develop their vocabulary knowledge by encountering unknown words in the textbooks. If so, the more important consideration may be how many high-frequency lemmas or other rarer words students can learn from textbooks. This kind of research will be needed in future.

4.2.4 The differences in the number of high-frequency lemmas between the Japanese and Taiwanese textbooks

This study revealed that there were some essential differences in the lexical variety students can encounter through their textbooks. Tables 4.3 and 4.4 show that the Taiwanese textbooks have approximately three times more lemmas than the textbooks the Japanese textbooks. Moreover, in terms of the variety of the high-frequency lemmas students can encounter in the Japanese and Taiwanese textbooks, this study found that some measurable differences existed. That is, the Taiwanese students can encounter most high-frequency lemmas (over 90% of the new-GSL-2500 and BNC-3000), while the Japanese students can encounter about 50% of the high-frequency lemmas through their textbooks. This 50% of high-frequency lemmas included in the Japanese textbooks is lower than the percentage found in other textbooks. For example, Eldridge and Neufeld

(2009) report that only 70% of the most frequent 2000 words (1400 out of the most frequent 2000 words) are shown in the *Success* coursebook series (Longman). Moreover, O'Loughlin (2012) concludes that three distinct levels of EFL textbooks from Oxford University Press's *New English File* present approximately 70% of high-frequency lemmas (1435 of the 2000 most frequent word families). Seeing the high-frequency word percentages of the Japanese textbooks in the most frequent 2000 words, they present only around 60% of the essential words. Therefore, much more various high-frequency lemmas were presented for the Taiwanese students, but not for the Japanese students. Moreover, this finding means that the Japanese students might be required to learn the rest of the high-frequency lemmas by themselves and from other language sources. As Nation (2006) states, high-frequency words could cover almost 80% of authentic English texts in both written or spoken form. Additionally, learning the high-frequency lemmas can allow students to have more English experience not only from their textbooks, but also other sources such as some Graded readers (e.g., Schmitt, 2000). Therefore, it should be important for student to have the opportunities to learn many high-frequency lemmas as much as possible through their textbooks. Thinking about the importance of the high-

frequency lemmas in English learning, the difference in the variety of the high-frequency lemmas between the Japanese and Taiwanese textbooks might become one of the possible causes in the difference of the English proficiency between Taiwan and Japan.

Moreover, the present study revealed that the Taiwanese textbooks could provide their students with more opportunities to encounter mid-frequency and low-frequency lemmas in the Japanese textbooks. According to the data presented in Tables 4.3 and 4.4, the capacity of Taiwanese textbooks to incorporate mid- or low-frequency lemmas into their content is approximately 70%. That is, out of the total 9039 lemmas in the Taiwanese textbooks, high-frequency lemmas only account for 2395 lemmas (26%) in the new-GSL-2500 and 2753 lemmas (30%) in the BNC-3000. Therefore, the Taiwanese textbooks can show approximately 7000 mid-or low-frequency lemmas to their students although many proper nouns are also included in this figure. By contrast, Japanese textbooks have a capacity of around 50% for such lemmas. Nation (2006) claims that learners need to know 98% of words to fully understand authentic texts in English without any assistance and the number of words for this became 8000-9000 word-family. It means, the high-frequency lemmas are not enough for

students to reach the ideal lexical coverage. Therefore, Schmitt and Schmitt (2014) suggest that knowing mid-frequency vocabulary, that is, words between 3000 and 9000 levels in frequency-based wordlists, is necessary to achieve 98% of coverage in authentic English use such as reading English novels or newspapers and watching movies. Consequently, the Taiwanese textbooks could provide students with more opportunities to encounter and learn more mid-frequency or low-frequency words than the Japanese textbooks.

In sum, the Taiwanese textbooks provide their students with more high-frequency lemmas. Moreover, since mid- or low-frequency words in textbooks can be also helpful for their students to potentially develop their vocabulary knowledge, students who are taught English through the Taiwanese textbooks benefit from being exposed to a wider variety of not only high-frequency but also mid- and low-frequency lemmas.

4.3 The second set of research questions

The second set of research questions aims at identifying the differences in the word frequency levels of newly-introduced lemmas and the use of them between the Japanese and Taiwanese textbooks. Milton (2009) claims that an

effective textbook introduces many new words to students because the more vocabulary is presented in a textbook, the more vocabulary students can potentially learn through the textbook. Moreover, research suggests that 95-98% of text coverage is required for learners to sufficiently understand the meaning of authentic English texts (e.g., Hirsh and Nation, 1992; Laufer, 1989; Nation, 2006; Schmitt et al., 2011). Knowing high-frequency words is essential to achieve this level of text coverage, but it is not enough. Schmitt and Schmitt (2014) suggest that knowing mid-frequency (3000-9000 lemmas in frequency-based wordlists) is also necessary, but they also claim that these words are unlikely to be paid enough attention in English teaching and learning. Therefore, it is significant to examine which word frequency levels of lemmas the Japanese and Taiwanese textbooks newly introduce to their students and how many newly-introduced lemmas are contained.

Moreover, vocabulary repetition can be one of the most important factors for students to develop their vocabulary knowledge (Webb and Nation, 2017). Webb (2007) suggests that 10 or more encounters with target words are needed for learners to show both receptive and productive knowledge of the words. Uchihara, Webb and Yanagisawa (2019) report that the effect of vocabulary

repetitions is something unignorable. The exact number of repetitions needed for successful vocabulary learning has remained unclear, but the common suggestion about vocabulary repetition is that the more learners encounter the target words, the more likely they can be to learn them. Therefore, it seems to be essential for EFL textbooks to provide students with many opportunities to repeatedly encounter the target lemmas so that their students can be more likely to learn them. Hence, the factor of vocabulary repetition is one of the most significant aspects at the vocabulary research on textbooks (Sun and Dang, 2020). The present study uses five repetition bands (“once”, “2 to 3”, “4 to 6”, “7 to 9” and “10 and more” occurrences) to find out how many times the newly-introduced lemmas appear in the Japanese and Taiwanese textbooks and to identify any differences between the two textbook series in terms of vocabulary repetition.

The analysis of the distances between vocabulary repetitions is an additional crucial element in the process of vocabulary learning, alongside the repetition itself. The learning and retention of newly-introduced lemmas seem to remain challenging for students, even when they encounter the target lemmas numerous times only in a single unit of a textbook. Vocabulary knowledge

students cannot re-encounter for a long time easily diminishes (Webb and Nation, 2017). Thus, textbooks should provide their students with opportunities to reencounter the target lemmas in other units of the textbook. In order to examine whether the Japanese and Taiwanese textbooks offer their students such vocabulary learning opportunities across units of each textbook, the present study examines how many of newly-introduced lemmas are reused throughout units of each textbook with three types of bands (“one unit”, “two units” and “three units and more”).

Finally, the present study counts the number of newly-introduced lemmas which are reused more than 10 times and appear in more than three units of the textbook. Webb (2007) suggests that 10 or more encounters with a word are needed to gain both receptive and productive knowledge incidentally. Additionally, the “more than three units” means that the lemmas in this band can appear in more than one-third or one-fourth of units in each textbook. Therefore, it is possible that students can have relatively many opportunities to re-encounter the lemmas in many of units in a textbook. By counting the newly-introduced lemmas which meet these two criteria, the present study attempts to estimate the number of the target lemmas students can theoretically learn through their textbooks.

Moreover, by comparing the results of the Japanese and Taiwanese textbooks, it is possible to elucidate which textbook series contributes more effectively to the expansion of students' vocabulary.

Thus, the present study examines the frequency characteristics of lemmas newly introduced in each unit of the Japanese and Taiwanese textbook series, the vocabulary repetition and its distance and the number of lemmas students can theoretically learn from their textbooks. The research questions in the second set of the present study are addressed following:

3. Which frequency bands in the BNC-6318 do newly-introduced lemmas in the Japanese and Taiwanese textbooks come from?
4. How many times do the newly-introduced lemmas occur in the Japanese and Taiwanese textbooks?
5. How many newly-introduced lemmas recur across units of the Japanese and Taiwanese textbooks?
6. How many of the newly-introduced lemmas are, in theory, learnable through the Japanese and Taiwanese textbooks?

4.4 Research question 3: Which frequency bands in the BNC-6318 do newly-introduced lemmas in the Japanese and Taiwanese textbooks come from?

To investigate the number and the L1-corpus frequency of the newly-introduced lemmas in the textbooks, this study used the lemmatized BNC frequency wordlist BNC-6318 (Kilgarriff, 2006). The wordlist contains not only high-frequency lemmas, but also some mid-frequency lemmas. Therefore, it is possible to observe how each textbook deal with mid-frequency lemmas, which are essential for students to achieve 95-98% of vocabulary coverages. This wordlist contains 6318 lemmas in a frequency order. This study separates the 6318 words into seven bands, from K1 to the last 318 lemmas (K7). The results are shown in Table 4.6.

Table 4.6: The frequency levels of the newly-introduced lemmas of the Japanese and Taiwanese textbooks in the BNC-6318; the percentages show how much each level band accounts for the total of the newly-introduced lemmas in each textbook; the number of the newly introduced lemmas in each band appears within parentheses

	New words	BNC K1	K2	K3	K4	K5	K6	K7	MISSING
TT1	491	23%(112)	21%(105)	17%(84)	12%(59)	9%(45)	5%(23)	1%(5)	12%(58)
TT2	498	11%(53)	20%(101)	18%(88)	12%(58)	10%(50)	9%(44)	2%(12)	18%(92)
TT3	511	6%(30)	15%(77)	17%(85)	12%(60)	11%(56)	9%(46)	2%(9)	29%(148)
TT4	505	5%(23)	9%(43)	14%(73)	12%(60)	10%(49)	10%(51)	3%(16)	38%(190)
TT5	409	2%(8)	6%(25)	11%(47)	14%(56)	12%(49)	10%(42)	4%(17)	40%(165)
TT6	376	1%(3)	5%(18)	10%(36)	10%(39)	13%(49)	12%(44)	4%(14)	46%(173)
TOTAL	2790	8%(229)	13%(369)	15%(413)	12%(332)	11%(298)	9%(250)	3%(73)	30%(826)
JT1	268	19%(52)	19%(50)	15%(41)	9%(25)	7%(20)	6%(17)	1%(4)	22%(59)
JT2	566	10%(59)	12%(69)	13%(74)	10%(55)	10%(58)	6%(33)	2%(12)	36%(206)
JT3	462	16%(76)	18%(84)	13%(58)	10%(47)	8%(35)	5%(24)	0%(0)	30%(138)
TOTAL	1296	14%(187)	16%(203)	13%(173)	10%(127)	9%(113)	6%(74)	1%(16)	31%(403)

Table 4.6 reports that the Taiwanese textbooks introduced 2,790 new lemmas through the three years of senior high school, while the Japanese textbooks introduce 1,296 lemmas. In other words, the newly-introduced lemmas in the Japanese textbooks are less than half of those in the Taiwanese textbooks. Moreover, Table 4.6 also indicates a significant difference in the frequency levels of the newly-introduced lemmas in BNC-6318 wordlist between the Japanese and Taiwanese textbooks. In TT1 and TT2, which are used during the first grade in senior high school, the Taiwanese textbooks introduced relatively high percentages of the high-frequency lemmas in the BNC-6318 wordlist. At the first

grade of Taiwanese senior high school, the percentages of the newly-introduced lemmas from K1 to K3 are 61% in total in TT1 and 49% in TT2. In other words, the high-frequency lemmas accounted for many of the newly-introduced lemmas in the first grade of the Taiwanese textbooks. However, from the second grade, the percentages start becoming lower in a regular pattern and at the third grade, the percentage is 19% in TT5 and 16% in TT6. That is, Table 4.6 revealed that the percentages of the high-frequency lemmas as the newly-introduced lemmas in the Taiwanese textbooks tended to decrease along with the grades of the textbooks. By contrast, the percentages of the newly-introduced lemmas in mid- or low-frequency levels (from K4 to the Missing column) gradually increased along with the grades of the textbooks. In accordance with the definition by Schmitt and Schmitt (2014), the lemmas in the "Missing" column of Table 4.6 are mid-frequency or low-frequency lemmas, as they fall outside the range of the 6318 lemmas in the frequency order. Thus, Table 4.6 indicates that the Taiwanese textbooks shift newly-introduced lemmas from the high-frequency levels to mid- or low-frequency ones during the three years of senior high school.

Conversely, it was difficult to find any pattern in the frequency level of the newly-introduced lemmas in the Japanese textbooks. Unlike the Taiwanese

textbooks, no regular increase or decrease of high-frequency lemmas was observed. However, by comparing the total percentages of high-frequency lemmas (K1 to K3) and mid- or low-frequency lemmas (K4 to “Missing”) in the Japanese textbooks, it points out that two textbooks out of the three Japanese textbooks introduce mid- or low-frequency lemmas more than high frequency lemmas. The percentage of newly-introduced lemmas in the JT1 is 53% for high-frequency levels and 47% for mid- or low-frequency levels. That is, at the first grade of Japanese senior high school, high-frequency lemmas receive relatively more attention than other frequency levels. However, in terms of the other two textbooks (JT2 and 3), the attention is reserved. Table 4.6 shows high-frequency lemmas account for 37% in the JT2 and 47% in the JT3 in total, whereas mid- or low-frequency lemmas have a higher proportion of newly-introduced lemmas (63% in the JT2 and 53% in the JT3). It might be true that frequent words and infrequent words should be introduced in roughly equal amount (Milton, 2009), but Nakayama (2022) claims that Japanese students tend not to be exposed to high-frequency words because their textbooks do not always provide them with such opportunities. In the case of the Taiwanese textbook, since there is a regular pattern and possible rationale for the decrease of high-frequency lemmas, it is

understandable that the Taiwanese textbooks require their students to learn less frequent vocabulary for their vocabulary development. However, in the case of the Japanese textbooks, the selection of newly-introduced lemmas does not seem to be based on information about the frequency of words in large L1 corpora such as the BNC.

4.4.1 The BNC frequency level of newly-introduced lemmas in the Japanese and Taiwanese textbooks

Table 4.6 shows that in the Taiwanese textbooks the BNC frequency of newly-introduced lemmas follows a pattern across textbook-series levels, while the Japanese textbooks do not. At the early grades of high school in Taiwan, high-frequency lemmas accounted for most of the newly-introduced lemmas in the textbooks; the number of the high-frequency lemmas in the newly-introduced lemmas decreased and the mid-frequency lemmas increased along with the grade in Taiwanese senior high school. In other words, the Taiwanese textbooks shifted their target vocabulary level from high-frequency to mid-frequency during the three years of senior high school. By contrast, neither the number of the high-frequency lemmas nor that of the mid-frequency lemmas among the newly-

introduced lemmas in the Japanese textbooks shows any increasing or decreasing pattern during the three years of high school. Based on this finding, it can be concluded that the Japanese textbook creators did not select the newly-introduced lemmas based on their frequency level in L1 English large corpora. By contrast, it is possible that the Taiwanese textbook creators selected the target words based on their L1 frequency because of the shift from the high-frequency to the mid-frequency levels. This shift in the Taiwanese textbooks might indicate that the Taiwanese students encounter most of the high-frequency lemmas by the third grade of high school and then move to learning mid-frequency lemmas through the textbooks. In fact, Tables 4.3 and 4.4 in the first set of research questions revealed that the Taiwanese textbooks offer opportunities to encounter most of the high-frequency lemmas in the new-GSL (Brezina and Gablasoba, 2015) and BNC-3000 (Kilgarriff, 2006). Consequently, the Taiwanese students are given the opportunity to learn both high-frequency and mid- or low-frequency lemmas through their textbooks more than the Japanese students.

Regarding the selection of the newly-introduced lemmas in the Japanese textbook series, according to the Japan Association of College English Teachers (JACET), the MEXT (Ministry of Education, Culture, Sport, Science and

Technology) did not provide any reference wordlist for textbooks publishers in Japan (JACET, 2016). That is, the selection of the newly-introduced lemmas depends on each textbook publisher in Japan. Moreover, the lack of a common wordlist in the Japanese school curriculum guideline causes some differences in the variation of the vocabulary students can learn through the textbooks. In other words, the vocabulary students can learn can vary widely depending on the textbook they use (JACET, 2016). In fact, Aizawa (2010, cited in JACET, 2016) suggests that the number of the new words shared among six popular Japanese English textbook series is small. This can cause very wide differences in students' English vocabulary level in Japan. Teaching every vocabulary in the textbooks seems to be impossible, but there is a widely spread agreement that the high-frequency lemmas should be taught in a deliberate way, especially for beginners or those who are not good at English (e.g., Schmitt, 2000; Webb and Nation, 2017). The textbook publishers should know which frequency level words the target students need to learn in the textbooks.

While the selection of vocabulary in the Japanese textbooks largely depends on each textbook publishers, a wordlist called "Reference Word List" (RWL), can be used as a guideline for the selection of vocabulary in Taiwanese

textbooks. There seem to be three selection criteria: the frequency, American and Chinese cultural background, rules for word formation and the life experience of Taiwanese senor-high school students by the college Entrance Examination Center (hence, CEEC). The RWL was created in order to facilitate senior high school students' English learning and help their preparation for the high-stakes college entrance examination (Reynolds, Shih and Wu, 2018). The RWL seems to be considered as a tailor-made learning tool. Among six levels of the RWL, the first four levels contain 4,320 words, which are used for the Scholastic Aptitude English Test (SAET)[3] development/preparation and 6,480 words are contained in the 6-level RWL to be used for the Department Required English Test (DRET) development/preparation in Taiwan (Reynolds, Shih and Wu, 2018). Therefore, the reason why the present study found a particular pattern in the newly-introduced lemmas of the Taiwanese textbooks in the word frequency levels could be because the RWL include the frequency as one of the criteria in the selection of newly-introduced lemmas.

[3] The SAET and DRET are the tests Taiwanese senior high school graduates have to take to enter university (the College Entrance Examination Center, 2014).

4.4.2 The differences in the number of newly-introduced lemmas between the Japanese and Taiwanese textbooks

Research question 3 examined the differences in the frequency levels of newly-introduced lemmas between the Japanese and Taiwanese textbooks. Additionally, as a results on the data process, the present study also identified the difference in the number of the newly-introduced lemmas in the Japanese and Taiwanese textbooks. Table 4.6 indicates that the Taiwanese textbooks seem to shift the target frequency level from the high-frequency to mid- or low-frequency levels during the high school years, while the Japanese textbooks do not show any particular pattern in terms of the L1-corpus frequency of newly-introduced lemmas. Moreover, it also becomes clear that the Taiwanese textbooks include more newly-introduced lemmas than the Japanese textbooks. At this point, the Taiwanese textbooks seem to be more helpful for students' vocabulary development than the Japanese textbooks.

Textbooks should introduce their newly-introduced lemmas as much as possible so that the students can have many opportunities to deliberately learn the target lemmas in the class or by themselves with the textbooks. Milton (2009) suggests that effective textbooks seemed to provide newly-introduced lemmas in

very large quantities. In terms of the vocabulary uptake from the classroom input, he claims that the more vocabulary textbooks provide, the more vocabulary learners seem to learn. Therefore, the Taiwanese textbooks seemed to be superior to the Japanese textbooks regarding to the total number of the newly-introduced lemmas students potentially learnt.

As Table 4.6 showed, while the Taiwanese textbooks provided 2,790 new lemmas in total during the three years of high school, the Japanese textbooks have 1,296 lemmas in total during the same period. Without textbooks, it might be possible for some students to use other sources such as Graded readers or English newspapers to increase their input. However, these additional sources of English language input are not available to all students because of differences in students' economic circumstances or their motivation toward English learning. Consequently, textbook play an important role as sources of English language input since every student has at least one.

However, even though the textbooks offer a lot of newly-introduced lemmas to students, it seems to be difficult for them to learn all of the target lemmas in the textbooks. According to Vassiliu (2001), even the best learners could not master every new word presented in the course. The study reports that

the highest uptake from the textbooks was 77% and on the average 55%. Moreover, as a similar study, Alsaif and Milton (2012) also reports that the vocabulary uptake from the textbooks in Saudi Arabia became around 40% on average. These studies indicates that it seems to be very difficult for students to memorize every newly-introduced lemma in their textbooks. Assuming these findings on vocabulary update from textbooks are applicable to the setting of the present study, the estimated vocabulary uptake would roughly be around 1395 words from the Taiwanese textbooks and 648 words from the Japanese textbooks. This estimation suggested that the Taiwanese students have more opportunities to potentially learn new lemmas through the textbooks than the Japanese textbooks, even though they cannot memorize all of the target words. In other words, to catch up with the Taiwanese textbooks, this study suggests that the Japanese textbooks need to increase the number of their newly-introduced lemmas.

4.5 Research question 4: How many times can students encounter the newly-introduced lemmas in the Japanese and Taiwanese textbooks?

Research question 4 in the second set examines the repetition of the newly-introduced lemmas in the Japanese and Taiwanese textbooks. The findings of RQ3 indicated that the selection of newly-introduced lemmas in the Taiwanese textbooks appeared to be based on vocabulary frequency more than that in the Japanese textbooks. In addition, the Taiwanese textbooks had more newly-introduced lemmas than the Japanese textbooks. However, students are unlikely to learn new lemmas unless they have sufficient encounters with them. Webb and Nation (2017) claim that vocabulary repetition can be a significant factor for students to learn the target lemmas. Moreover, various empirical studies indicate that the more students encounter the target lemmas, the more likely they are to learn them (e.g., Teng, 2016; Waring and Takaki, 2003; Webb, 2007). Therefore, when this study attempts to estimate how helpful textbooks can be for their students for the vocabulary development, it seems to be important to examine how many times students can encounter target lemmas in each textbook.

Webb (2007) reports that Japanese English learners can acquire

vocabulary through incidental learning, but to show sizable learning gains for both receptive and productive vocabulary knowledge, they need to encounter target lemmas at least 10 times. This study also indicates that learners can acquire some vocabulary knowledge through a few encounters. Although there are studies which indicate that more than 10 encounters are needed for learners to learn target lemmas, but because textbooks are a teaching material, the amount of texts in textbooks is limited and it seems to be difficult to present many of target lemmas more than 10 times. Thus, examining differences in vocabulary repetitions within 10 times between the Japanese and Taiwanese textbooks can be more meaningful than examining the number of newly-introduced lemmas presented more than 10 times in the textbooks. Thus, based on the relationships between the learning gain and the repetition rate by Webb (2007), this study made five repetition rate bands, “once”, “2 to 3” times, “4 to 6” times, “7 to 9” times and “10 times and more”.

Figures 4.1 and 4.2 below show how many of newly-introduced lemmas are presented in each repetition band and its percentages. Figure 4.1 illustrates the frequency of occurrence of the newly-introduced lemmas in the Taiwanese textbooks. The once band is eliminated because the newly-introduced lemmas

are presented in Taiwanese textbooks at least twice. Figure 4.2 shows the frequency of occurrence of the newly-introduced lemmas in the Japanese textbooks. The comparison of Figures 4.1 and 4.2 indicates differences in the number of encounters with target lemmas that each textbook series offers to students.

Figure 4.1: The number and percentage of newly-introduced lemmas occurring “2 to 3”, “4 to 6”, “7 to 9” times and “10 and more” times in the Taiwanese textbooks.

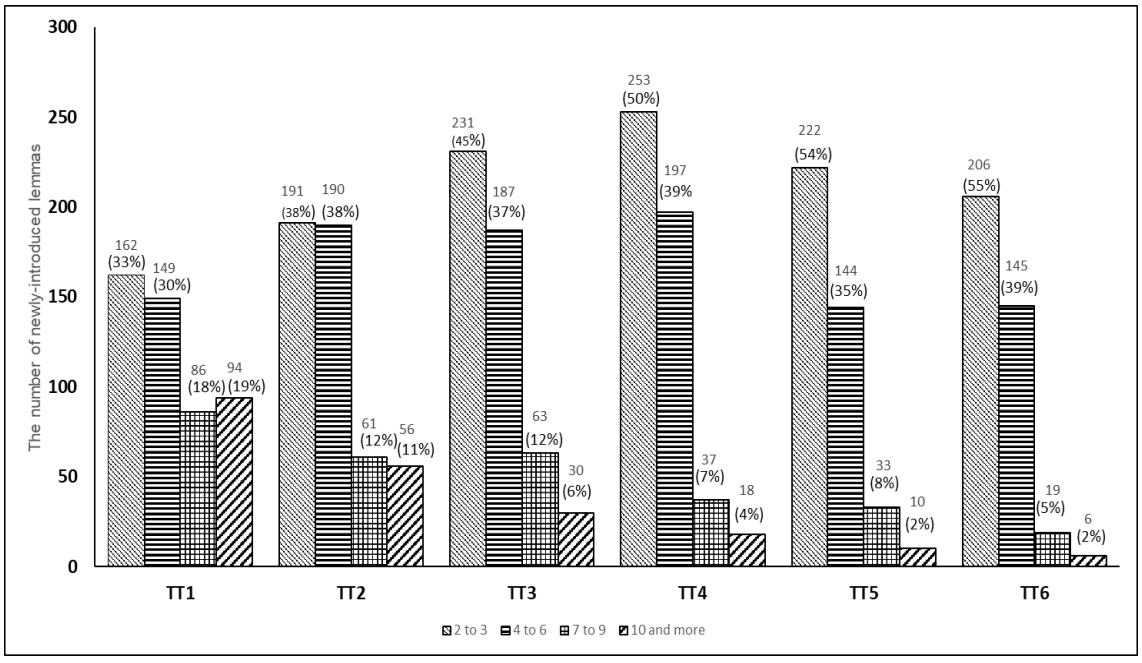
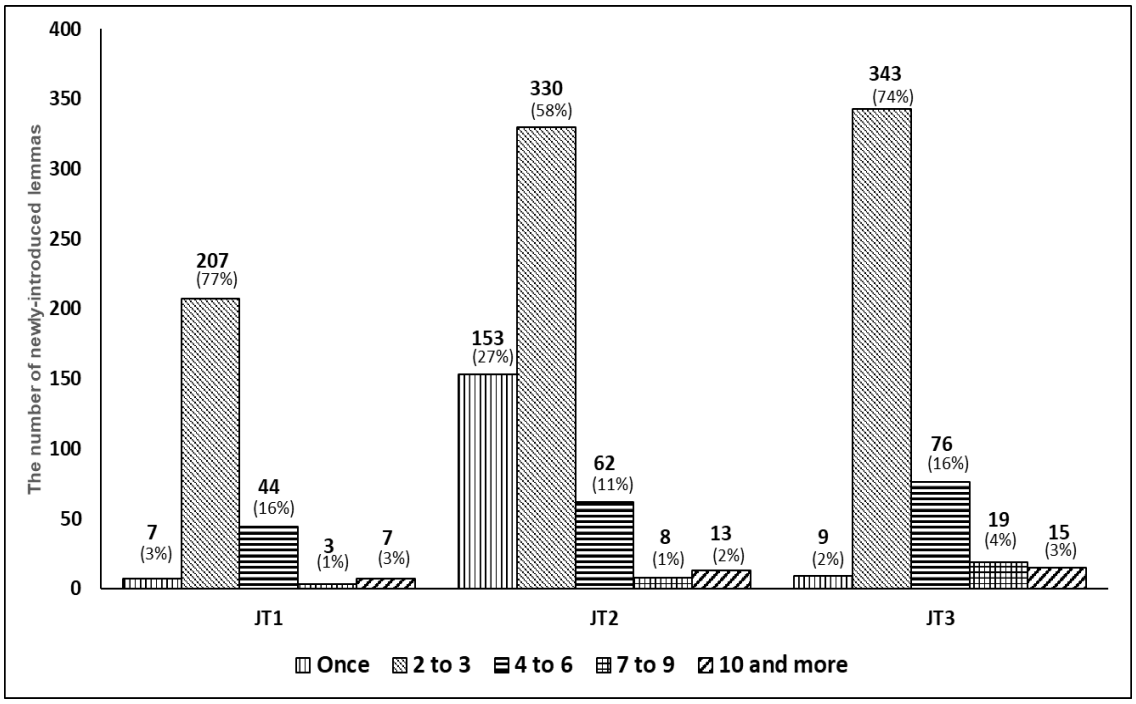


Figure 4.1 indicates that across all the Taiwanese textbooks the repetition band which contains most of newly-introduced lemmas is the “2 to 3” band, followed by the “4 to 6” times band. In addition, in terms of “7 to 9” and “10 and more”

bands, the numbers of the new lemmas in the two bands decreased along with the grades of senior high school. For example, TT1 includes 18% of the newly-introduced lemmas (86 words) in the “7 to 9” band and 19% (94 words) “10 and more” band, these numbers became much smaller in TT6: 5% (19 words) in the “7 to 9” and 2% (6 words) in “10 and more” band. Moreover, Figure 4.1 also shows that the Taiwanese textbooks provided the newly-introduced lemmas without large differences in the “2 to 3” and ”4 to 6” bands during the first grade of high school (TT1 and TT2), but the differences in the two bands also tend to be larger from the second grade to the third grade. Besides, the number of the newly-introduced lemmas in the “10 and more” band, which are the words students can potentially learn sizable receptive and productive vocabulary knowledge, is 214 words.

Figure 3.2: The number and percentage of newly-introduced lemmas occurring once, 2–3, 4-6, 7-9 times and 10 or more times in the Japanese textbooks



The most obvious difference between Figures 4.1 and 4.2 is the existence of the "once" band in Figure 4.2. While the Taiwanese textbooks introduced the new lemmas at least twice, the Japanese textbooks include some words in the "once" band and JT2 shows many words in this band. Moreover, Figure 4.2 shows that the "2 to 3" band is the highest band and there are large gaps between the "2 to 3" band and other repetition bands. Figure 4.1 showed that the "2 to 3" band was the highest and the "4 to 6" band was the second highest, but the gaps between the two bands in the Taiwanese textbooks were not as large as the ones in the

Japanese textbook.

4.5.1 The “once” band in the Japanese textbooks

One of the most obvious differences between the Japanese and Taiwanese textbooks is the existence of the “once” bands. While the Taiwanese textbooks present their newly-introduced lemmas at least twice, the Japanese textbooks, especially the JT2, include many new lemmas in the “once” band. Since this study included all English words used in the textbooks, if the textbooks use the newly-introduced lemmas in the main reading text in each unit, the words should be counted at least twice or more in the result because it appears in the reading text and introduction of new vocabulary section of each unit. Thus, having the “once” band means that the Japanese textbooks just show some newly introduce lemmas outside the reading texts. Then, it might be meaningful to examine how the kind of words are introduced in the Japanese textbooks so that the present study can check whether the way seems to be appropriate or problematic.

The present study found that the Japanese textbooks used some pictures to introduce some newly-introduced lemmas which appeared only once. Figures

4.3 and 4.4 below shows some examples of how the Japanese textbooks show the newly-introduced lemmas belonging to the “once” band. It shows that the lemmas in the “once” band are introduced with a picture relating with the target lemmas. According to Milton (2009) and Vassiliu (2001), even though some words are not used a lot in the textbooks, students can learn them if the words are truly imageable. Nouns such as some job names might seem to be easier to make a connection between the meaning and the picture. However, it seems to be doubtful whether the same thing can be done for some words for feelings in the second example.

Figure 4.3: Example of how the words in the “once“ band were introduced in the Japanese textbooks for nouns (e.g., JT2, p.80)

Figure 4.4: Example of how “once“ words were introduced in the Japanese textbooks for adjectives (e.g., JT2, p.22)

Figures 4.3 and 4.4 are activities in the Japanese textbooks and other unit and students are required to investigate some words they do not know with a dictionary, check the accent and make one sentence using the target lemmas. In the first activity, students are required to say *"I want to be _____ in the future."* and *"I am/feel _____."* in the second one. It seems to be very questionable how much this kind of activity can help students learn the target lemmas and whether it can be defined as a communication activity. Therefore, this study suggests that this way of introducing the newly-introduced lemmas should be examined very carefully for the effect on students' learning gains.

4.5.2 The differences in the "2 to 3" and "4 to 6" bands between the Japanese and Taiwanese textbooks

This study found that the "2 to 3" and "4 to 6" bands are the highest bands in the Japanese and Taiwanese textbooks. However, this study concluded that the Taiwanese textbooks tend to provide students with opportunities to encounter the target lemmas 2 to 6 times, but the opportunities the Japanese students can gain through their textbooks are only "2 to 3" encounters. This is because while the Taiwanese textbooks present their newly-introduced lemmas in the "2 to 3"

and “4 to 6” bands with only small gaps (7.8% gaps on average), the Japanese textbooks have large gaps between the two bands (55% gaps on average). That is, the Japanese students can encounter most of the target lemmas only two or three times through their textbooks during the senior high school years. Thus, it might be possible that the Japanese textbooks limit their students’ opportunity to learn the target lemmas more than the Taiwanese textbooks. To maximize the effect of a textbook, it will be necessary to think how the Japanese textbooks can provide more opportunities to learn the newly-introduced lemmas for their students.

In terms of the large gap between the “2 to 3” and ““4 to 6”” bands in the Japanese textbooks, this study hypothesizes that one of the possible reasons is the way target lemmas are introduced in the Japanese and Taiwanese textbooks. Figures 4.5 and 4.6 are the examples of how the Japanese and Taiwanese textbooks present their newly-introduced lemmas in each unit of textbooks.

Figure 4.5: Example of the introduction of new words in the Taiwanese textbooks[4]

1. **nervous** [ν3ϖ1σ] adj. *showing or having feelings of being worried, excited, afraid, or troubled* 緊張的，擔憂的
   - Stella was so nervous about the final exam that she didn't have a good sleep the night before the test.
2. **pound** [πα5νδ] vi. (*of hearts*) *to beat hard, loudly, and fast* (心臟)狂跳，怦怦跳
   - Mr. Green's heart was pounding as he waited for the result of the job interview.

   vi. *to hit someone or something hard again and again, especially in a way that makes a lot of noise* 敲打，重擊
   - When Hank found that his new motorcycle had been stolen, he pounded on the table in anger.
3. **communicate** [κ1 μ̄φυν1&κετ] vi. *to share feelings and ideas with someone in order to have a good relationship* 溝通
   - It is difficult to communicate with Carlos since he can't speak either English or Chinese.

   vi. *to give information to someone by speaking or writing* 交流，聯繫
   - Josh usually communicates with his cousins by e-mail once a week.

   communication [κ1&μφυν1 κ̄εΣ1ν] n. [U] 溝通
   - Good communication between parents and children helps make a happy family.
4. **hall** [ηΟλ] n. [C] *a passage in a building that leads to rooms* (建物內的)走廊
   - The hostess welcomed the guests and led them down the hall to the dining room.
5. **medal** [μ̄ΕδΛ] n. [C] *something that is shaped like a large coin and is given to someone who wins a game or does something brave* 獎牌，獎章
   - The firefighter was given a medal for saving ten children from the burning school bus.
6. **costume** [κ̄Αστφυμ] n. [C] *clothes that someone wears in a play, movie, etc.* 戲服
   - On Halloween, kids usually go trick-or-treating in costumes from door to door.
7. **role** [ρολ] n. [C] *the part of an actor or an actress in a movie, play, etc.* (演員的)角色
   - Bob worked very hard to get leading roles in many famous plays.

   n. [C] *the job or purpose that someone or something has in a situation or activity* 職責，作用
   - Parents play an important role in teaching their children.
8. **drama** [δ̄ρΑμ1] n. [C] *a play for the TV, theater, etc.* 戲劇
   - My favorite actress is going to star in a new TV drama. I believe it will become a big hit.

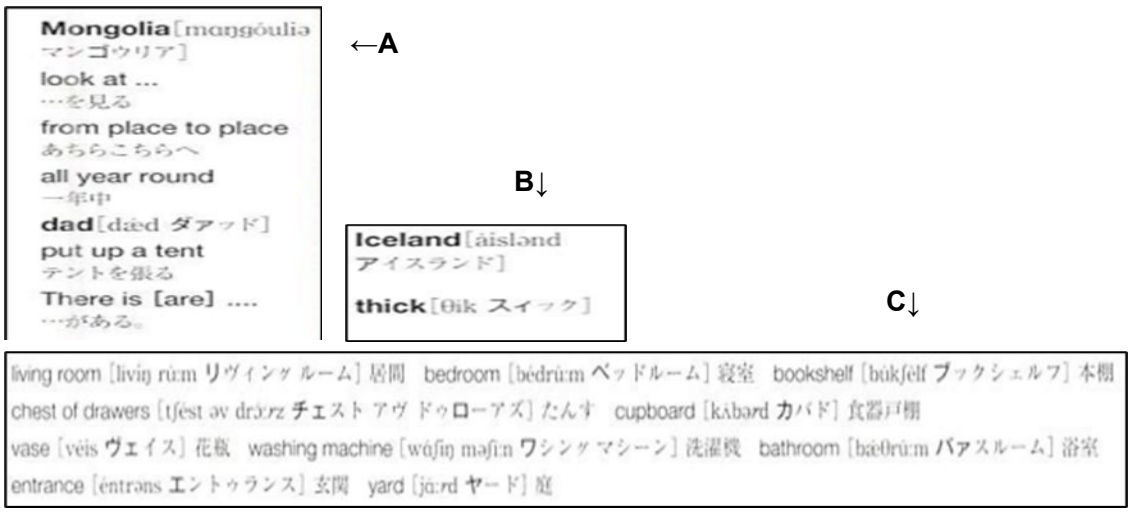
   n. [C] *an exciting event or situation* 戲劇性的事件
   - A big drama happened when I was making breakfast this morning. The pan caught fire, and the kitchen was almost burned down.

   dramatic [δρ1 μ̄8τΙκ] adj. 戲劇性的，誇張的
   - Come on. It's no big deal—don't be so dramatic.
9. **wonder** [ω̄⊥νδ2] vt. *to think about something and want to know more about it* 想知道，感到納悶
   - "Why didn't half of the class pass the test? Was it too difficult?" the teacher wondered.

   vi. *to be surprised at something* 對…感到驚訝
   - The audience (觀眾) wondered at the magician's tricks, especially when he made a car

[4] *The phonetic symbols were not recognized by my computer.

Figure 4.6: Examples of the introduction of new words in the Japanese textbooks



Figures 4.5 and 4.6 illustrate the introduction of new lemmas in Unit 1 in the Japanese and Taiwanese textbooks, respectively. Figure 4.5 shows that Taiwanese textbooks provided a lot of information about each new lemma such as the English definition, the part of speech, the accent, whether a noun is countable or uncountable and the L1 translation of the lemma. In addition, all the newly-introduced lemmas are introduced with an example sentence. Therefore, it is possible to say that the Taiwanese students can see how to use the newly-introduced lemmas through the examples and gain much information about target lemmas. The way of the introduction in the Taiwanese textbooks resembles an English dictionary. By contrast, Figure 4.6 shows that the Japanese textbooks provide only the accent of the newly-introduced lemmas. Some compound words

were introduced with their meaning in box A in Figure 4.6. However, the meaning of the newly-introduced lemmas is not always shown. For example, “thick” does not have the meaning in box B. More specifically, the Japanese textbooks show the meaning for only the words used in an activity in box C. However, the meaning of the newly-introduced lemmas used in reading texts are not shown in the unit. This difference in terms of example-sentence provision is one of the reasons why the Taiwanese textbooks can show the newly-introduced lemmas more times than the Japanese textbooks. Moreover, the lack of information about the target lemmas indicates that Japanese students have to check the new lemmas by themselves. Therefore, the learning gains are more likely to depend on their own effort than the textbooks provide students with sufficient information. If students have to look up each newly-introduced lemmas individually, it could increase the learning burden for students, potentially making vocabulary learning more challenging. Therefore, the present study advocate for a more thorough explanation of newly-introduced lemmas in the Japanese textbooks, aiming to ensure that students can effectively grasp the target lemmas from the contents. This might be able to increase the vocabulary repetitions students can have for the target lemmas.

4.5.3 The differences in the number of the newly-introduced lemmas in "10 and more" band between the Japanese and Taiwanese textbooks

Even though it can be difficult for textbooks to present their new target lemmas 10 or more times, examining how many new words students can encounter 10 times or more in a textbook can help us to estimate how much a textbook supports vocabulary learning. Studies on the effect of vocabulary repetition on learning have shown that the more encounters with target lemmas students have, the more likely they are to learn them (e.g., Teng, 2016; Waring and Takaki, 2003; Webb, 2007).

According to Webb (2007), after 10 encounters, learners can show sizable learning gains for both receptive and productive vocabulary knowledge of target lemmas. The findings indicate that out of the 2,790 newly-introduced lemmas, total 214 lemmas are in the "10 and more" bands in the Taiwanese textbooks. The adequacy of this proportion remains uncertain due to the limited number of studies that have examined the repetition of newly-introduced lemmas in textbooks to date. However, through a comparison with the Japanese textbooks, there seems to be a clear and important difference between the

Japanese and Taiwanese textbooks. In the case of the Japanese textbooks, out of the 1,297 newly-introduced lemmas, only 35 lemmas in total appeared in the “10 and more” band. This means that the Taiwanese textbooks provide their students with more opportunities to develop students’ vocabulary knowledge than the Japanese textbooks.

However, the number of the new lemmas in the “10 and more” bands across the Taiwanese textbooks gradually decreased along with the school grades. For example, TT1 and TT2 at the first grade contain 150 words in the “10 and more” band, but in TT3 and 4, the number of the lemmas becomes 48 words and in TT5 and 6 at the third grade show only 16 lemmas. The same patterns can be also observed in the “7 to 9” bands in the Taiwanese textbooks. However, in the Japanese textbooks, the number of the words in the “7 to 9” and “10 and more” bands slightly increased along with the school grade. This study hypothesizes that the cause for these patterns might be on the vocabulary frequency levels targeted by the textbooks. That is, for example, the research question 3 proved that the TT1 and TT2, which are used at the first grade of Taiwanese senior high school, contains many high-frequency lemmas in their target lemmas. However, the number of them gradually decreased along with the school grade and TT5

and TT6 at the last grade show only a few high-frequency lemmas. Instead of them, the number of mid- or low-frequency lemmas increased. Yet, because mid- or low-frequency lemmas are unlikely to be used as frequently as high-frequency lemmas, it might be possible that such decreasing on high repetition bands (“7 to 9” and “10 and more”) among the Taiwanese textbooks happens. In order to ascertain this hypothesis, the present study examined the frequency levels of lemmas in the “7 to 9” and “10 and more” bands in the Japanese and Taiwanese textbooks.

The total number of the newly-introduced lemmas in the “7 to 9” is 299 lemmas and “10 and more” band contains 214 lemmas in the Taiwanese textbooks. On the other hand, 30 lemmas in the “7 to 9” band and 35 lemmas in the “10 and more” band in the Japanese textbooks. The BNC-6318 (Kilgarriff, 2006) is used to check the frequency levels of newly-introduced lemmas in the target bands. Tables 4.7 and 4.8 show the breakdowns of the frequency levels of newly-introduced lemmas in the two target repetition bands in the Japanese and Taiwanese textbooks.

Table 4.7: Breakdown of frequency levels of newly-introduced lemmas in the “7 to 9” band of the Japanese and Taiwanese textbooks

	K1	K2	K3	K4	K5	K6	the last 318	Not in BNC	Total
TT1	28	24	12	6	5	4	1	6	86
TT2	14	15	13	3	5	4	1	6	61
TT3	8	18	16	3	6	4	1	7	63
TT4	7	8	9	3	1	1	0	8	37
TT5	1	4	10	6	4	3	1	4	33
TT6	0	0	5	1	5	0	1	7	19
	K1	K2	K3	K4	K5	K6	the last 318	Not in BNC	Total
JT1	1	0	0	0	0	1	1	0	3
JT2	2	0	1	0	2	0	0	3	8
JT3	6	3	2	1	1	1	0	5	19

Table 4.8: Breakdown of frequency levels of newly-introduced lemmas in the “10 and more” band of the Japanese and Taiwanese textbooks

	K1	K2	K3	K4	K5	K6	the last 318	Not in BNC	**Total**
TT1	39	21	12	9	4	2	0	7	94
TT2	23	15	12	1	1	0	1	3	56
TT3	6	7	4	2	2	3	0	6	30
TT4	3	2	3	1	1	2	0	6	18
TT5	2	2	2	2	0	0	0	2	10
TT6	0	2	0	0	2	0	0	2	6
	K1	K2	K3	K4	K5	K6	the last 318	Not in BNC	**Total**
JT1	0	1	1	1	0	1	0	3	7
JT2	3	3	1	0	0	0	0	6	13
JT3	3	2	0	0	1	3	0	6	15

Tables 4.7 and 4.8 reveal that the Taiwanese textbooks use their newly-introduced lemmas more systematically than the Japanese textbooks. The Taiwanese textbooks during the first and second grades in senior high school relatively have many high-frequency lemmas in the “7 to 9” and “10 and more”

bands. However, the number of the high-frequency lemmas in both bands gradually decreased along with the grade of the textbooks. Additionally, the number of the newly-introduced lemmas in both bands are also decreased on Tables 4.7 and 4.8. These patterns in the “7 to 9” and “10 and more” bands correspond with the shift in the research question 3. As the finding in the research question 3 revealed, the frequency levels of target lemmas in the Taiwanese textbooks shifted from the high-frequency to mid- or low-frequency levels during senior high school. Consequently, since TT1 and 2 contain relatively many high-frequency lemmas in their newly-introduced lemmas, they can easily reuse the target lemmas. However, the TT5 and 6 pay more attention to mid- or low-frequency lemmas. Thus, it might become more difficult for the TT5 and 6 to present many of their new lemmas in the “7 to 9” and “10 and more” bands than the TT1 and 2. It should be natural that high-frequency lemmas are more likely to be reused in English than mid- or low-frequency lemmas.

In terms of the Japanese textbooks, there are no notable changes across all the frequency levels on Tables 4.7 and 4.8. In strict terms, the number of high-frequency lemmas (K1 to K3) slightly increase, but this change might be resulted from the increase of word tokens in JT2 and 3. Since these frequency levels

lemmas are frequently used in English, the more word tokens the textbook has, the more frequently they can appear. Additionally, unlike the Taiwanese textbooks, the Japanese textbooks do not have many high-frequency lemmas in the “7 to 9” and “10 and more” bands even though the Japanese textbooks had relatively many high-frequency lemmas as their newly-introduced lemmas shown in Table 4.6 (JT1: 143 high-frequency lemmas in total; JT2: 210 words; JT3: 218 words). Thus, it is possible to say that regardless of the level of newly-introduced lemmas, the repetition rates of the newly-introduced lemmas in the Japanese textbooks are low. In other words, the Japanese textbooks tend not to reuse their newly-introduced lemmas. Therefore, from the findings at the point of vocabulary repetition, it can be concluded that the Taiwanese textbooks can provide their students with more solid opportunities to repeatedly encounter the newly-introduced lemmas with than the Japanese textbooks.

4.6 Research question 5: How many newly-introduced lemmas are reused across units of the Japanese and Taiwanese textbooks?

The research question 5 in the second set examines the distances

between the repetitions of newly-introduced lemmas, which means that how the target lemmas are recycled throughout the units of the textbook. Even though students have some encounters with newly-introduced lemmas only in one unit of a textbook where the word is introduced for the first time, it should be difficult for students to retain the knowledge longer if they do not encounter the words again in other units. Webb and Nation (2017) claims that if learners do not revisit a target word for a long time after they learn it, the knowledge easily decayed. Therefore, in addition to the repetition of the newly-introduced lemmas in the RQ4, it can be also essential to research how the newly-introduced lemmas are recycled throughout the units of the textbook so that students can revisit newly-introduced lemmas within the textbook.

Table 4.9 shows how the textbooks recycle the newly-introduced lemmas across units of each Japanese and Taiwanese textbook. To analyse how they are reused in other units, the present study made three bands, “one unit”, “two units” and “more than three units”. The “two units” and “more than three units” columns mean that the lemmas in the two columns are recycled in different units from the one originally introduced the target lemma. For example, the words in the “one unit” column appeared only in one unit of each textbook. Because the Japanese

textbooks have 10 units and the Taiwanese textbook have 10 or 12 units in a textbook, the words which appeared in more than three units are used in at least a third part or a quarter of a textbook's units.

Table 4.9: The percentages of recycling newly-introduced lemmas across the units of each textbook; the number of the newly introduced lemmas in each band appears within parentheses

	NEW	one unit	two units	more than three units
TT1	491	45%(220)	25%(123)	30%(148)
TT2	498	56%(281)	20%(98)	24%(119)
TT3	511	67%(341)	18%(92)	15%(78)
TT4	505	73%(367)	19%(98)	8%(40)
TT5	409	77%(315)	18%(73)	5%(21)
TT6	376	81%(303)	16%(59)	4%(14)
JT1	268	83%(223)	13%(36)	3%(9)
JT2	556	87%(486)	9%(49)	4%(21)
JT3	454	84%(380)	13%(58)	4%(16)

Table 4.9 indicates that the rates of the words in the "one unit" column exhibit a progressive increase in Taiwanese textbooks, whereas the rates of the lemmas in the "two units" and "more than three units" columns demonstrate a gradual decrease. For example, in TT1 out of 491 newly-introduced lemmas, 45% (220 lemmas) belong to the "one unit" column, 25% (123 lemmas) to the "two units"

column and 30% (148 lemmas) to the “more than three units’ column. Thus, TT1 seems to recycle 55% of the newly-introduced lemmas across some units of TT1. Therefore, the TT1 can provide their students with opportunities to reencounter the target lemmas with some spaces so that students can remember the target lemmas when they reencounter in other units. However, the trends of the “two units” and “more than three units” columns from TT1 to TT6 indicate that the recycling rates in the Taiwanese textbooks gradually decreased. For example, while TT1 recycles 55% of the newly-introduced lemmas in the two bands of the textbook, TT6 recycles only 20% of them.

In terms of the Japanese textbooks, most of the newly-introduced lemmas (around 80%) appear in the “one unit” column and a few appear in the other two columns across textbook levels. Therefore, Table 4.9 indicates that the Japanese textbooks tend to use their newly-introduced lemmas in the original Unit where they are introduced for the first time, whereas the Taiwanese textbooks tend to recycle newly-introduced lemmas across textbook units.

4.6.1 The differences in the number of newly-introduced lemmas, which are reused across units of a textbook between the Japanese and Taiwanese textbooks.

Table 4.9 showed that the “one unit” columns in the Japanese and Taiwanese textbooks always account for the highest percentages. As Nation (1993) suggested, it would be difficult to systematically recycle every newly-introduced vocabulary in a textbook. However, Table 4.9 also indicates that many newly-introduced lemmas appeared in the “two units” and “more than three units” bands in the Taiwanese textbooks, especially in the textbooks used in early grades (TT1 and TT2). Moreover, the higher the grades, the lower the proportion of lemmas that appeared in more than three units. This pattern seems similar to the finding of the research question 4, where early-grade Taiwanese textbooks showed many newly-introduced lemmas in the high repetition bands (“7 to 9” and “10 and more” occurrences bands) but later-grade textbooks could not provide students with opportunities to encounter the target lemmas in the two bands. The present study identified that the main reason for this was the amount of the high-frequency lemmas in the target lemmas. The early-grade Taiwanese textbooks included many high-frequency lemmas as target lemmas, but the later-grade

textbooks did not because the target frequency level was shifted from high-frequency to mid- or low-frequency level. This change made it more difficult to present or recycle the target lemmas in the later-grade Taiwanese textbooks in the “7 to 9” or “10 and more” bands. This might be able to affect the result of research question 5 as well. Thus, a decrease in the quantity of newly-introduced lemmas in the "more than three units" bands of Taiwanese textbooks may be attributed to an increase in the number of mid- or low- frequency lemmas in the newly-introduced lemmas of the Taiwanese textbooks.

To test this hypothesis, this study examined the frequency levels of the newly-introduced lemmas in the “more than three units” column in the Japanese and Taiwanese textbooks. Through this process, it is possible to identify whether the decreases in “more than three units” band in the Taiwanese textbooks were truly caused by the change from high-frequency to mid-frequency level lemmas across grades. Table 4.10 analysed frequency levels of the newly-introduced lemmas belonging to the “more than three units” in Table 4.9.

Table 4.10: Breakdown of the BNC frequency level of the newly-introduced lemmas in the “more than three units” in both Japanese and Taiwanese textbooks.

	K1	K2	K3	K4	K5	K6	ast 318 w	Others
TT1	58	39	24	13	5	1	1	7
TT2	36	29	25	7	7	4	2	9
TT3	15	28	18	2	2	5	1	7
TT4	13	10	7	1	1	3	0	5
TT5	2	3	8	4	0	3	0	1
TT6	1	3	4	1	0	1	0	4
	K1	K2	K3	K4	K5	K6	ast 318 w	Others
JT1	2	1	2	1	0	1	0	2
JT2	7	4	2	1	2	0	1	4
JT3	4	3	2	0	1	0	0	6

Table 4.10 showed similar pattern with Table 4.7 which revealed the cause of the decreases in the high repetition bands in the research question 4. That is, the early-grade Taiwanese textbooks, which include many high-frequency lemmas as their target lemmas, could provide more newly-introduced lemmas across the units of each textbook than the latter Taiwanese textbooks. Thus, it might be possible to think that the high-frequency lemmas are more likely to be recycled among the units of the textbooks than the mid-frequency lemmas. From this finding, it became clear that the Taiwanese textbooks could systematically present their newly-introduced lemmas in the textbooks to some extent.

In terms of the Japanese textbooks, both Tables 4.9 and 4.10 did not

show any noticeable pattern in the recycling of newly-introduced lemmas. Table 4.9 showed that the Japanese textbooks had over 80% of the newly-introduced lemmas in the “one unit” band. Some appeared in “two units” or “more than three units”, but compared with the Taiwanese textbooks, these lemmas account for smaller percentages of all newly-introduced lemmas in a textbook. In the comparison of the later-grade Japanese and Taiwanese textbooks, the Japanese one contains more high-frequency lemmas in their newly-introduced lemmas in Table 4.6. While the TT6 contains only 57 lemmas in total of high-frequency levels (K1 to K3), the JT3 does 218 lemmas. Following the Taiwanese textbooks’ pattern, at least the high-frequency lemmas can be recycled more than the other levels of words in the Japanese textbooks. Nevertheless, the results of the Japanese textbooks on Table 4.10 showed that both high-frequency and mid-frequency lemmas are not recycled. Therefore, it is possible to confirm that the Japanese textbooks seem not to recycle the newly-introduced lemmas. In other words, they tend to show the target lemmas only in the unit in which the words appeared for the first time. It is not sure whether teachers do something to recall students’ memory for the newly-introduced lemmas in the class, but this study revealed that at least this textbook series does not recycle their newly-introduced lemmas

across the units of each textbook. That means, the students who use these Japanese textbooks can encounter the newly-introduced lemmas only in one unit and with limited encountering rates. Thus, this Japanese textbook series requires students or teachers to compensate for the limited word encounters it offers.

4.7 Research question 6: How many of newly-introduced lemmas are theoretically learnable through the Japanese and Taiwanese textbooks?

The last research question 6 aims at estimating the number of the newly-introduced lemmas, which students can theoretically learn from their textbooks. Webb and Nation (2017) suggest that conditions such as the repetition, retrieval, and noticing need to be met to achieve successful vocabulary learning. There are some conditions, which the present study did not focus on such as the noticing, varied encounters and use and the elaboration. Yet, the research question 4 and 5 showed the repetition rates of the newly-introduced lemmas in the Japanese and Taiwanese textbooks and how the newly-introduced lemmas were recycled across the units of each textbook. Because some lemmas, which fulfil these conditions in the target textbooks are the ones students are more likely to

encounter. In other words, students can have more opportunities to learn the lemmas through their textbooks. Hence, it might be possible to estimate how many newly-introduced lemmas students can potentially and theoretically learn from their textbooks through the vocabulary repetitions and the distances between them. Webb (2007) reports that learners can show their sizable learning gains of vocabulary after at least 10 encounters with target lemmas. In addition, it is unsure how long spaced repetition can enhance learners' vocabulary retention, but revisiting target lemmas after some time can be a powerful learning condition (e.g., Barcroft, 2007; Webb and Nation, 2017). Therefore, by examining the number of the target lemmas which meet these conditions, the present study attempts to estimate the number of newly-introduced lemmas students can potentially and theoretically learn through the Japanese and Taiwanese textbooks. Table 4.11 below integrates the newly-introduced lemmas, which appear in the textbook more than 10 times and are recycled across more than three units of each textbook and reveals the number of the lemmas, which fulfil both conditions.

Table 4.11: Newly-introduced lemmas that occurred 10 or more times, words that occurred in more than three units and words that both occurred 10 or more times and occurred in more than three units in each Japanese and Taiwanese textbook

	10 and more	more than three units	Words shared
TT1	94	148	76
TT2	56	119	50
TT3	30	78	22
TT4	18	40	10
TT5	10	21	4
TT6	6	14	2
Total			164
JT1	7	9	1
JT2	13	21	5
JT3	15	16	0
Total			6

Table 4.11 revealed that the Taiwanese textbooks had more newly-introduced lemmas, which students could theoretically learn through their textbooks, than the Japanese textbooks. 164 newly-introduced lemmas can be theoretically learnt by the Taiwanese students during three years of high school. The number of the words shared in the Japanese textbooks, however, were significantly small in the comparison of that in the Taiwanese textbooks. That is, the number of newly-introduced lemmas the Japanese students can, in theory, learn incidentally from the textbook input is only six words. Because the Japanese textbooks did

not have many words neither in the “10 and more” repetition bands nor the “more than three units” recycling column at the first point, it might be natural that the number of the words shared becomes smaller.

4.7.1 The differences in the number of the theoretically learnable new lemmas between the Japanese and Taiwanese textbooks

The results seemed to be similar with previous findings in the research question 4 and 5. That is, in terms of the Taiwanese textbooks, the early-grade textbooks including many high-frequency lemmas in their newly-introduced lemmas tend to show relatively many lemmas in the “words shared” column in Table 4.11. By contrast, the later-grade Taiwanese textbooks, which aim at providing mid-frequency words, showed the small number of words in the column. The decrease in the words shared column can be also caused by the shift from the high-frequency lemmas to the mid-frequency lemmas in the newly-introduced lemmas. Regarding to the Japanese textbooks, unlike the Taiwanese textbooks, there was not a certain pattern in the “words shared”. However, this finding also indicates that the Japanese textbooks do not provide their students with rich vocabulary learning opportunities through the contents. Both vocabulary

repetition and opportunities to reencounter the target lemmas in multiple units of a textbooks are strictly limited. As the result, students using this Japanese textbook series are hardly provided with sufficient exposure to new lemmas from their textbooks. Based on the findings of research question 6, the present study concludes that the Taiwanese textbooks contain more newly-introduced lemmas that students can, in theory, learn through textbook exposure than the Japanese textbooks do.

Chapter 5 Conclusion

5.1 Introduction

The purpose of this study was to compare the learning opportunities for high-frequency lemmas and newly-introduced lemmas that Japanese and Taiwanese English textbooks offer EFL students at senior high schools. Words were identified in the corpora created out of these textbooks using two wordlists, new-GSL (Brezina and Gablasova, 2015) and BNC-3000 and 6318 (Kilgarriff, 2006). The answers this study provides to its research questions are summarized and discussed in this chapter. Moreover, this study proposes possible improvements to the Japanese textbooks utilized in this study in light of the findings. The study's limitations will be discussed in the last section along with implications for future research.

5.2 High-frequency vocabulary learning opportunities offered in Japanese and Taiwanese textbooks

The first set of research questions aimed at investigating the learning opportunities for high-frequency lemmas provided by the Japanese and

Taiwanese EFL textbooks targeted in this study. The research questions contained:

1. How much do high-frequency lemmas account for the lexical coverage in the Japanese and Taiwanese textbook series examined in this study?

2. How many of high-frequency lemmas are in the Japanese and Taiwanese textbook series examined in this study?

In order to identify the differences, new-GSL (Brezina and Gablasova, 2015) and BNC-3000, which extracted the most frequent 3000 lemmas from the BNC-6318 (Kilgarriff, 2006), were employed.

Results indicated that the Taiwanese textbooks are superior to the Japanese textbooks in terms of the vocabulary learning opportunities for high-frequency lemmas. With regard to the research question 1, results indicate that high-frequency lemmas provide higher coverage of the Taiwanese textbooks than of the Japanese textbooks (Finding 1). Moreover, the present study also revealed that the Taiwanese textbooks have approximately ten times more word tokens than the Japanese textbooks (Finding 2). Considering these findings of the first research question together, this study reveals that Taiwanese textbooks are able to provide their students with a greater amount of language input that contains

more high-frequency lemmas than Japanese textbooks.

Through the second research question, this study examined how many of high-frequency lemmas are shown in the Japanese and Taiwanese textbooks and the number of lemmas used in the textbooks. This study found that the Taiwanese textbooks contained over 90% of high-frequency lemmas in the new-GSL-2500 and BNC-3000 wordlists, while the Japanese textbooks can provide around half of the lemmas in each wordlist (Finding 3). Additionally, the lemma analysis has also revealed that the Taiwanese textbooks have nearly three times more lemmas than the Japanese textbooks (Finding 4). Thus, since the Taiwanese textbooks contain much more word tokens and various lemmas, the Taiwanese students are given more opportunities to develop their vocabulary knowledge than the Japanese students.

Based on the findings of this first set of research questions, this study concludes that Taiwanese textbooks provide their students with more vocabulary input than Japanese textbooks in terms of both quantity (word tokens) and variety (lemmas). According to Findings 1 and 2, in addition to containing 10 times more word tokens than Japanese textbooks, Taiwanese textbooks also contain a greater proportion of high-frequency lemmas. Finding 1 indicated that the

average percentages of high-frequency lemmas in Taiwanese textbooks made their superiority over Japanese textbooks even more pronounced. In addition, Finding 2 showed that during the three years of senior high school, Taiwanese students were given the opportunity to read 283,895 word tokens in their textbooks, while Japanese students only 28,985. Linguistic input can be beneficial not only for learning something new, but also for enhancing previously acquired knowledge. Consequently, these findings indicate that Taiwanese students have more opportunities to learn new English vocabulary and deepen their vocabulary knowledge than Japanese students.

Moreover, Findings 3 and 4 also proved that the Taiwanese textbooks were more helpful for students' vocabulary development, not only for high-frequency lemmas, but also others. Finding 3 revealed that the percentages of the usages of high-frequency lemmas in new-GSL-2500 and BNC-3000 in the Taiwanese textbooks were much higher than that of the Japanese textbooks. The Taiwanese textbooks showed over 90% of high-frequency lemmas in both wordlists during the senior high school years, whereas the percentages in the Japanese textbooks were around 50%. Thus, while the Taiwanese students can encounter most high-frequency lemmas through their senior high school

textbooks, the Japanese students can only encounter about half of them. This may allow Taiwanese students to comprehend the contents of their readings more thoroughly and perceive them as more readable than Japanese students. Additionally, Finding 4 revealed that the total number of lemmas (including proper nouns) in Taiwanese textbooks was 9039, while the total number of lemmas in Japanese textbooks was 2936. Consequently, Taiwanese students are given the opportunity to learn three times more lemmas than Japanese students from their textbooks during their senior high school years. Even though proper nouns are included in the present study, the difference seems huge.

5.3 Newly-introduced lemmas in the Japanese and Taiwanese textbooks

The second set of research questions aims at investigating differences in learning opportunities of newly-introduced lemmas in the Japanese and Taiwanese textbooks. This set contained the following research questions:

3. Which frequency bands in the BNC-6318 do newly-introduced lemmas in the Japanese and Taiwanese textbooks come from?

4. How many times do the newly-introduced lemmas occur in the Japanese

and Taiwanese textbooks?

5. How many newly-introduced lemmas recur across units of the Japanese and Taiwanese textbooks?

6. How many of the newly-introduced lemmas are, in theory, learnable through the Japanese and Taiwanese textbooks?

For this set of research questions, the BNC-6318 wordlist (Kilgarriff, 2006) was employed because the wordlist could show some mid-frequency lemmas in the newly-introduced lemmas although it did not include all of the mid-frequency lemmas. By comparing Japanese and Taiwanese textbooks in terms of frequency levels, vocabulary repetitions, and the distances between each repetition of newly-introduced lemmas within units of textbooks, this study determined which textbook series provides students with more robust opportunities for vocabulary acquisition. Besides, from the findings on lemma repetition in general and across units, it estimated how many of newly-introduced lemmas are theoretically learnable.

Results indicated that Taiwanese textbooks were more advantageous than Japanese textbooks in terms of the introduction of new lemmas. In response to the first research question of the second set, this study identified a pattern in

the frequency levels of newly-introduced lemmas in Taiwanese textbooks. Across the senior high school years, the frequency levels of newly-introduced lemmas in the Taiwanese textbooks shifted from high-frequency to mid-frequency levels. At the first year of senior high school in Taiwan, for instance, 61% of all newly-introduced lemmas are high-frequency lemmas. However, the number of high-frequency lemmas becomes only 16% in the Taiwanese textbooks used in the last grade. This decline in high-frequency lemmas is due to an increase in mid-frequency lemmas as students progressed through school. The percentage of mid-frequency lemmas increased from 27% in the first year to 39% in the last grade. By contrast, although the Japanese textbooks introduce some BNC-6318 lemmas, no pattern of increase or decrease in terms of high-frequency lemmas was observed in these textbooks. Thus, this study concluded that Taiwanese textbooks select newly-introduced lemmas based on their frequency levels at least as one of the criteria of vocabulary selection, whereas Japanese textbooks appear not to focus on the word frequency levels in the selection of newly-introduced lemmas (Finding 5).

The present study also compared the number of newly-introduced lemmas between the Taiwanese and Japanese textbooks. During the three-year

period, Taiwanese textbooks introduced 2,790 new words, while Japanese textbooks introduced 1,296 new words. Thus, Taiwanese students are given the opportunity to learn more than twice as many new words as Japanese students (Finding 6).

The second research question in this set examined differences between Japanese and Taiwanese textbooks regarding the vocabulary repetitions of newly-introduced lemmas. This study grouped the lemmas across five repetition bands: "once", "2 to 3" times, "4 to 6" times, "7 to 9" times, and "10 or more" times. Webb (2007) reports that vocabulary learning can occur even with a few repetitions, but 10 or more encounters appear to be necessary for learners to master both receptive and productive knowledge. Taiwanese textbooks presented every newly-introduced lemma at least twice, whereas Japanese textbooks presented a large number of new lemmas only once. Furthermore, the majority of newly-introduced lemmas in Japanese textbooks fall under the "2 to 3" band, whereas in Taiwanese textbooks, many lemmas appeared in both the "2 to 3" and "4 to 6" bands without large gaps. Thus, Taiwanese textbooks tend to display newly-introduced vocabulary more than Japanese textbooks.

In addition, the number of new words in the "7 to 9" and "10 and more"

bands gradually decreased in Taiwanese textbooks as students advanced through school. The analysis of mid-frequency vocabulary in the textbook indicates that the aforementioned decrease in 7 or more lemma repetitions in the Taiwanese textbooks could be because the Taiwanese textbooks offer their students more mid-frequency level lemmas along with students' grade in school. Since mid-frequency lemmas are not likely to be reused frequently, by increasing the number of them, it is possible that the Taiwanese textbooks do not repeat these lemmas as often as high-frequency lemmas. This fact aligns with the shift in the frequency levels of newly-introduced lemmas in Finding 5. However, the bands on Japanese textbooks were consistently very low during senior high school. Therefore, this study concluded that Taiwanese textbooks tend to provide their students with more vocabulary repetitions for newly-introduced lemmas than Japanese textbooks (Finding 7).

The third research question examined how many newly-introduced lemmas were recycled throughout different units of each Japanese and Taiwanese textbook. The recycling of lemmas was categorised across three bands, "one unit," "two units," and "more than three units". Results indicate that the Taiwanese textbooks provide more opportunities for students to revisit

newly-introduced lemmas across units than Japanese textbooks. The number of new lemmas in "more than three units" of Taiwanese textbooks was significantly greater than that of Japanese textbooks, particularly in the first grade. While the Japanese textbooks contained 3-4% of newly-introduced lemmas in the "more than three units" band during the senior high school years, the Taiwanese textbooks contained 30 and 24% at the first grade, 15 and 8% at the second grade and 5 and 4% at the last grade. This finding indicates that Taiwanese textbooks tend to reuse newly-introduced lemmas throughout each unit of the textbook, whereas Japanese textbooks tend to use newly-introduced lemmas only in the unit where target lemmas are introduced (Finding 8).

Although Taiwanese textbooks display high percentages of newly-introduced lemmas reused in "more than three units" in the first grade, the percentages decrease as students' progress through the grades. Thus, this study analysed the distribution of the level of newly-introduced lemmas in "more than three units" in Taiwanese textbooks to determine the cause of the decline during the three years. The results indicated that the majority of lemmas in the "more than three units" band of Taiwanese textbooks were high-frequency level lemmas. As the frequency level research question revealed, the number of high-frequency

lemmas in the newly-introduced lemmas decreased gradually. As a result, it is conceivable that the number of lemmas belonging to the "more than three units" band decreased as a result of the reduction of high-frequency lemmas. This phenomenon parallels the decrease in vocabulary repetition in the "7 to 9 units" and "10 or more units" bands of Taiwanese textbooks. This study concluded that the decrease in Taiwanese textbooks appears to have reasonable causes, whereas there was no discernible pattern in Japanese textbooks.

Lastly, given the findings relating to vocabulary repetition and the recycling of newly-introduced lemmas, this study attempts to estimate the number of lemmas that students can, in theory, learn from the textbooks. Webb (2007) suggests that 10 or more encounters are necessary for effective vocabulary learning, and Webb and Nation (2017) note that the space between repetitions can enhance vocabulary learning. Assuming these claims are correct, one can determine how many newly-introduced lemmas in Japanese and Taiwanese textbooks meet the two essential conditions of appearing at least ten times and being recycled in more than three units of each textbook. The results revealed that 164 new lemmas out of 2,790 lemmas in Taiwanese textbooks and six new lemmas out of 1,296 lemmas in Japanese textbooks met both criteria. This finding

highlights the fact that newly-introduced lemmas in Japanese textbooks tend to occur and re-occur only within the unit where they are introduced. Hence, this study concluded that Taiwanese students can learn more newly-introduced lemmas thanks to repeated encounters with them in their textbooks, whereas the opportunities to learn new words through repeated encounters in the Japanese textbooks are limited (Finding 9).

5.4 Summary of the findings in this study

In conclusion, this study has revealed the following differences between the Japanese and Taiwanese textbooks:

1. The word token analysis revealed that high-frequency lemmas provided greater coverage of the Taiwanese textbooks than the Japanese textbooks.
2. The Taiwanese textbooks contain roughly ten times as many word tokens as Japanese textbooks.
3. The Taiwanese textbooks provided their students with over 90% of the lemmas in the new-GSL-2500 and BNC-3000 wordlists, whereas the Japanese textbooks provide approximately 50% of the wordlists in each.
4. The Taiwanese textbooks contain almost three times as many lemmas as

Japanese textbooks.

5. The Taiwanese textbooks appear to select newly-introduced lemmas based on the frequency levels of vocabulary, whereas vocabulary selection in Japanese textbooks seems not.
6. During senior high school, the Taiwanese textbooks introduce approximately twice as many new words as Japanese textbooks.
7. The Taiwanese textbooks are more likely than Japanese textbooks to repeat newly-introduced lemmas.
8. The Taiwanese textbooks appear to recycle newly-introduced vocabulary in later units, whereas the Japanese textbooks tend not to recycle newly-introduced vocabulary throughout the units.
9. In contrast to their Japanese counterparts, the Taiwanese textbooks appear to possess newly-introduced lemmas with theoretically learnable components by satisfying two crucial learning conditions: vocabulary repetition and its distance.

5.5 Pedagogical implications

Based on the aforementioned findings, this section of the study proposes

potential improvements to the Japanese textbooks. There appear to be two significant differences between these two countries that can impact students' vocabulary learning the most. The amount of vocabulary input that students can obtain from their textbooks is one consideration. The second is the vocabulary that students are expected to learn from their textbooks. The following two sections discuss these points and offer some suggestions for improving the effectiveness and utility of Japanese textbooks for teaching vocabulary to Japanese students.

5.5.1 The need to increase the amount of vocabulary input in the Japanese textbooks

It has been extensively discussed and acknowledged that language input is essential for the success of language learners (e.g., Krashen, 1985; Nation, 2007; Nation and Webb, 2017). In addition, EFL students must receive a substantial amount of input to develop and consolidate their vocabulary knowledge. According to Milton (2009), EFL students' textbooks can be a primary source of language input. Therefore, limited language input in textbooks can have a direct impact on the vocabulary development of EFL students. This study

reveals that Taiwanese textbooks contain roughly ten times as many running words or language input as Japanese textbooks. This difference may be one of the causes for certain findings in this study, as the number of running words can influence the variety of words students can encounter in the text. For instance, Findings 3 and 4 indicate a distinction between the number of lemmas and the variety of high-frequency lemmas that students encounter in textbooks. The Taiwanese textbooks contain three times as many lemmas and greater coverage of new-GSL-2500 (Brezina and Gablasova, 2015) and BNC-3000 (Kilgarriff, 2006) than the Japanese counterparts. Furthermore, Finding 7 suggests that the Japanese textbooks repeat newly-introduced lemmas less frequently than Taiwanese textbooks. Because Taiwanese textbooks contain ten times more running words than Japanese textbooks, there are ten times more opportunities to show various words and reuse them in Taiwanese textbooks. However, a large number of running words does not always ensure that students will encounter a wide variety of words. If textbook creators aim to include a variety of words on purpose, the large number of running words can provide more room to include them than the small number of running words. Therefore, it is necessary to aim for a large amount of running words in textbooks so that students will have many

opportunities to encounter a rich variety of words and acquire or learn them via textbooks.

The use of L1 in Japanese textbooks contributes to the possibility of such a distinction. While Taiwanese textbooks provide the entire contents of textbooks in English, the Japanese textbooks utilize L1 to introduce practical questions, explain grammar, and elicit answers to practical questions. In Chapter 4, this study concluded that the majority of L1 words in Japanese textbooks appears reasonable. However, it is essential to compensate for the lack of English language input resulting from L1 usage in the Japanese textbooks through increasing the number of words in the reading passages in textbooks or through teachers utilizing external materials such as Graded readers.

Therefore, although it should be true that Taiwanese students also need more language input, the first pedagogical implication is that the Japanese textbooks must increase their English running words so that students have more opportunities to encounter and learn various words. The Japanese textbooks examined in this study do not fulfil their function as the most significant source of language input (Milton, 2009). The tenfold difference in the amount of input between Japanese and Taiwanese textbooks suggests that the Japanese

textbooks may contain fewer examples of English usage. In terms of introducing new vocabulary, Taiwanese textbooks provide an English definition and example sentence for each target word, whereas Japanese textbooks simply list the target vocabulary in each unit only with their pronunciation symbols. In addition, this study reveals that numerous newly-introduced lemmas appear only once in the Japanese textbooks because the lemmas are only presented with their image in an activity. Providing the English definition and a few example sentences of target vocabulary, as is the case in the Taiwanese textbooks, will increase the likelihood that students will encounter and comprehend target lemmas. Moreover, increasing the number of running words in each unit's reading content can help students acquire a great deal of language input through textbooks. The running words in each reading section of Japanese textbooks are relatively few compared to those in the Taiwanese textbooks. Extending the reading texts in the Japanese textbooks allows students to gain much language input through their textbooks, and if textbook publishers include vocabulary they want students to learn, such as high-frequency lemmas or newly-introduced lemmas, Japanese students will have more opportunities to learn the words. Moreover, it should be possible that textbook publishers adjust the level of the reading text so that students can

comprehensively read the text at the “i+1” level by controlling the vocabulary used in the text as Graded readers do.

Unit 5 and Unit 10 of each Japanese textbook contain two longer reading passages that serve as a reading activity. It appears feasible to implement such reading activities in every Unit or to use them as the reading content of each Unit in order to increase the number of running words in textbooks. Thus, it appears that there are numerous ways to increase the amount of English in Japanese textbooks. Since language input is something students must have in order to improve their English and the textbooks they use tend to be the ones they rely on in an EFL environment, it is crucial to consider how to maximize language input in the content.

Utilizing external reading materials, such as Graded readers, is one way to increase the amount of language input in Japanese EFL teaching. Graded readers are a form of reading material designed for extensive reading by learners of English. They restrict the vocabulary used in the texts, control the grammatical structures in the text, and adjust the length of the text in accordance with the vocabulary and grammar controls (Nation, 1999). Moreover, according to Hirano

(2019)[5], Japanese textbooks and the Yo-han ladder series, one of the graded readers produced in Japan, have a high level of compatibility with regard to the use of vocabulary. In the event that it is difficult for Japanese textbooks to increase the number of language input due to constraints such as national curriculum guidelines or rules governing the creation of textbooks, graded readers may be useful for Japanese students to increase their language input or improve their vocabulary learning. Even though graded readings may have some shortcomings, such as the inauthentic English use, they can increase the encounters English learners have with useful vocabulary.

5.5.2 The need for standards in vocabulary selection for textbooks

The frequency counts of vocabulary can be a useful tool for selecting textbook terms and newly-introduced lemmas for inclusion in content. According to Nation (1990), word-frequency counts can assist teachers and course

[5] Unpublished master dissertation, which examined the effectiveness of Yo-han ladder series as an external source of vocabulary introduced in Japanese textbooks

designers in determining which words in textbooks should be emphasized and offered to students. In addition, he mentions that EFL students need to learn the most frequent 2000 words and some strategies for dealing with unknown words. In addition, Schmitt (2008) asserts that learning the most frequent 2000 words can help EFL students increase their comprehensible input and establish a foundation for learning English. O'Dell (1997) suggests a number of criteria for material designers to consider when choosing textbook vocabulary, including frequency of use, coverage, range, availability, learnability, opportunism, and centres of interest (See more detail in Chapter 2 at section 3). Yet, Milton (2009) asserts that focusing solely on the frequency of words can render students' learning ineffective, as it is difficult to create English texts that contain only high-frequency lemmas. Therefore, it should be true that textbooks need to show their students mid-frequency or infrequent words as well, while maintaining a balance between high-frequency and other words such as topic-related vocabulary. However, because learning high-frequency lemmas can be extremely beneficial for students, textbooks must provide opportunities for students to learn these useful words prior to entering EAP or ESP stages at university.

In comparison to Taiwanese textbooks, this study has revealed that

Japanese textbooks do not pay enough attention to words included in the new-GSL and BNC-3000 wordlists. While Taiwanese textbooks provide students with a large quantity and variety of high-frequency lemmas in both running words and newly-introduced lemmas, Japanese textbooks contain fewer high-frequency lemmas and approximately 50 percent of the lemmas in the new-GSL and BNC-3000 wordlists. In other words, Japanese students have limited opportunities to learn high-frequency lemmas, whereas Taiwanese students are exposed to nearly all high-frequency lemmas in textbooks. In addition, this study reveals that the frequency levels of newly-introduced lemmas in Taiwanese textbooks decrease from a high frequency level to a mid- or low-frequency level during the senior high school years. However, such a pattern cannot be observed in this study for newly-introduced lemmas in Japanese textbooks. Thus, this study concludes that Taiwanese textbooks pay more attention to the frequency levels of vocabulary within their content than Japanese textbooks.

This study proposes employing frequency-based wordlists as one of the criteria in the selection of vocabulary in Japanese textbooks or at least emphasizes a heightened awareness of the importance of word frequency in the process. Milton (2009) suggests that textbooks should include a large number of

new words so that students can encounter and learn a wide range of new words by including both high-frequency and low-frequency words roughly equal. In addition, he notes even though textbooks limit the number of newly-introduced lemmas, students cannot master them all. However, the greater the number of words introduced in textbooks, the more words students tend to learn through textbooks (Milton, 2009). Furthermore, as stated by Nation (1990), the use of such frequency-based wordlists can be beneficial until students learn the most important words, and the words they learn from wordlists can assist them in dealing with unknown words. From the points, this study suggests that a reference wordlist for junior and senior high school English textbooks should be created or that Japanese textbook creators should use frequency wordlists such as the new-GSL, the BNC wordlist, the BNC/COCA list or the NGSL in order to choose words to include in textbooks.

5.6. Limitations of the study

There seem to be some limitations in the present study. First, this study examined every English word used in the textbooks because it assumed that every English word in the textbooks constitutes English language input for

students. It is reasonable to assume that all English vocabulary included in a textbook may be attended to by students and, therefore, constitute an opportunity for vocabulary learning. However, due to this decision, the actual number of the words in each reading text of each unit became unclear. From the findings, this study concluded that the Taiwanese textbooks provided their students with much more English input than the Japanese textbooks. However, the difference seemed to be caused by the use of L1 in the Japanese textbooks. Therefore, if the present study had focused solely on the reading texts, the results could be different from this time. Given that reading texts in the textbooks entail students actively engaging with and striving to comprehend English consciously, there might be a heightened significance as language input compared to other English passages such as example sentences of new words or introduction of each practice questions. Moreover, the Taiwanese textbooks provide each newly-introduced lemmas with an example sentence, while the Japanese textbooks provide their students with only the meaning or just form. The amount of the vocabulary used in each example sentence and the vocabulary itself may be able to be a good language input for the Taiwanese students. However, since this study focused on the repetition and recycling rates of newly-introduced lemmas,

the impact of the example sentences as language input cannot be analysed. Third, since proper nouns were also counted in the present study, it was quite difficult to show the exact number of lemmas except proper nouns in this study. Fourth, the lack of a frequency-based lemmatized wordlist, which contains every mid-frequency lemma, the present study could not show the exact details of mid-frequency lemmas in the textbooks. If there was, the present study could show more detailly the use of mid-frequency lemmas and newly-introduced lemmas. Future research needs to cover these points so as to more deeply analyse the vocabulary learning opportunities on textbooks between Japan and Taiwan.

5.6.1 Avenues for future research

Although this study focused on the opportunities for learning high-frequency and newly-introduced lemmas from Japanese and Taiwanese textbooks, their actual learnability remains unknown. In other words, this study identified differences in their vocabulary learning opportunities from a theoretical perspective, such as the coverage and variety of high-frequency lemmas in their running words and the repetition and re-encountering rates of newly-introduced lemmas throughout textbook units as the first step to identify the differences of

the vocabulary learning opportunities between the Japanese and Taiwanese textbooks. However, the number of words students can actually learn from textbooks is not addressed in this project. Consequently, this study needs further research to examine student's vocabulary learning gains through the textbooks.

Furthermore, only one textbook series from Taiwan and Japan was analysed in this study. The textbook series selected for this study were chosen because they are widely used in each country. Due to the fact that there are other textbooks in both countries, it is unclear whether the same results will be observed if a similar study focuses on other textbooks. Thus, similar studies need to be conducted on more various English textbooks used in Japanese and Taiwanese senior high school in order to ascertain whether similar findings in this study can be observed in other English textbooks of both countries. In addition, the Japanese English education includes other classes for English expressions and grammar. That is, the analysed textbooks in the present study pertain solely on one of the English classes in senior high school. Since Japanese students can gain some English input from other classes of English in school, in order to comprehensively address senior high school English education, it might be imperative to investigate other English textbooks employed in other courses as

well.

At last, this study did not compare vocabulary learning activities in Japanese and Taiwanese textbooks. This study confirmed that Taiwanese textbooks provide students with more opportunities to encounter target lemmas in the content than Japanese textbooks, but it is unclear how Japanese and Taiwanese textbooks differ in terms of qualitative opportunities, such as opportunities to practice the use of target lemmas or retrieve them on certain activities. Activities can play an important role in vocabulary learning in order for students to learn and retain the vocabulary. Thus, future research needs to examine whether the target textbooks in this study provide their students with consolidate vocabulary learning activities of the textbooks.

References

Ahmed Alshumrani, H., & Masoud Al-Ahmadi, N. (2022). The Representation of Vocabulary Knowledge Aspects in Saudi EFL Textbooks. *Arab World English Journal, 13*(4), 325–340. https://doi.org/10.24093/awej/vol13no4.21

Alcaraz, G. (2009). Frequency and Functionality: Two Keys for L2 Coursebooks. www.um.es/ijes

Alsaif, A. & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by EFL learners in *Saudi Arabia. Language Learning Journal. 40*(1). 21-33. 10.1080/09571736.2012.658221

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), Comprehension and Teaching: *Research Reviews* (pp. 77-117). Newark, DE: International Reading Association.

Anthony, L. (2014). AntConc (Version 3.4.4) [computer software]. Tokyo: Waseda University.

Anthony, L. (2022). AntFileConverter (Version 2.0.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Atikah, D., & Rezki, A. (2018). Repetition Facilitates Retrieval Opportunity in Vocabulary Learning. *IOP Conference Series: Earth and Environmental Science, 175*(1). https://doi.org/10.1088/1755-1315/175/1/012148

Atkinson, D. (1987). The mother tongue in the classroom: a neglected resource? https://academic.oup.com/eltj/article/41/4/241/356996

Ayu, M., & Inderawati, R. (2018). EFL Textbook Evaluation: The Analysis of Tasks Presented in English Textbook. *TEKNOSASTIK*, *16*(1). 21-25.

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning, 57*(1), 35–56. https://doi.org/10.1111/j.1467-9922.2007.00398.x

Barcroft, J. (2015). Can Retrieval Opportunities Increase Vocabulary Learning During Reading? *Foreign Language Annals, 48*(2), 236–249. https://doi.org/10.1111/flan.12139

Bergström, D., Norberg, C., & Nordlund, M. (2023). “The Text Comes First”–Principles Guiding EFL Materials Developers’ Vocabulary Content Decisions. *Scandinavian Journal of Educational Research, 67*(1), 154–

168. https://doi.org/10.1080/00313831.2021.1990122

Biber, D., Conrad. S. and Reppen, R. (1998). *Corpus Linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498–520.

Bouhlal, F., Horst, M., & Martini, J. (2018). Modality in ESL Textbooks: Insights from a Contrastive Corpus-Based Analysis. *The Canadian Modern Language Review / La revue canadienne des langues vivantes, 74*, 227 - 252.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the new general service list. *Applied Linguistics, 36*(1), 1–22. https://doi.org/10.1093/applin/amt018

Brown, D. (2011). What aspects of vocabulary knowledge do textbooks give attention to? *Language Teaching Research, 15*(1), 83–97. https://doi.org/10.1177/1362168810383345

Brown, D., Stoeckel, T., Mclean, S., & Stewart, J. (2022). The Most Appropriate Lexical Unit for L2 Vocabulary Research and Pedagogy: A Brief Review of the Evidence. *Applied Linguistics, 43*(3), 596–602. https://doi.org/10.1093/applin/amaa061

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a foreign language, 20*(2), 136-163.http://nflrc.hawaii.edu/rfl

Browne, C. (1996). Japanese High School Textbooks: How Readable Are They? Temple University Japan. Working Paper in *Applied Linguistics, 8,* 28-41. Available at: https://www.wordengine.jp/research/pdf/we30/Japanese_High_School_Textbooks-How_Readable_Are_They.pdf

Browne, C., Culligan, B., and Phillips, J. (2013). The new general service list. Available at: https://www.newgeneralservicelist.com/new-general-service-list

Cao, T.H.P. (2018). Vocabulary in EFL Textbook: An Analysis of “Life A2-B1” Coursebook Used for Vietnamese Tertiary Students. Proceedings of the 7thVietnamese Young Researchers Conference in Education at Hanoi

National University of Education, Vietnam (section 3, 548-556). Hanoi National University of Education Publishing House.

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713. https://doi.org/10.1093/applin/amq031

Cheung, H., Chung, S.F., & Skoufaki, S. (2010). Indexing second language vocabulary in the Intermediate GEPT. In Proceedings of the Twelfth Academic Forum on English Language Testing in Asia (Language Testing in Asia: Continuity, Innovation and Synergy), March, 5-6, 2010 (pp.118-136). Taipei: Language Training and Testing Center.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology, 11*, 38-64. http://llt.msu.edu/vol11num3/cobb/

Cobb, T. (2009). Web VocabProfile (Version: VP English v.3). Available at: http://www.lextutor.ca/vp.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213–238

Criado, R., & Sánchez, A. (2009). Vocabulary in EFL Textbooks. A Contrastive Analysis against Three Corpus-Based Word Ranges. AELINCO.

Crothers, E., & Suppes, P. (1967). *Experiments in Second-Language Learning.* New York: Academic Press.

D'Anna, C. A., Zechmeister, E. B., and Hall, J. W. (1991). Toward a Meaningful Definition of Vocabulary Size. *Journal of Reading Behavior*, *23*(1), 109-122. https://doi.org/10.1080/10862969109547729

Daigaku-eigokyoikugakkai-kihonngokaitei-tokubetuiinkai. (The Japan Association of College English Teachers: JACET) (2016). New JACET 8000- The new JACET list of 8000 basic words. Kiriharasyoten

Dang, T. N. Y. and Webb, S. (2016). Making an essential word list for beginners. In: Nation, ISP, (Ed.) *Making and Using Word Lists for Language Learning and Testing.* John Benjamins. 153-167. https://doi.org/10.1075/z.208.15ch15

Dang, T. N. Y., & Webb, S. (2016b). Evaluating lists of high-frequency words. *ITL - International Journal of Applied Linguistics, 167*(2), 132–158. https://doi.org/10.1075/itl.167.2.02dan

Dang, T. N. Y., Webb, S., & Coxhead, A. (2022). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research, 26*(4), 617–641. https://doi.org/10.1177/1362168820911189

Daulton, E. (2008). *Japan's built-in Lexicon of English-based loanwords.*

Multilingual Matters.

Demetriou, L. (2017). Selection, presentation and recycling of phrasal verbs in ESL textbooks in Cyprus. Paper presented at the 2017 BAAL vocabulary special interest group Conference, Reading.

Educational Testing Service. (2018). TOEFL iBT test and score data summary 2018. ETS.

Educational Testing Service. (2019). TOEFL iBT test and score data summary 2019. ETS.

Educational Testing Service. (2020). TOEFL iBT test and score data summary 2020. ETS.

Educational Testing Service. (2021). TOEFL iBT test and score data summary 2021. ETS.

Eldridge, J.E., & Neufeld, S. (2009). The Graded Reader Is Dead, Long Live the Electronic Reader. *The Reading Matrix: an International Online Journal, 9*, 224-244.

Ellis, R. (1997). *Second language acquisition.* Oxford University Press.

Fan, M. (2000). How big is the gap and how to narrow it? an investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal, 31*(2), 105–119. https://doi.org/10.1177/003368820003100205

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Goulden, R., Nation, I. S. P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*(4), 341–363.

Harris, J. (2019). Issues related to the presence of Japanese loanwords of English origin in vocabulary size tests. *The Asian Journal of Applied Linguistics, 6*(1), 1–13.

Heatley, A., Nation, I.S.P. & Coxhead, A. (2002). RANGE and FREQUENCY programs. Retrieved from: http://www.vuw.ac.nz/lals/staff/Paul_Nation

Henriksen, B. (1999). Three dimensions of vocabulary development. Studies in *Second Language Acquisition, 21*(2), 303–317. https://doi.org/10.1017/s0272263199002089

Hirano, T. (2019). The examination of the vocabulary acquisition through Japanese English textbooks and the Yo-han ladder series. [Unpublished Master Dissertation]. The University of Essex.

Hirsh, D.& Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689–696.

https://doi.org/10.26686/wgtn.12560417

Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a foreign language, 11*, 207-223.

Hsu, W. (2009). Measuring the Vocabulary of College General English Textbooks and English-medium Textbooks of Business Core Courses. *Electronic Journal of Foreign Language Teaching, 6*(2). 126-149. http://e-flt.nus.edu.sg/ https://www.researchgate.net/publication/328494810

Hu, M., & Nation, P. (2020). Unknown vocabulary density and reading comprehension. *Reaing in a Foreign Language, 13*(1), 403–430. https://doi.org/10.26686/wgtn.12560354.v1

Hunston. S. (2002). *Corpora in Applied Linguistics. Cambridge:* Cambridge University Press

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language, 17*(1). http://nflrc.hawaii.edu/rfl

Hyland, K., & Tse, P. (2004). Metadiscourse in Academic Writing: *A Reappraisal. Applied Linguistics, 25*, 156-177. http://dx.doi.org/10.1093/applin/25.2.156

Jordan, G., & Gray, H. (2019). We need to talk about coursebooks. *ELT Journal, 73*(4), 438–446. https://doi.org/10.1093/elt/ccz038

Karami, A., & Bowles, F. A. (2019). Which Strategy Promotes Retention? Intentional Vocabulary Learning, Incidental Vocabulary Learning, or a Mixture of Both? *Australian Journal of Teacher Education, 44*(9). https://doi.org/10.14221/ajte.2019v44.n9.2

Kilgarriff, A. (2006). BNC database and word frequency lists. Read-me for Kilgarriff's BNC word frequency lists. http://Kilgarriffff.co.uk/bnc-readme.html

Krashen, S. D. (1981). *Second language acquisition and Second language learning.* Pergamon Press.

Lado, R., Baldwin, B., & Lobo, F. (1967). Massive vocabulary in a foreign language beyond the basic course: the effects of stimuli, timing and order of presentation. Washington, D.C.: U.S. Department of Health, Education, and Welfare.

Laufer, B. (1989). What percentage of text lexis is necessary for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters

Laufer, B. (1992). How Much Lexis is Necessary for Reading Comprehension? In P. J. L., Arnaud, & H. Béjoint. (Eds.). *Vocabulary and Applied Linguistics.* Palgrave Macmillan. https://doi.org/10.1007/978-1-349-12396-4_12

Laufer, B. (1996). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy* (pp. 20–34). Cambridge University Press.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*(2), 255–271. https://doi.org/10.1093/applin/19.2.255

Laufer, B. (2003). Vocabulary Acquisition in a Second Language: Do Learners Really Acquire Most Vocabulary by Reading? Some Empirical Evidence. *The Canadian Modern Language Review, 59,* 567-587. http://dx.doi.org/10.3138/cmlr.59.4.567

Laufer, B. (2006). Comparing focus on form and focus on FormS in second-language vocabulary learning. *Canadian Modern Language Review*, *63*(1), 149–166. University of Toronto Press. https://doi.org/10.3138/cmlr.63.1.149

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Laufer, B., & Paribakht, T. S. (1998). The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning, 48*(3), 365–391. https://doi.org/10.1111/0023-8333.00046

Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. *Language Teaching Research, 15*(4)*,* 391–411. https://doi.org/10.1177/1362168811412019

Laufer, B., & Rozovski-Roitblat, B. (2015). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge. Language Teaching Research, 19(6), 687–711. https://doi.org/10.1177/1362168814559797

Le, N. T. M., & Dinh, H. T. (2022). Vocabulary Coverage in a High School Vietnamese EFL Textbook: A Corpus-based Preliminary Investigation. *Vietnam Journal of Education, 6*(2). 102-113 https://doi.org/10.52296/vje.2022.187

Lemhöfer, K. & Mirjam B. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods 44*(2).

325–343.

Lightbown, P., & Spada, N. M. (2013). How languages are learned. Oxford University Press.

Liu, J., & Zhang, J. (2018). The Effects of Extensive Reading on English Vocabulary Learning: A Meta-analysis. *English Language Teaching, 11*(6), 1-15. https://doi.org/10.5539/elt.v11n6p1

Marzban, A., & Hadipour, R. (2012). Depth Versus Breadth of Vocabulary Knowledge: Assessing their Roles in Iranian Intermediate EFL Students' Lexical Inferencing Success Through Reading. *Procedia - Social and Behavioral Sciences, 46*. https://doi.org/10.1016/j.sbspro.2012.06.426

Matsuoka, W., & Hirsh, D. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language, 22*(1). http://nflrc.hawaii.edu/rfl

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics 39*. 823–45.

McLean, S., Stewart, J., & Batty, A. O. (2020). Predicting L2 reading proficiency with modalities of vocabulary knowledge: A bootstrapping approach. *Language Testing, 37*(3), 389–411. https://doi.org/10.1177/0265532219898380

Mehrpour, S., Razmjoo, S., & Kian, P. (2011). The Relationship between Depth and Breadth of Vocabulary Knowledge and Reading Comprehension among Iranian EFL Learners. *Journal of English Language Teaching and Learning, 222*(53). 97-127.

Milton, J. (2006). Language lite: learning French vocabulary in school. *Journal of French Language Studies, 16*(2), 187–205.

Milton, J. (2009). *Measuring second language vocabulary acquisition.* Multilingual Matters.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. In Source: *Reading Research Quarterly, 20*(2). Winter.

Nakayama, S. (2021). A Quantitative Analysis of Vocabulary Taught in Japanese EFL Textbooks. *Research Square.* https://doi.org/10.21203/rs.3.rs-688772/v1

Nakayama, S. (2022). A Close Examination of Vocabulary in Japanese EFL Textbooks. *JALT Postconference Publication, 2021*(1), 209. https://doi.org/10.37546/jaltpcp2021-24

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines 5*(1): 12-25.

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Heinle & Heinle Publishers.

Nation, I. S. P., & Webb, S. A. (2011). *Researching and analyzing vocabulary.* Cengage Learning.

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

Nation, I.S.P. (2007) The four strands. *Innovation in Language Learning and Teaching 1*(1) 1-12.

Nation, I.S.P. (2012). The BNC/COCA word family lists. Available at: http://www.victoria.ac.nz/

Nation, P. (1993). Vocabprofile. (IBM Computer Program). Nation: New Zealand.

Nation, P. (2003). The role of the first language in foreign language learning. *Asian EFL Journal.* Available at: https://www.asian-efl-journal.com/june_2003_pn.pdf

Nation, P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer. (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 3–13). John Benjamins.

Nation, P. (2013). *Learning vocabulary in another language (2nd ed.).* Cambridge University Press.

Nation, P. and Newton, J. (1997). Teaching vocabulary. In: J., Coady & T., Huckin (Eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*. Cambridge Applied Linguistics. Cambridge University Press; 238-254. doi:10.1017/CBO9781139524643.017

Nerbonne, J. & Hinrichs, E. (2006). Linguistic Distances. Proceedings of the Workshop on Linguistic Distances. *Association for Computational Linguistics*, 1- 6.

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research, 17*(2), 164–187. https://doi.org/10.1177/1362168812460814

NishiKawa-Van Eester, M. (2016). “Linguistic Distance” and Its Influence on TESOL. Kokusaiseikei, 22. 21-32. https://core.ac.uk/download/pdf/229746826.pdf

O’Dell, F. (1997). Incorporating vocabulary into the syllabus. In Schmitt, N. & McCarthy, M. (2011) *Vocabulary: Description, Acquisition & Pedagogy (11th printing).* Cambridge University Press. 258-278.

O’Loughlin, R. (2012). Tuning in to vocabulary frequency in coursebooks. *RELC*

Journal, 43(2), 255–269. https://doi.org/10.1177/0033688212450640

Orosz, A. (2009). The growth of young learners' English vocabulary size. In M. Nikolov (Eds.), *Early learning of modern foreign languages.* 181–194. Multilingual Matters.

Pellicer-Sánchez, A. (2014). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking Study. *Studies in Second Language Acquisition, 38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381–402. https://doi.org/10.1177/1362168815618428

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do Things Fall Apart? *Reading in a Foreign Language, 22*(1). http://nflrc.hawaii.edu/rfl

Peters, E. (2014). The effects of repetition and time of post-test administration on EFL learners' form recall of single words and collocations. *Language Teaching Research, 18*(1), 75–94. https://doi.org/10.1177/1362168813505384

Peters, E., & Webb, S. (2018). Incidental vocabulary Acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551–577. https://doi.org/10.1017/S0272263117000407

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1). http://nflrc.hawaii.edu/rfl

Qian, D. D. (1989). Depth of vocabulary knowledge: Assessing its role in adult's reading comprehension in English as a Second Language. Doctoral thesis. The University of Toronto. National Library of Canada. https://tspace.library.utoronto.ca/bitstream/1807/12079/1/NQ33914.pdf

Rahman, S., Iqbal, K., & Nadeem, A. (2019). Effect of Working Capital Management on Firm Performance: The Role of Ownership Structure. *Global Social Sciences Review, 4,* 108-119.https://doi.org/10.31703/gssr.2019(IV-I).10

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10,* 355–371.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (Ed.), *Validation in language assessment*. 41–60. Routledge.

Read, J. (2000). *Assessing vocabulary.* Cambridge University Press.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics, 24,* 146–161. https://doi.org/10.1017/s0267190504000078

Reynolds, B., Shih, Y., & Wu, W. (2018). Modeling Taiwanese adolescent learners' English vocabulary acquisition and retention: The washback effect of the College Entrance Examination Center's reference word list. *English for Specific Purposes. 52*. 47-59. 10.1016/j.esp.2018.08.001.

Reynolds, B., Wu, W., Liu, H., Kuo, S. & Yeh, C. (2015). Towards a Model of Advanced Learners' Vocabulary Acquisition: An Investigation of L2 Vocabulary Acquisition and Retention by Taiwanese English Majors. *Applied Linguistics Review*, *6*(1), 121-144. https://doi.org/10.1515/applirev-2015-0006

Rogers, J., Webb, S., & Nakata, T. (2015). Do the cognacy characteristics of loanwords make them more easily learnt than noncognates? *Language Teaching Research, 19*(1), 9–27. https://doi.org/10.1177/1362168814541752

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*(4), 589–619. https://doi.org/10.1017/S0272263199004039

Schmid. H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.

Schmidt, R.W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 129-158. https://doi.org/10.1093/applin/11.2.129

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning, 64*(4), 913–951. https://doi.org/10.1111/lang.12077

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484–503. Cambridge University Press. https://doi.org/10.1017/S0261444812000018

Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *Modern Language Journal, 95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Schmitt, D. and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language*

Testing 18(1). 55-88.

Shin, D., Chon, Y. V., and Kim, H. (2011). Receptive and productive vocabulary sizes of high school learners: What next for the basic word list? *English Teaching, 66*(3), 123-148. http://dx.doi.org/10.15858/engtea.66.3.201109.123

Sinclair, J. (2005). Corpus and text: Basic principles, in M. Wynne (ed.) *Developing Linguistic Corpora: A guide to good practice.* Oxbow Books, 1-16.

Skoufaki, S. & Petrić, B. (2021). Exploring polysemy in the Academic Vocabulary List: A lexicographic approach. J*ournal of English for Academic Purposes 54*. https://doi.org/10.1016/j.jeap.2021.101038

Skoufaki, S., & Petrić, B. (2021b). Academic vocabulary in an EAP course: Opportunities for incidental learning from printed teaching materials developed in-house. English for Specific Purposes, 63, 71–85. https://doi.org/10.1016/j.esp.2021.03.002

Soars, L., & Soars, J. (2005). *New headway student's book* (Upper–Intermediate). Oxford University Press.

Stoeckel, T., Ishii, T., & Bennett, P. (2020). Is the lemma more appropriate than the flemma as a word counting unit? In *Applied Linguistics. Vol. 41, Issue 4,* pp. 601–606). Oxford University Press. https://doi.org/10.1093/applin/amy059

Sun, Y., & Dang, T. N. Y. (2020). Vocabulary in high-school EFL textbooks: Texts and learner knowledge. *System, 93*. https://doi.org/10.1016/j.system.2020.102279

Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass and C. Madden (eds), *Input in Second Language Acquisition*. 235-253. Rowley, MA: Newbury House.

Swain, M. (1995) Three functions of output in second language learning. In G. Cook and B. Seidelhofer (eds) *Principle and Practice in Applied Linguistics: Studies in Honour of H.G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research.* Routledge.

Teng, F. (2014). Strategies for teaching and learning vocabulary. *Beyond Words, 2*(2), 40–56. http://jurnal.wima.ac.id/index.php/BW/article/view/592

Teng, F. (2015). The Effectiveness of Extensive Reading on EFL Learners' Vocabulary Learning: Incidental versus Intentional Learning. *BELT – Brazilian English Language Teaching Journal, 6*(1), 66. https://doi.org/10.15448/2178-3640.2015.1.20033

Teng, M., & Xu, J. (2022). Pushing vocabulary knowledge from receptive to productive mastery: Effects of task type and repetition frequency. *Language Teaching Research, 0*(0). https://doi.org/10.1177/13621688221077028

Teng, Mark Feng. (2016). The Effects of Word Exposure Frequency on Incidental Learning of the Depth of Vocabulary Knowledge. *Gema Online Journal of Language Studies. 16.* 53-70. 10.17576/gema-2016-1603-04

The Ministry of Education, Culture, Sport, Science and Technology. (2018). Senior high school curriculum guidelines section 13 English. Available at: 20230328-mxt_kyoiku02_100014466_002.pdf

The Ministry of Education, Culture, Sports, Science and Technology of Japan. (2018). High school curriculum guidelines Chapter 2 Section 13 English. https://www.mext.go.jp/content/20230328-mxt_kyoiku02_100014466_002.pdf

Thornbury, S. (2002). *How to teach vocabulary.* Pearson Education Limited.

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System, 21*(3), 371–380.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research, 13*(2), 138–163.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work.* Amsterdam and Philadelphia: John Benjamins.

Tokyo Metropolitan Board of Education. (2021). Results of textbook selection by subject for metropolitan high schools and metropolitan secondary schools used in 2022. Retrieved from: https://www.kyoiku.metro.tokyo.lg.jp/press/press_release/2021/files/release20210826_01/reference.pdf

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The Effects of Repetition on Incidental Vocabulary Learning: A Meta-Analysis of Correlational Studies. *Language Learning, 69*(3), 559–599. https://doi.org/10.1111/lang.12343

Urdaniz, R. P., & Skoufaki, S. (2022). Spanish L1 EFL learners' recognition knowledge of English academic vocabulary: The role of cognateness, word frequency and length. *Applied Linguistics Review, 13*(4), 661–703.

https://doi.org/10.1515/applirev-2018-0109

van Zeeland, H. and Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*(3), pp. 609–624. http://dx.doi.org/10.1016/j.system.2013.07.012

Vidal, K. (2003). Academic Listening: A Source of Vocabulary Acquisition? Applied Linguistics, 24, 56-89.

Vidal, K. (2011). A Comparison of the Effects of Reading and Listening on Incidental Vocabulary Acquisition. *Language Learning, 61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Vilkaitė-Lozdienė, L., & Schmitt, N. (2020). Frequency as a guide for vocabulary usefulness. In S. Webb. (Eds.), *The Routledge Handbook of Vocabulary Studies,* 81–96. Routledge. https://doi.org/10.4324/9780429291586-6

Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System, 37*(3), 461–469. https://doi.org/10.1016/j.system.2009.01.004

Waring, N., & Nation, I. S. P. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (2011), *Vocabulary: Description, acquisition and pedagogy (11th printing)* (pp.6-19). Cambridge University Press.

Waring, R. (1997). A study of receptive and productive learning from word cards. *Studies in Foreign Languages and Literature (Notre Dame Seishin University, Okayama), 21*(1), 94–114.

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*(2). http://nflrc.hawaii.edu/rfl

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition, 27*(1), 33–52. https://doi.org/10.1017/S0272263105050023

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition, 30*(1), 79–95. https://doi.org/10.1017/S0272263108080042

Webb, S. (2008b). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2). http://nflrc.hawaii.edu/rfl

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.) *The*

Routledge Handbook of Vocabulary Studies, 225-239. Routledge. https://doi.org/10.4324/9780429291586-6

Webb, S. A. (2002). Investigating the effects of learning tasks on vocabulary knowledge. [Unpublished PhD Thesis]. Victoria University of Wellington, Wellington, New Zealand. file:///C:/Users/hiran/Downloads/thesis.pdf

Webb, S. A., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal, 43*(1), 113–126.

Webb, S., & Chang, A. C.-S. (2015). Second language vocabulary learning through extensive reading: how does frequency and distribution of occurrence affect learning? *Language Teaching Research, 19*(6), 667–686.

Webb, S., & Nation, P. (2017). *How vocabulary is learnt.* Oxford University Press.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test. *ITL - International Journal of Applied Linguistics, 168*(1)*,* 33–69. https://doi.org/10.1075/itl.168.1.02web

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, *56*(2), 161–180. doi:10.1017/S0261444822000507

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary learning activities? A meta-analysis. *Modern Language Journal, 104*(4), 715-738.

West, M. (1953). *A general service list of English Words.* Longman, Green & Co.

Xinxin, C., Xiaolan, L., & Ahmed, M. (2022). Relationship of the language distance to English ability of a country. doi:10.48550/arXiv.2211.07855

Yanagisawa, A., & Webb, S. (2020). Measuring depth of vocabulary knowledge. In S. Webb. (Eds.), *The Routledge Handbook of Vocabulary Studies,* 371–386. https://doi.org/10.4324/9780429291586-24

Yang, L., & Coxhead, A. (2022). A Corpus-based Study of Vocabulary in the New Concept English Textbook Series. *RELC Journal, 53(*3), 597–611. https://doi.org/10.1177/0033688220964162

Zhang, P. (2022). How does repetition affect vocabulary learning through listening to the teacher's explicit instruction? The moderating role of listening proficiency and preexisting vocabulary knowledge. *Language Teaching Research, 0*(0). https://doi.org/10.1177/13621688221140521

Zhang, S., & Zhang, X. (2020). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research, 26*(4), 696–725.

https://doi.org/10.1177/1362168820913998

Zimmerman, C. B. (2014). Teaching and learning vocabulary for second language learners. In M. CelceMurcia, D. M. Brinton, & M. A. Snow (Eds.), *Teaching English as a second or foreign language (4th ed.),* 288-302. National Geographic Learning.

Zipf, G. (1936). The Psychobiology of Language. London: Routledge.

Zipf, G. (1949). Human Behavior and the Principle of Least Effort. New York: Addison-Wesley.

Appendix

The lists of newly-introduced lemmas used in this study

All Aboard! 1

bathroom_n
bedroom_n
bookshelf_n
cheerful_adj
cupboard_n
dad_n
drama_n
entrance_n
fast-food_adj
Iceland_n
mathematics_n
shy_adj
talkative_adj
thick_adj
vase_n
yard_n
actress_n
admire_v
aggressively_adv
amazing_adj
athlete_n
attractive_adj
beautifully_adv
dancer_n
fantastic_adj
focus_v
good-looking_adj
gracefully_adv
Hero_n
passionately_adv
pose_v
powerful_adj
professional_adj
talent_n
toast_n
tough_adj
truly_adv
actually_adv
adult_n
anime_n
annual_adj
aquarium_n
character_n
clothing_n
crane_n
crepe_n
defence_n
experience_v
Manga_n
method_n
object_n
raw_adj
similar_adj
vinegared_adj
appear_v
castle_n
championship_n
flat_n
fridge_n
highly_adv
Miracle_n
pond_n
rainwater_n
rarely_adv
recommended_adj
reflect_v
scene_n
shallow_adj
snowfield_n
Spanish_n
speechless_adj
splendid_adj
surface_n
agricultural_adj
airplane_n
although_con
approach_v
attach_v
attend_v
ballet_n
behind_con
cattle_n
chase_v
coach_n
costume_n
cow_n
dairy_n
death_n
delivery_n
eel_n
feed_v
firefighter_n
hammer_n
health_n
hobby_n

lend_v
living_adj
nail_n
owner_n
pay_v
pig_n
pilot_n
realize_v
record_v
responsibility_n
scratch_n
shake_v
smell_v
specific_adj
stingy_adj
thief_n
till_con
translator_n
trumpet_n
value_n
veterinarian_n
amazed_adj
awesome_adj
belong_v
Blog_n
breathtaking_adj
create_v
else_adv
funny_adj
historical_adj
humorous_adj
monster_n
notice_v
period_n
pioneer_n
Portuguese_n
print_n
above_adv
Antarctic_n
decrease_v
dolphin_n
ecofriendly_adj
enough_adj
extinct_adj
female_n
hardship_n
krill_n
male_n
marine_adj
melt_v
penguin_n
protect_v
recycle_v
therefore_adv
upon_con
astronaut_n
boring_adj
commander_n
communicate_v
control_n
crew_n
delay_v
develop_v
difficulty_n
dinnertime_n
discount_n
duty_n
enhance_v
enthusiasm_n
experiment_n
express_v
opportunity_n
overcome_v
pianist_n
regularly_adv
relax_v
repair_v
scarf_n
scientific_adj
senior_n
separately_adv
share_v
success_n
supply_n
teamwork_n
temperature_n
toward_con
trust_n
valuable_adj
various_adj
worker_n
battle_n
broaden_v
cross_v
deepen_v
effort_n
exchange_v
homeland_n
homestay_n
hunger_n
importance_n
Japanese-American_n
journey_n
local_adj
moreover_adv
rapid_adj
reach_v
relationship_n

result_n
several_adj
situation_n
soldier_n
source_n
subway_n
successful_adj
suffer_v
survive_v
terrible_adj
track_n
unique_adj
account_n
alive_adj
appoint_v
attack_v
basket_n
belief_n
billionaire_n
brave_adj
bright_adj
broadcast_v
cabinet_n
cheer_v
claim_v
continue_v
courage_n
destination_n
discuss_v
donate_v
education_n
elephant_n
empty_adj
equality_n
fear_n
flag_n
frog_n
graffiti_n
headquarters_n
iced_adj
increase_v
injure_v
lottery_n
manner_n
meadow_n
miraculous_adj
onto_con
pitcher_n
pollution_n
poster_n
prefecture_n
president_n
pull_v
receive_v
recovery_n
resident_n
salary_n
scared_adj
seriously_adv
shine_v
shortage_n
threat_n
toad_n
turtle_n
wave_v
(268 words)

All Aboard! 2
boxer_n
carnival_n
colorful_adj
compete_v
crowded_adj
earn_v
elegant_adj
enjoyable_adj
fantastic_adj
firework_n
futsal_n
goldfish_n
mysterious_adj
old-fashioned_adj
presentation_n
samba_n
scoop_v
smiling_adj
unfashionable_adj
unfortunately_adv
annoyed_adj
ashamed_adj
backache_n
bold_adj
bored_adj
brave_adj
calm_adj
careful_adj
childhood_n
chilly_adj
comfortable_adj
confident_adj
confused_adj
contest_n
cough_n
cowardly_adj
dancer_n
delighted_adj
depressed_adj
diligently_adv
dissatisfied_adj
dizzy_adj
dull_adj
dynamic_adj
ecstatic_adj
embarrassed_adj
envious_adj
excited_adj
exhausted_adj
express_v
fever_n
flu_n
freedom_n
frustrated_adj
furious_adj
gradually_adv
graduate_v
headache_n
impatient_adj
inspiration_n
itchy_adj
jealous_adj
joke_n
leader_n
lonely_adj
machine_n
mad_adj
matter_n
meal_n
medal_n
medicine_n
millionaire_n
miserable_adj
motto_n
naturally_adv
nervous_adj
overseas_adj
patient_adj
personality_n
positively_adv
proud_adj
quietly_adv
relaxed_adj
rest_v
restless_adj
rhythm_n
risk_n
satisfied_adj
scared_adj
sore_adj
stomachache_n
surprised_adj
thirsty_adj
throat_n
timid_adj
toothache_n
trampolinist_n
uncomfortable_adj
unique_adj
upset_adj
website_n
active_adj
boyfriend_n
centimeter_n
checkout_n
deforestation_n
disappear_v

emission_n
fluffy_adj
fur_n
gram_n
hunting_n
illegal_adj
killifish_n
length_n
lifestyle_n
nowadays_adv
overhunting_n
pollution_n
precious_adj
protected_adj
rare_adj
reduce_v
rhinoceros_n
serve_v
shocking_adj
solution_n
species_n
spirit_n
tail_n
watermelon_n
weigh_v
apron_n
attention_n
autograph_n
blackboard_n
blanket_n
blazer_n
chalk_n
clearly_adv
comb_n
creepy_adj
daily_adj
dishcloth_n
expression_n
facial_adj
faucet_n
greet_v
heal_v
helmet_n
homeroom_n
housework_n
humanoid_adj
imaginary_adj
instead_adv
interact_v
novelist_n
outlet_n
pajamas_n
pepper_n
pillow_n
platform_n
pour_v
preparation_n
probably_adv
reality_n
rescue_n
resident_n
robot_n
ruler_n
sausage_n
schoolbag_n
scissors_n
sink_n
slipper_n
sprinkle_v
stapler_n
teapot_n
tie_n
toilet_n
tone_n
towel_n
triangle_n
weapon_n
add_v
ahead_adv
ancient_adj
ankle_n
anybody_pron
audience_n
backpack_n
basically_adv
blazer_n
bracelet_n
breathe_v
butter_n
buttock_n
calf_n
calmly_adv
cardigan_n
central_adj
cheek_n
chest_n
chin_n
clown_n
contrary_n
cousin_n
crouch_v
darkness_n
elbow_n
era_n
evil_n
eyebrow_n
eyelash_n
faint_adj

fingernail_n
firefly_n
forefinger_n
forehead_n
form_n
glove_n
golf_n
gum_n
handkerchief_n
heavily_adv
heel_n
hide_v
hip_n
impact_n
iris_n
justice_n
Kabuki_n
lantern_n
lightly_adv
loafer_n
lonesome_adj
main_adj
makeup_n
merchant_n
moat_n
moment_n
muscle_n
necklace_n
nobody_pron
novel_n
ocean_n
originate_v
palm_n
pant_n
perform_v
performance_n
principal_n
publisher_n
pupil_n
represent_v
reveal_v
ribbon_n
rise_v
robber_n
row_n
scare_v
scary_adj
scream_v
shin_n
shoulder_n
skirt_n
sleeve_n
slope_n
sneaker_n
sob_v
stomach_n
streetlamp_n
stroke_v
sunset_n
superhero_n
surprise_v
surround_v
thigh_n
thumb_n
tickle_v
tie_n
toe_n
tongue_n
vest_n
villain_n
weep_v
worldwide_adv
wrist_n
amount_n
appreciate_v
atmosphere_n
bleach_v
consist_v
coral_n
crisis_n
decade_n
electricity_n
erosion_n
explore_v
fascinating_adj
fuel_n
generation_n
Including_con
Italian_adj
microcosm_n
opinion_n
optional_adj
overnight_adv
phenomenon_n
pollution_n
polyp_n
prevent_v
reduce_v
reef_n
refuse_v
scientist_n
shell_n
stamp_n
summarize_v
surcharge_n
valley_n
wonder_v
accountant_n

attract_v
author_n
babysitter_n
baker_n
barber_n
beg_v
broadcaster_n
businessperson_n
butcher_n
chef_n
clerk_n
confectioner_n
customer_n
dentist_n
dressmaker_n
earthquake_n
economy_n
editor_n
fisher_n
hairdresser_n
hairstylist_n
happiness_n
hometown_n
hula_n
interpreter_n
judge_n
local_n
mechanic_n
occupation_n
pharmacist_n
photographer_n
politician_n
postal_adj
professor_n
programmer_n
prosecutor_n
purpose_n
receptionist_n
related_adj
reporter_n
restoration_n
salesclerk_n
secretary_n
sever_n
shy_adj
sightseeing_n
stimulate_v
struggle_v
tailor_n
technical_adj
temporary_adj
tourist_n
toward_con
transit_v
tropical_adj
wedding_n
appealing_adj
beat_v
beautifully_adv
cathedral_n
charm_v
couple_n
deeply_adv
depart_v
downtown_adj
film_v
flavor_n
forward_adv
gentle_adj
Halloween_n
haunted_adj
honey-colored_adj
illumination_n
knock_v
limestone_n
location_n
lovely_adj
magical_adj
parade_n
proud_adj
recommend_v
remind_v
safely_adv
scenery_n
series_n
according to_con
allow_v
alternative_adj
annually_adv
average_n
base_v
bitter_adj
buyer_n
cacao_n
cheese_n
choice_n
condition_n
consumer_n
data_n
easily_adv
everyday_adj
exam_n
figure_n
guarantee_v
import_v
improve_v
income_n
issue_n

kilogram_n
label_v
labor_n
mark_n
minimum_adj
mostly_adv
offer_v
partnership_n
percentage_n
physical_adj
poverty_n
price_n
product_n
solution_n
soybean_n
spaceship_n
spaghetti_n
superman_n
taste_n
though_con
ton_n
trading_n
truth_n
volume_n
wise_adj
academy_n
aloud_adv
assistance_n
bamboo_n
bat_n
bathtub_n
behavior_n
bite_v
brush_v
cafeteria_n
certain_adj
cheek_n
cheerfully_adv
cherish_v
childhood_n
confidence_n
cost_v
crosswalk_n
curly_adj
design_n
design-based_adj
designer_n
detergent_n
device_n
dollar_n
donation_n
ease_n
economical_adj
eggplant_n
expect_v
explode_v
fair_adj
fate_n
first-grade_adj
fork_n
freely_adv
front_n
gain_v
garlic_n
gaze_v
girlfriend_n
graduation_n
ill_adj
improvement_n
industry_n
iron_adj
kettle_n
knit_v
lamp_n
landmine_n
lunchtime_n
might_mod
microwave_n
miserable_adj
mistake_n
misbehave_v
nap_n
naughty_adj
object_n
ordinary_adj
paper-made_adj
peaceful_adj
pedestrian_n
perhaps_adv
pointer_n
pot_n
potholder_n
prepare_v
produce_v
projector_n
puppy_n
quietly_adv
refrigerator_n
refugee_n
roll_v
rug_n
sauna_n
select_v
serious_adj
shelf_n
shelter_n
smartphone_n
sniff_v

solve_v
soothe_v
sponge_n
spoon_n
stray_adj
suit_n
surf_v
sweat_v
tear_n
therapy_n
tight_adv
tool_n
toothbrush_n
toothpaste_n
tray_n
trickle_v
troublemaker_n
universe_n
unwell_adj
usual_adj
vacant_adj
veterinarian_n
victim_n
wallet_n
whole_adj
wind-powered_adj
wipe_v
(566 words)

All Aboard! 3
Antarctica_n
bill_n
capybara_n
center_n
circle_n
complete_v
discover_v
discovery_n
diver_n
duck_n
inspect_v
mammal_n
mystery_n
newly_adv
ocean_n
pattern_n
pelican_n
platypus_n
resemble_v
rodent_n
round-shaped_adj
seem_v
shining_adj
spot_n
type_n
unusual_adj
wingspan_n
surprisingly_adv
though_con
curved_adj
motif_n
characteristic_n
creativity_n
masterpiece_n
lizard_n
warmly_adv
fountain_n
stair_n
wavy_adj
artwork_n
beauty_n
sculptor_n
futuristic_adj
maximum_n
huge_adj
mausoleum_n
marble_n
hang_v
bronze_n
medalist_n
congratulation_n
coincidence_n
currency_n
European_adj
Union_n
despite_con
cultural_adj
diversity_n
euro_n
common_adj
unite_v
trade_v
image_n
symbolize_v
unity_n
contrast_n
Austrian_adj
Finnish_adj
purchase_v
prefer_v
Canadian_adj
yen_n
reverse_adj
Beaver_n
staple_n
identity_n
addition_n
antique_adj
archaeological_adj
armor_n
arrival_n
artifact_n
atomic_adj
atomic_adj
battlefield_n
bomb_n
bomb_n
bombing_n
Buddhist_n
burn_n
close_adj
clothing_n
consider_v
destroy_v
disaster_n
encourage_v
evacuate_v
exhibit_n
fade_v
feature_v
important_adj
inhabitant_n
memorial_adj
memory_n
missing_adj
operate_v
passenger_n

passenger_n
prefectural_adj
repeatedly_adv
replace_v
resume_v
ruin_n
service_n
standstill_n
statue_n
streetcar_n
surviving_adj
swell_v
tag_n
vital_adj
western_adj
wish_v
witness_n
anyway_adv
apology_n
arithmetic_n
attic_n
awfully_adv
bedroom_n
blame_v
blog_n
clever_adj
comfort_v
contact_v
county_n
depressed_adj
diary_n
differently_adv
disappointed_adj
disease_n
dusty_adj
except_con
flash_v
fraction_n
gently_adv
geography_n
handwritten_adj
hate_v
inspector_n
lie_n
light_v
lonely_adj
long-term_adj
loudly_adv
mama_n
mechanical_adj
medical_adj
message_n
mind_n
neighborhood_n
part-time_adj
persuade_v
playground_n
pray_v
prefer_v
reader_n
receiver_n
reply_v
request_v
round_adj
score_n
search_n
sender_n
sign_n
stick_v
stupid_n
suit_v
suppose_v
telebook_n
text_n
typical_adj
whistle_v
accidentally_adv
accompany_v
achievement_n
affect_v
anywhere_adv
appreciate_v
awareness_n
ban_v
butterfly_n
cargo_n
central_adj
cicada_n
connect_v
conserve_v
continent_n
damage_v
designation_n
dolphin_n
dragonfly_n
due to_con
ecosystem_n
evolution_n
evolve_v
explore_v
follow_v
further_adj
green_adj
guide_n
indigenous_adj
insect_n
invade_v
invasion_n

isolated_adj
lizard_n
locate_v
mud_n
negatively_adv
nonnative_adj
process_n
seed_n
site_n
topic_n
tourism_n
treasure_n
visitor_n
vulnerable_adj
whale_n
wipe_v
adventure_n
afterlife_n
army_n
bone_n
bury_v
charcoal_n
clay_n
clothing_n
contain_v
contribute_v
designate_v
dig_v
emperor_n
empire_n
ethnic_adj
exactly_adv
foundation_n
fully_adv
greatness_n
hairstyle_n
Inca_adj
include_v
including_con
investigation_n
jewel_n
kingdom_n
life-size_adj
magnificent_adj
mainly_adv
mascot_n
mineral_n
modern_adj
palace_n
Peruvian_adj
possibility_n
purple_n
replica_n
research_n
researcher_n
standardize_v
starry_adj
structure_n
terracotta_n
tomb_n
unify_v
warrior_n
well_n
whole_adj
approach_n
available_adj
brand_n
cent_n
chemical_adj
collapse_v
collection_n
community_n
cotton_n
damage_v
developing_adj
dye_n
enable_v
entrepreneur_n
ethical_adj
fabric_n
factory_n
fair_adj
handmade_adj
harmful_adj
highlight_v
human_n
incident_n
instance_n
item_n
liter_n
mass-produce_v
material_n
organic_adj
per_con
provide_v
recently_adv
recycled_adj
respect_v
right_n
sew_v
skill_n
skillful_adj
skillfully_adv
strongly_adv
tailor_n
trap_v
trend_n
vintage_adj

wage_n
waste_n
worker_n
working_adj
abolish_v
accident_n
aim_v
biomass_n
carbon-neutral_adj
challenge_n
coal_n
comparison_n
consume_v
consumption_n
continually_adv
eruption_n
feature_n
fossil_n
fuel_n
generate_v
geographical_adj
geothermal_adj
graph_n
hydropower_n
investment_n
limited_adj
manage_v
military_adj
nearly_adv
non-renewable_n
nuclear_adj
proportion_n
region_n
renewable_adj
resource_n
restart_v
slightly_adv
social_adj
solar_adj
spending_n
sunlight_n
total_adj
Volcanic_adj
volcano_n
absence_n
absently_adv
accept_v
achieve_v
administration_n
aware_adj
barrier_n
beach_n
budding_adj
Challenger_n
convey_v
crowd_n
crowded_adj
crown_n
debut_n
desperately_adv
destiny_n
determine_v
development_n
direction_n
discrimination_n
drag_n
drowsy_adj
edge_n
entirely_adv
equal_adj
equally_adv
excitement_n
exhausted_adj
familiar_adj
federal_adj
foolishly_adv
force_v
forehead_n
frankly_adv
government_n
grader_n
grant_v
honestly_adv
hungrily_adv
immediately_adv
integration_n
interrupt_v
kindergarten_n
link_n
literature_n
major_v
marshal_n
mention_v
movement_n
neglect_v
neither_con
novel_n
opposite_adj
paragraph_n
passion_n
Pip_n
portrait_n
positive_adj
prejudice_n
previous_adj
racial_adj
reading_n
regrettably_adv

retire_v
rob_v
romantic_adj
segregation_n
sensation_n
sentence_n
settle_v
sightseeing_n
skin_n
sleepless_adj
slip_v
society_n
somebody_pron
spit_v
steal_v
strangely_adv
subway_n
swear_v
tear_v
throughout_con
tip_n
toe_n
tragic_adj
treat_v
unusual_adj
used_adj
waiter_n
whatever_x
wine_n
within_con
yell_v
(462 words)

Taiwanese textbook 1
nervous_adj
pound_v
communicate_v
communication_n
hall_n
medal_n
costume_n
role_n
drama_n
dramatic_adj
wonder_v
wonder_n
chat_v
chat_n
principal_n
schedule_n
schedule_v
personal_adj
belonging_n
confused_adj
confuse_v
confusion_n
suddenly_adv
sudden_adj
seem_v
obvious_adj
reply_v
reply_n
interrupt_v
interruption_n
plenty_n
plentiful_adj
remind_v
forgiveness_n
forgive_v
deliver_v
delivery_n
community_n
roll_v
edge_n
shoot_v
smooth_adj
smooth_v
grasp_v
grasp_n
slip_v
crash_n
crash_v
frightened_adj
fright_n
escape_v
escape_n
guilty_adj
guilt_n
greet_v
greeting_n
comfortable_adj
comfort_n
comfort_v
calculate_v
calculation_n
repair_v
repair_n
creep_v
relieved_adj
relieve_v
relief_n
straight_adv
straight_adj
return_v
return_n
receive_v
discover_v
discovery_n
invent_v
invention_n
inventor_n
saying_n
necessity_n
necessary_adj
solution_n
solve_v
explain_v
explanation_n
create_v
creation_n
creator_n
mechanic_n
mechanical_adj
design_v
design_n
designer_n
automatically_adv
automatic_adj
control_v
control_n
device_n
represent_v
representative_n
exhibition_n
exhibit_v
international_adj
brilliant_adj
successful_adj
success_n
succeed_v
result_n

result_v
failed_adj
fail_v
failure_n
experiment_n
experiment_v
experimental_adj
chemical_n
chemical_adj
expectation_n
expect_v
disappointed_adj
disappoint_v
disappointment_n
instead_adv
convince_v
frustrate_v
frustration_n
damage_v
damage_n
spread_v
spread_n
colorful_adj
desire_n
desire_v
desirable_adj
diet_n
mood_n
influential_adj
influence_v
influence_n
imagine_v
imagination_n
describe_v
description_n
ordinary_adj
besides_adv
besides_con
apply_v
application_n
daily_adj
attract_v
attraction_n
indicate_v
indication_n
warning_n
warn_v
affect_v
appetite_n
probably_adv
probable_adj
environment_n
environmental_adj
decorate_v
decoration_n
increase_v
increase_n
natural_adj
nature_n
raise_v
spirit_n
spiritual_adj
spoil_v
aware_adj
routine_n
routine_adj
recipe_n
friendship_n
friendly_adj
chef_n
satisfying_adj
satisfy_v
pour_v
similar_adj
similarity_n
mix_v
mix_n
compliment_n
compliment_v
support_n
support_v
trust_n
trust_v
hug_n
hug_v
encouragement_n
encourage_v
properly_adv
proper_adj
poem_n
poet_n
effort_n
tasty_adj
require_v
requirement_n
ingredient_n
fond_adj
stove_n
therefore_adv
worth_adj
worth_n
local_adj
local_n
nod_v
frown_v
throughout_con
gesture_n
vary_v

various_adj
variety_n
positive_adj
traveler_n
travel_v
travel_n
culture_n
cultural_adj
express_v
expression_n
expressive_adj
approval_n
approve_v
tilt_v
disagreement_n
disagree_v
threat_n
threaten_v
offer_v
offer_n
palm_n
insulting_adj
insult_v
insult_n
criminal_n
criminal_adj
punishment_n
punish_v
tradition_n
traditional_adj
impolite_adj
polite_adj
outward_adv
rescue_n
rescue_v
negative_adj
realize_v
dessert_n
item_n
perform_v
performance_n
crisp_adj
absorb_v
moisture_n
method_n
powder_n
impress_v
impressive_adj
impression_n
stain_n
stain_v
remove_v
bare_adj
stubborn_adj
effectively_adv
effective_adj
effect_n
gently_adv
gentle_adj
amazingly_adv
amazing_adj
amaze_v
amazement_n
situation_n
squeeze_v
rub_v
gradually_adv
gradual_adj
secret_n
secret_adj
substance_n
content_n
laboratory_n
located_adj
locate_v
tribe_n
nearby_adv
legend_n
generation_n
century_n
seriously_adv
serious_adj
decrease_v
decrease_n
volunteer_v
volunteer_n
destroy_v
destruction_n
tear_v
fierce_adj
foolish_adj
fool_v
fool_n
greed_n
greedy_adj
empty_v
empty_adj
hunger_n
shock_v
shocked_adj
embarrassed_adj
embarrass_v
embarrassment_n
elder_n
elderly_adj
decision_n
avoid_v
extinct_adj

harmony_n
resource_n
diary_n
engaged_adj
engagement_n
scream_v
scream_n
immediately_adv
immediate_adj
familiar_adj
origin_n
original_adj
custom_n
evil_adj
evil_n
attend_v
ceremony_n
superstition_n
superstitious_adj
belief_n
event_n
prevent_v
prevention_n
related_adj
relate_v
relation_n
regard_v
regarding_con
scientific_adj
scientist_n
proof_n
prove_v
actually_adv
actual_adj
tend_v
tendency_n
associate_v
association_n
sneeze_v
nowadays_adv
artificial_adj
predict_v
virtue_n
honesty_n
honest_adj
fellow_adj
policy_n
error_n
admit_v
admission_n
deceive_v
irresponsible_adj
responsible_adj
responsibility_n
involved_adj
involve_v
involvement_n
manage_v
management_n
recently_adv
recent_adj
project_n
project_v
reveal_v
confident_adj
confidence_n
consequence_n
consequent_adj
convenient_adj
convenience_n
contrast_n
contrast_v
consider_v
consideration_n
reaction_n
react_v
maintain_v
maintenance_n
protective_adj
protect_v
protection_n
circumstance_n
harm_n
harm_v
determine_v
determination_n
specific_adj
motive_n
motivate_v
expose_v
ability_n
heal_v
percent_n
percentage_n
emotion_n
emotional_adj
mental_adj
attitude_n
weaken_v
focus_v
focus_n
response_n
respond_v
laughter_n
claim_n
claim_v
relax_v
relaxation_n

muscle_n
complicated_adj
complicate_v
equipment_n
equip_v
afterward_adv
provide_v
rate_n
normal_n
normal_adj
internal_adj
production_n
produce_v
product_n
comedy_n
comedian_n
include_v
including_con
pain_n
painful_adj
reduction_n
reduce_v
enthusiastic_adj
enthusiasm_n
concept_n
colony_n
colonial_adj
politician_n
politics_n
political_adj
rural_adj
society_n
social_adj
remain_v
litter_n
litter_v
trail_n
achievement_n
achieve_v
valuable_adj
value_n
moreover_adv
degree_n
sort_v
sort_n
fund_n
fund_v
worthless_adj
gain_v
advantage_n
handle_v
career_n
pursue_v
pursuit_n
benefit_n
benefit_v
beneficial_adj
absolutely_adv
absolute_adj
explore_v
(491 words)

Taiwanese textbook 2
characteristic_n
characteristic_adj
character_n
clumsy_adj
loyalty_n
loyal_adj
violence_n
violent_adj
cruelty_n
cruel_adj
authority_n
likewise_adv
choice_n
journalist_n
journalism_n
customer_n
mall_n
herd_n
useless_adj
apparently_adv
apparent_adj
fortune_n
outstanding_adj
appreciate_v
appreciation_n
audience_n
insect_n
coward_n
dare_v
diligence_n
diligent_adj
additionally_adv
additional_adj
addition_n
flock_n
flock_v
spot_n
indeed_adv
apologize_v
apology_n
delay_v
delay_n
continuously_adv
continuous_adj
continual_adj
ignore_v
furious_adj
crush_n
crush_v
separate_v
separate_adj
separation_n
download_v
dependence_n
technological_adj
technology_n
luxury_n
luxurious_adj
addict_n
addiction_n
psychological_adj
psychology_n
psychologist_n
term_n
isolated_adj
isolate_v
isolation_n
access_n
access_v
contact_n
contact_v
emergency_n
promptly_adv
prompt_adj
suffer_v
frequently_adv
frequent_adj
message_n
messenger_n
insist_v
clever_adj
concentrate_v
concentration_n
praise_v
praise_n
tough_adj
overcome_v
pile_n
pile_v
childhood_n
lung_n
infection_n
infect_v
theme_n
vehicle_n
military_adj
vessel_n
typical_adj
reflect_v
reflection_n
reality_n
realistic_adj
wicked_adj
progress_v
progress_n
tension_n
tense_adj

complexity_n
complex_adj
unique_adj
critic_n
criticize_v
critical_adj
alike_adv
alike_adj
feature_n
feature_v
imaginative_adj
imaginary_adj
awe_n
awe_v
tip_n
improve_v
improvement_n
memory_n
memorize_v
develop_v
development_n
information_n
inform_v
informative_adj
assist_v
assistance_n
assistant_n
goal_n
particular_adj
active_adj
activity_n
challenge_n
challenge_v
challenging_adj
input_n
input_v
brain_n
recognize_v
recognition_n
arrange_v
arrangement_n
phrase_n
process_n
mention_v
mention_n
organize_v
organization_n
recall_v
occasionally_adv
occasional_adj
musical_adj
musical_n
instrument_n
strengthen_v
strength_n
connection_n
connect_v
concerning_con
concern_v
blame_v
blame_n
mess_n
mess_v
tremble_v
tremble_n
rage_n
rage_v
nag_v
stressed_adj
stress_v
stress_n
silence_n
silent_adj
channel_n
complain_v
complaint_n
expert_n
expert_adj
frankly_adv
frank_adj
theory_n
clip_n
humorously_adv
humorous_adj
humor_n
gender_n
scan_n
scan_v
sex_n
sexual_adj
quarrel_n
quarrel_v
unconsciously_adv
unconscious_adj
conscious_adj
surf_v
tidy_v
tidy_adj
strike_v
fortunately_adv
fortunate_adj
attempt_v
attempt_n
gap_n
roast_adj
roast_v
acceptable_adj
accept_v

acceptance_n
issue_n
factor_n
major_adj
geography_n
geographical_adj
essential_adj
essential_n
contain_v
container_n
gather_v
gathering_n
ancient_adj
wealth_n
wealthy_adj
satisfaction_n
satisfactory_adj
insure_v
insurance_n
prosperous_adj
prosper_v
prosperity_n
crop_n
crop_v
flourish_v
climate_n
average_adj
average_n
average_v
temperature_n
classic_adj
classic_n
staple_n
adapt_v
understandable_adj
opportunity_n
press_n
freedom_n
democratic_adj
democracy_n
significant_adj
significance_n
government_n
govern_v
individual_adj
individual_n
publish_v
publisher_n
detail_n
detail_v
private_adj
privacy_n
illustrate_v
illustration_n
poverty_n
struggle_n
struggle_v
pregnant_adj
pregnancy_n
urgent_adj
economic_adj
economy_n
nonetheless_adv
fame_n
overnight_adv
overnight_adj
sum_n
exclusive_adj
manufacturer_n
manufacture_v
manufacture_n
advertise_v
advertisement_n
discuss_v
discussion_n
possibility_n
curiosity_n
curious_adj
gossipy_adj
gossip_n
resident_n
lean_v
annoy_v
rough_adj
roughly_adv
roar_v
roar_n
awake_adj
awake_v
haunt_v
pride_n
suppress_v
conscience_n
mature_adj
maturity_n
incident_n
cabinet_n
remedy_n
calm_v
calm_adj
upset_adj
upset_v
soothe_v
scrape_v
scrape_n
drift_v
drift_n
hesitantly_adv

hesitate_v
hesitation_n
ungratefully_adv
gratefully_adv
grateful_adj
regretful_adj
regret_v
regret_n
involuntarily_adv
involuntary_adj
voluntary_adj
seize_v
portray_v
portrait_n
melt_v
drown_v
pollution_n
pollute_v
amount_n
population_n
rapidly_adv
rapid_adj
victim_n
fatal_adj
discharge_v
discharge_n
atmosphere_n
fuel_n
region_n
regional_adj
coastal_adj
coast_n
severe_adj
continent_n
continental_adj
property_n
estimate_v
estimate_n
billion_n
fertile_adj
fertilizer_n
barren_adj
desert_n
expand_v
expansion_n
starvation_n
starve_v
horrible_adj
furry_adj
fur_n
shortage_n
tragedy_n
tragic_adj
enviable_adj
envy_v
envy_n
opinion_n
official_adj
official_n
sample_v
sample_n
headquarters_n
appearance_n
texture_n
flavor_n
discipline_n
discipline_v
thorough_adj
preparation_n
attractive_adj
skip_v
plastic_n
metal_n
regular_adj
dull_v
dull_adj
swallow_v
swallow_n
introduce_v
introduction_n
lick_v
convey_v
ideal_adj
ideal_n
balance_n
balance_v
spit_v
sacrifice_n
sacrifice_v
skinny_adj
alternative_adj
alternative_n
delightful_adj
delight_n
delight_v
numerous_adj
research_n
research_v
researcher_n
impact_n
impact_v
grab_v
ridiculous_adj
inspiring_adj
inspire_v
inspiration_n
amusing_adj
amuse_v

amusement_n
visual_adj
conclude_v
conclusion_n
employ_v
employee_n
employment_n
skillful_adj
skill_n
considerable_adj
entertainment_n
entertain_v
entirely_adv
entire_adj
tolerate_v
tolerance_n
tolerant_adj
bulb_n
nevertheless_adv
willingness_n
willing_adj
accomplish_v
accomplishment_n
desperately_adv
desperate_adj
founder_n
found_v
foundation_n
signify_v
element_n
appeal_v
appeal_n
majority_n
novel_n
novelist_n
title_n
incredible_adj
grave_n
fascinate_v
creature_n
contrary_adj
contrary_n
monster_n
monstrous_adj
alarmed_adj
alarm_v
alarm_n
flee_v
horror_n
horrify_v
wander_v
hatred_n
murder_v
murder_n
murderer_n
being_n
depression_n
depress_v
otherwise_adv
miserable_adj
misery_n
request_n
request_v
promise_n
revenge_n
misfortune_n
disappear_v
(498 words)

Taiwanese textbook 3

chain_n
chain_v
battered_adj
batter_v
closure_n
chill_n
chill_v
chilly_adj
dim_adj
dim_v
strand_v
approach_v
approach_n
rob_v
robber_n
robbery_n
freezing_adj
freeze_v
flat_adj
aid_n
trunk_n
owe_v
reward_n
reward_v
repay_v
content_adj
content_v
contentment_n
shabby_adj
towel_n
erase_v
thoughtful_adj
slide_v
slide_n
blur_v
blur_n
vision_n
due_adj
whisper_v
whisper_n
mutual_adj
background_n
definitely_adv
definite_adj
crucial_adj
file_n
file_v
virtual_adj
clue_n
flesh_n
howl_v
howl_n
commit_v
suicide_n
bully_v
bully_n
meanwhile_adv
delete_v
account_n
switch_v
switch_n
site_n
site_v
nasty_adj
remark_n
remark_v
remarkable_adj
abusive_adj
abuse_v
abuse_n
torture_n
torture_v
comment_v
comment_n
efficient_adj
efficiency_n
agent_n
agency_n
violate_v
violation_n
procrastinate_v
procrastinator_n
procrastination_n
extremely_adv
extreme_adj
extreme_n
practical_adj
acquire_v
acquisition_n
statistics_n
statistical_adj
university_n
define_v
definition_n
chronic_adj
persistently_adv
persistent_adj
persist_v
assignment_n
assign_v
further_adv
overwhelm_v
suggest_v
suggestion_n
defeat_v
defeat_n
correspond_v

correspondence_n
identify_v
identification_n
priority_n
capable_adj
capability_n
budget_v
budget_n
sufficient_adj
adequate_adj
counselor_n
counsel_v
counsel_n
mild_adj
romance_n
romantic_adj
odds_n
pressure_n
pressure_v
pretend_v
optimistic_adj
optimism_n
cinema_n
awful_adj
gallery_n
energetic_adj
carefree_adj
lively_adj
awkward_adj
arrogant_adj
arrogance_n
insecure_adj
secure_adj
security_n
sympathetic_adj
sympathy_n
sympathize_v
yawn_v
advisable_adj
advise_v
advice_n
demand_v
demand_n
flow_v
flow_n
steady_adj
slight_adj
previous_adj
respect_v
respect_n
respectful_adj
sincerity_n
sincere_adj
dispute_n
dispute_v
launch_v
launch_n
immortal_adj
mortal_adj
mortal_n
quote_n
quote_v
kidnap_v
declare_v
declaration_n
passion_n
passionate_adj
jealous_adj
jealousy_n
mythology_n
conflict_n
conflict_v
erupt_v
eruption_n
refuse_v
refusal_n
judge_v
judge_n
judgment_n
command_v
command_n
bribe_v
bribe_n
wisdom_n
eventually_adv
eventual_adj
fulfill_v
fulfillment_n
ally_n
ally_v
alliance_n
conquer_v
conquest_n
cunning_adj
hollow_adj
haul_v
feast_v
feast_n
rejoice_v
attack_n
attack_v
destination_n
somewhat_adv
behavior_n
behave_v
offend_v
offense_n
offensive_adj

fundamental_adj
fundamental_n
verbally_adv
verbal_adj
upward_adv
upward_adj
journey_n
gratitude_n
religion_n
religious_adj
sacred_adj
precaution_n
caution_n
cautious_adj
conservative_adj
sensitive_adj
rashly_adv
rash_adj
display_v
display_n
affection_n
affectionate_adj
enormously_adv
enormous_adj
inappropriate_adj
appropriate_adj
funeral_n
fantastic_adj
genuinely_adv
genuine_adj
reverse_n
reverse_v
identity_n
vital_adj
survey_n
survey_v
conduct_v
conduct_n
prefer_v
preferable_adj
preference_n
commercial_adj
commercial_n
commerce_n
celebrity_n
endorsement_n
endorse_v
ambassador_n
suitably_adv
suitable_adj
refreshment_n
refresh_v
ironically_adv
ironic_adj
irony_n
contribute_v
contribution_n
establishment_n
establish_v
signal_v
signal_n
purchase_v
purchase_n
enable_v
originality_n
original_adj
origin_n
phenomenon_n
evident_adj
evidence_n
tremendous_adj
version_n
garment_n
exceed_v
logo_n
memorable_adj
prominent_adj
moderately_adv
moderate_adj
rent_v
rent_n
rental_n
furnished_adj
furnish_v
tenderly_adv
tender_adj
oblige_v
obligation_n
crack_v
crack_n
progressively_adv
progressive_adj
progress_v
progress_n
mock_v
ridicule_v
deny_v
denial_n
terror_n
wing_n
vague_adj
accompany_v
orchestra_n
pause_v
pause_n
announce_v
announcement_n
explode_v

explosion_n
explosive_adj
anxiously_adv
anxious_adj
anxiety_n
imitation_n
imitate_v
deafening_adj
deafen_v
applause_n
applaud_v
scar_n
scar_v
objective_n
observe_v
observation_n
physical_adj
defect_n
participant_n
participate_v
participation_n
sticky_adj
stick_v
clinic_n
peep_v
peep_n
assumption_n
assume_v
inevitably_adv
inevitable_adj
inferiority_n
inferior_adj
bound_adj
confront_v
confrontation_n
crisis_n
bruise_v
bruise_n
ego_n
hence_adv
limb_n
intend_v
intention_n
intent_n
beam_v
despite_con
deficient_adj
deficiency_n
embrace_v
embrace_n
mere_adj
pimple_n
steer_v
tourism_n
tourist_n
carve_v
confine_v
inspection_n
inspect_v
inspector_n
ruin_n
ruin_v
grave_adj
gravity_n
march_n
march_v
tramp_v
tramp_n
tomb_n
delicate_adj
scratch_v
scratch_n
treasure_v
treasure_n
remote_adj
industry_n
industrial_adj
industrialize_v
neighborhood_n
landscape_n
coastline_n
reservation_n
reserve_v
preserve_v
preservation_n
entry_n
permit_n
permit_v
permission_n
section_n
undisturbed_adj
disturb_v
disturbance_n
measure_n
restore_v
restoration_n
conservation_n
conserve_v
morality_n
moral_adj
moral_n
privilege_n
species_n
plot_n
plot_v
organism_n
structure_n
structural_adj

fossil_n
decade_n
survive_v
survival_n
survivor_n
advance_n
advance_v
advanced_adj
rate_n
native_n
native_adj
perish_v
revive_v
revival_n
accurately_adv
accurate_adj
accuracy_n
forecast_v
forecast_n
disaster_n
disastrous_adj
migrate_v
migrant_n
migration_n
flap_v
flap_n
vanish_v
massive_adj
scale_n
habitat_n
consumption_n
consume_v
consumer_n
wildlife_n
poacher_n
poach_v
proceed_v
arrest_v
arrest_n
petty_adj
jail_n
exhausted_adj
exhaust_v
accommodation_n
accommodate_v
suspicion_n
suspect_v
suspect_n
suspicious_adj
inquire_v
inquiry_n
convict_n
convict_v
humiliation_n
humiliate_v
exchange_v
exchange_n
release_v
release_n
shiver_v
shiver_n
pity_n
pity_v
dine_v
precious_adj
shelter_n
shelter_v
stunned_adj
stun_v
possession_n
possess_v
deserve_v
solemnly_adv
solemn_adj
roam_v
dreadful_adj
brood_v
gloom_n
gloomy_adj
somehow_adv
(511 words)

Taiwanese textbook 4

disease_n
potentially_adv
potential_adj
potential_n
excessive_adj
excess_n
label_v
label_n
alert_v
alert_n
alert_adj
nutrient_n
pedestrian_n
respectively_adv
respective_adj
serving_n
quantity_n
overall_adv
overall_adj
devour_v
aspect_n
instantly_adv
instant_adj
instant_n
furthermore_adv
universally_adv
universal_adj
barrier_n
currently_adv
current_adj
opposition_n
oppose_v
figure_n
advocate_n
advocate_v
counter_v
core_n
interpret_v
interpretation_n
enroll_v
enrollment_n
campus_n
imprison_v
imprisonment_n
dizzy_adj
pace_n
pace_v
competitive_adj
competition_n
compete_v
competitor_n
constant_adj
academic_adj
standard_n
standard_adj
obtain_v
phase_n
attain_v
attainment_n
simplify_v
simplicity_n
compromise_v
compromise_n
dependent_adj
suppose_v
catastrophe_n
justified_adj
justify_v
repress_v
reject_v
rejection_n
cherish_v
recommend_v
recommendation_n
professional_adj
profession_n
sheer_adj
ecstasy_n
covet_v
joyful_adj
lifetime_n
preach_v
raid_v
raid_n
slavery_n
slave_n
slave_v
shepherd_n
captivity_n
captive_adj
captive_n
perilous_adj
peril_n
bound_adj
priesthood_n
priest_n
convert_v
labor_v
labor_n
modesty_n
modest_adj
magnetic_adj
magnet_n
mission_n
missionary_n
retire_v
retirement_n

commemorate_v
surround_v
surroundings_n
sermon_n
entity_n
adopt_v
icon_n
holy_adj
evolve_v
evolution_n
parade_n
pinch_v
pinch_n
immigrant_n
immigrate_v
immigration_n
humanity_n
tug_v
tug_n
sleeve_n
plantation_n
inject_v
injection_n
pest_n
boost_v
boost_n
harvest_n
harvest_v
reload_v
load_v
load_n
barrel_n
spray_v
infertility_n
infertile_adj
disclose_v
disclosure_n
ban_v
ban_n
register_v
register_n
registration_n
export_v
export_n
yield_n
yield_v
decline_v
decline_n
poison_n
poison_v
poisonous_adj
staff_n
reluctant_adj
exposure_n
bubble_v
bubble_n
disregard_v
illuminating_adj
illuminate_v
innocent_adj
innocence_n
compel_v
campaign_n
campaign_v
dumb_adj
surgery_n
restrict_v
restriction_n
heel_n
undergo_v
astonishing_adj
astonish_v
astonishment_n
operation_n
operate_v
drill_v
drill_n
insert_v
rod_n
guarantee_v
guarantee_n
endanger_v
cripple_v
cripple_n
contemporary_adj
contemporary_n
hazard_v
hazard_n
perceive_v
pale_adj
tan_adj
tan_v
tan_n
pill_n
injurious_adj
injury_n
injure_v
contract_v
contract_n
itchy_adj
itch_v
reasonable_adj
cheerful_adj
excluding_con
exclude_v
multiple_adj
racist_adj
racism_n

occupy_v
occupation_n
badge_n
readily_adv
concealment_n
conceal_v
duration_n
pastime_n
namely_adv
wage_n
monotonous_adj
monotony_n
giggly_adj
giggle_v
brutal_adj
backpack_n
snatch_v
sealed_adj
seal_v
dread_n
dread_v
dreadful_adj
sow_v
slaughter_v
slaughter_n
gnaw_v
lessen_v
throng_n
throng_v
soberly_adv
sober_adj
sober_v
betray_v
capture_v
capture_n
sole_adj
diverse_adj
diversify_v
prediction_n
qualify_v
qualification_n
reproduce_v
unavailable_adj
available_adj
abruptly_adv
abrupt_adj
arise_v
exaggeration_n
exaggerate_v
utmost_adj
utmost_n
steep_adj
agricultural_adj
agriculture_n
pesticide_n
propose_v
proposal_n
feasible_adj
grain_n
attach_v
attachment_n
acute_adj
ecology_n
formation_n
elaborate_adj
network_n
sustain_v
famine_n
emigrate_v
emigration_n
emigrant_n
resistant_adj
resist_v
resistance_n
retain_v
offspring_n
suburb_n
suburban_adj
density_n
dense_adj
crammed_adj
cram_v
boom_v
boom_n
weave_v
primary_adj
vast_adj
errand_n
withdraw_v
deposit_v
deposit_n
install_v
installation_n
range_v
range_n
cuisine_n
strategy_n
strategic_adj
translator_n
translate_v
translation_n
document_n
document_v
documentary_adj
gum_n
counter_n
reliable_adj
supply_n

supply_v
stack_v
stack_n
combine_v
combination_n
promotion_n
promote_v
tempt_v
temptation_n
stimulate_v
stimulation_n
maximize_v
maximum_n
maximum_adj
profit_n
profitable_adj
ski_n
ski_v
leather_n
kit_n
vocal_adj
immensely_adv
immense_adj
mislead_v
classification_n
classify_v
stereotype_v
stereotype_n
overlook_v
identical_adj
series_n
neglect_v
neglect_n
adolescent_n
adolescence_n
manner_n
sloppy_adj
publication_n
unethical_adj
ethical_adj
ethic_n
merchant_n
distort_v
basis_n
nationality_n
orientation_n
orient_v
prejudice_n
acknowledge_v
acknowledgement_n
monopoly_n
peculiar_adj
accent_n
glamour_n
paste_v
stare_v
stare_n
vote_v
vote_n
elect_v
election_n
trample_v
vein_n
quiver_n
quiver_v
breathing_n
breathe_v
breath_n
glare_v
glare_n
troublesome_adj
strangle_v
inanimate_adj
animate_adj
paralyzed_adj
paralyze_v
frantically_adv
frantic_adj
despair_n
despair_v
obstacle_n
hushed_adj
hush_v
drained_adj
drain_v
electronic_adj
electronics_n
solo_adj
solo_n
surge_v
surge_n
vibrate_v
vibration_n
burst_v
burst_n
dose_n
realization_n
glow_n
glow_v
comprehend_v
comprehension_n
bold_adj
substantially_adv
substantial_adj
universe_n
lifespan_n
span_n
span_v

collapse_v
collapse_n
compose_v
composition_n
chunk_n
particle_n
dust_n
dust_v
dusty_adj
stray_adj
stray_v
stray_n
atom_n
atomic_adj
mystery_n
mysterious_adj
realm_n
astronomy_n
astronomer_n
volume_n
gleam_n
gleam_v
triple_adj
triple_v
plunge_v
bizarre_adj
permanent_adj
detention_n
detain_v
strand_n
shadow_n
emerge_v
dimension_n
source_n
assassinate_v
ascend_v
throne_n
accordingly_adv
vow_v
vow_n
stepfather_n
contemplate_v
contemplation_n
burden_n
burden_v
agony_n
hasty_adj
haste_n
crown_n
crown_v
endure_v
endurance_n
torment_n
torment_v
intolerable_adj
tolerable_adj
profound_adj
adviser_n
constructive_adj
construct_v
construction_n
heed_v
heed_n
intimate_adj
intimacy_n
prudent_adj
royal_adj
royalty_n
status_n
diplomatic_adj
diplomacy_n
diplomat_n
magnificent_adj
literature_n
(505 words)

Taiwanese textbook 5

college_n
tray_n
tedious_adj
elegant_adj
spill_v
spill_n
colleague_n
graduate_v
graduate_n
graduation_n
orchard_n
prune_v
vine_n
blazing_adj
blaze_v
blaze_n
bend_v
bend_n
cooperation_n
cooperate_v
cooperative_adj
resume_n
applicant_n
proficiency_n
eligible_adj
supplement_v
supplement_n
youngster_n
globe_n
global_adj
broaden_v
horizon_n
horizontal_adj
enhance_v
enhancement_n
worthwhile_adj
racial_adj
race_n
discrimination_n
discriminate_v
whistle_n
whistle_v
torch_n
growl_v
folklore_n
vivid_adj
drip_v
drip_n
jaw_n
lunar_adj
terrifying_adj
terrify_v
weirdly_adv
weird_adj
stiff_adj
penalty_n
abnormally_adv
abnormal_adj
regulation_n
regulate_v
lunatic_n
lunatic_adj
accumulate_v
accumulation_n
inference_n
infer_v
rotate_v
rotation_n
tide_n
tidal_adj
exert_v
ebb_v
confirm_v
confirmation_n
stable_adj
stability_n
stabilize_v
sway_v
adjust_v
adjustment_n
harsh_adj
condition_n
ritual_n
ancestor_n
refugee_n
refuge_n
execute_v
execution_n
guerrilla_n
muse_n
session_n
grammar_n
grammatical_adj
literary_adj
perspective_n
spur_v
spur_n
logic_n
logical_adj
stern_adj
pad_n
pad_v
bud_n
submit_v
contest_n
contestant_n
finance_v

finance_n
financial_adj
scholarship_n
scholar_n
summit_n
arm_n
monument_n
spark_n
spark_v
rebellious_adj
rebel_n
rebel_v
alternately_adv
alternate_adj
alternate_v
charm_v
charm_n
grief_n
grieve_v
emphatically_adv
emphatic_adj
distinguished_adj
distinguish_v
capacity_n
resolve_v
resolve_n
midst_n
crater_n
mournful_adj
mourn_v
abandon_v
smash_v
smash_n
cellar_n
bullet_n
dignity_n
civilization_n
civilize_v
compassion_n
compassionate_adj
extraordinary_adj
entitle_v
solitary_adj
grip_v
grip_n
electric_adj
electricity_n
unashamedly_adv
unashamed_adj
ashamed_adj
flawless_adj
flaw_n
mustache_n
frame_v
frame_n
soak_v
soak_n
strip_v
encounter_v
encounter_n
defy_v
mighty_adj
might_n
unite_v
unity_n
condense_v
demonstration_n
demonstrate_v
score_n
assemble_v
assembly_n
historic_adj
historical_adj
flame_n
flame_v
joyous_adj
daybreak_n
lonesome_adj
exile_v
exile_n
highlight_v
highlight_n
shameful_adj
shame_n
shame_v
urgency_n
capital_n
essence_n
oppression_n
oppress_v
transform_v
transformation_n
oasis_n
vicious_adj
governor_n
sibling_n
crooked_adj
crook_v
glory_n
glorious_adj
Lord_n
discord_n
symphony_n
liberty_n
dweller_n
dwell_v
dwelling_n
centigrade_adj

machinery_n
generate_v
generator_n
reflectiveness_n
reflective_adj
detect_v
trend_n
satellite_n
evaluate_v
evaluation_n
decisive_adj
interact_v
interaction_n
humid_adj
humidity_n
disperse_v
vegetation_n
opposite_n
opposite_adj
notable_adj
trap_v
trap_n
shallow_adj
layer_n
annoyance_n
pave_v
pavement_n
emission_n
emit_v
implementation_n
implement_v
investment_n
invest_v
anticipate_v
anticipation_n
radically_adv
radical_adj
discourteous_adj
courteous_adj
courtesy_n
foul_adj
incense_v
assault_n
assault_v
supervisor_n
supervise_v
supervision_n
subordinate_n
subordinate_adj
subordinate_v
deadline_n
taunt_v
harass_v
harassment_n
scold_v
passive_adj
discount_v
discount_n
plague_v
plague_n
impose_v
uncover_v
disabled_adj
disable_v
trauma_n
halt_v
halt_n
diminish_v
reassess_v
assess_v
assessment_n
wound_n
wound_v
heal_v
thereby_adv
utter_v
hurl_v
client_n
retort_v
defend_v
defense_n
defensive_adj
irritate_v
irritation_n
professor_n
therapy_n
therapist_n
grin_v
grin_n
outlook_n
consensus_n
twist_v
twist_n
outrage_v
outrage_n
outrageous_adj
ponder_v
versus_con
underlying_adj
underlie_v
downward_adj
spiral_n
distress_v
distress_n
confess_v
confession_n
reckon_v
strive_v

concede_v
motivation_n
fury_n
grumble_n
grumble_v
dissolve_v
mend_v
cement_v
punctual_adj
monitor_n
monitor_v
infant_n
choke_v
artery_n
transport_v
transportation_n
oxygen_n
metaphor_n
imply_v
implication_n
bloom_v
bloom_n
ward_n
cradle_v
cradle_n
accuse_v
accusation_n
bind_v
diagnose_v
diagnosis_n
deteriorate_v
hum_v
dough_n
pastry_n
lecture_n
lecture_v
portable_adj
hiss_n
hiss_v
click_n
click_v
clip_v
clip_n
presume_v
neutral_adj
neutrality_n
legal_adj
inconceivable_adj
conceivable_adj
conceive_v
reputation_n
legendary_adj
poetry_n
poetic_adj
verse_n
manifest_v
manifest_adj
aptitude_n
prime_adj
tread_v
sigh_n
sigh_v
hence_adv
stroll_v
stroll_n
route_n
companion_n
companionship_n
mutter_v
gorgeous_adj
scenery_n
scenic_adj
initial_adj
depict_v
dilemma_n
narrator_n
narrate_v
narrative_n
superficially_adv
superficial_adj
recount_v
inner_adj
arouse_v
hardship_n
option_n
optional_adj
(409 words)

Taiwanese textbook 6
flicker_v
flicker_n
speculate_v
architect_n
architecture_n
veterinarian_n
enterpriser_n
enterprise_n
toil_v
toil_n
taxing_adj
tax_v
tax_n
tuition_n
meditate_v
meditation_n
literally_adv
literal_adj
trifling_adj
trifle_n
minimum_adj
minimum_n
minimal_adj
minimize_v
consolation_n
console_v
destined_adj
inequality_n
equality_n
deem_v
perseverance_n
persevere_v
unfavorable_adj
favorable_adj
incentive_n
boundary_n
intense_adj
intensity_n
intensify_v
debate_n
debate_v
excellence_n
expertise_n
mastery_n
renowned_adj
mount_v
invasion_n
invade_v
string_n
string_v
alter_v
prior_adj
author_n
author_v
biography_n
formula_n
formulate_v
stagger_v
stream_n
stream_v
alien_adj
alien_n
retreat_n
retreat_v
ceaselessly_adv
ceaseless_adj
cease_v
differentiate_v
formidable_adj
marvelous_adj
marvel_n
exception_n
blessing_n
adulthood_n
accustomed_adj
accustom_v
worthy_adj
symmetry_n
awaken_v
bush_n
brook_n
miracle_n
miraculous_adj
grant_v
gaze_v
gaze_n
outer_adj
outline_v
outline_n
upright_adv
upright_adj
thrilling_adj
thrill_v
thrill_n
glimpse_n
glimpse_v
investigate_v
investigation_n
investigator_n
serene_adj
serenity_n
canvas_n
slum_n
pathetic_adj
hint_n
hint_v
strain_n

strain_v
perfume_n
perfume_v
relish_n
relish_v
scrap_n
abortion_n
abort_v
peek_v
peek_n
weary_adj
weary_v
physician_n
bluntly_adv
blunt_adj
surgeon_n
beforehand_adv
whisk_v
whisk_n
outcome_n
commitment_n
weep_v
notoriously_adv
notorious_adj
couch_v
pure_adj
purity_n
purify_v
outweigh_v
pardon_v
pardon_n
vulnerable_adj
disposable_adj
extend_v
extension_n
terminally_adv
terminal_adj
lawsuit_n
compensation_n
compensate_v
minor_adj
angel_n
rumble_v
rumble_n
boulevard_n
craft_n
dodge_v
distract_v
distraction_n
harbor_n
stimulus_n
mimic_v
mimic_n
videotape_v
videotape_n
analyze_v
analysis_n
startle_v
besiege_v
ambush_n
ambush_v
forthcoming_adj
prey_n
tiresome_adj
filter_n
filter_v
coordination_n
coordinate_v
abundance_n
abundant_adj
tackle_v
navigate_v
navigation_n
marine_adj
combat_n
combat_v
staircase_n
doorway_n
intact_adj
rubbish_n
passage_n
triumph_n
triumph_v
triumphant_adj
mist_n
statue_n
eternity_n
eternal_adj
burial_n
bury_v
chamber_n
solid_adj
coffin_n
arrow_n
jewel_n
jewelry_n
chariot_n
intrude_v
intruder_n
hut_n
erect_v
erect_adj
obscure_v
obscure_adj
assert_v
grim_adj
tag_v
tag_n

curse_n
curse_v
nonsense_n
acquainted_adj
acquaint_v
acquaintance_n
tobacco_n
alcohol_n
alcoholic_adj
alcoholic_n
persuade_v
persuasion_n
persuasive_adj
peer_n
lure_v
lure_n
urge_v
forbid_v
chew_v
slam_v
slam_n
bulk_n
prescribe_v
prescription_n
undermine_v
ample_adj
gang_n
stink_v
oddly_adv
odd_adj
sniff_v
stuff_n
senator_n
athlete_n
astronaut_n
nourishing_adj
nourish_v
nourishment_n
ambitious_adj
ambition_n
bunch_n
bronze_adj
bronze_n
dynamic_adj
tumor_n
beneath_con
caption_n
snap_v
innumerable_adj
biology_n
biological_adj
instructor_n
instruct_v
instruction_n
descendant_n
spectacular_adj
vaccine_n
cloning_n
clone_v
clone_n
infinite_adj
finite_adj
expire_v
expiration_n
consent_n
consent_v
ultimately_adv
ultimate_adj
halt_v
halt_n
unprecedented_adj
precedent_n
trace_v
trace_n
deadly_adj
fume_v
deed_n
exploit_v
bang_v
bang_n
whereas_con
nominate_v
nomination_n
nominee_n
sorrow_n
sorrowful_adj
perfectionism_n
perfection_n
revise_v
revision_n
curved_adj
curve_v
curve_n
sprint_v
sprint_n
pump_v
pump_n
strikingly_adv
striking_adj
mobility_n
mobile_adj
subtly_adv
subtle_adj
paradox_n
yearn_v
simultaneously_adv
simultaneous_adj
endeavor_v

endeavor_n
pedal_v
pedal_n
breakthrough_n
milestone_n
sentimental_adj
sentiment_n
prospect_n
prospective_adj
beloved_adj
depart_v
departure_n
guardian_n
conscientious_adj
rigorous_adj
congratulation_n
congratulate_v
diploma_n
vacancy_n
vacant_adj
institution_n
bleak_adj
plight_n
renewal_n
renew_v
suitcase_n
extensive_adj
discouraged_adj
discourage_v
discouragement_n
substitute_n
substitute_v
substitution_n
alley_n
chip_n
chip_v
catalog_n
catalog_v
thesis_n
irrelevant_adj
relevant_adj
chart_v
chart_n
philosopher_n
philosophy_n
philosophical_adj
commence_v
commencement_n
chest_n
synonym_n
barbarian_n
succession_n
successive_adj
(376 words)