



A Corpus Analysis of Vocabulary Coverage and Vocabulary Learning Opportunities within a Children's Story Series

Sun, Yu-Chih

National Chiao Tung University, Taiwan

ABSTRACT

Extensive reading for second language learners have been widely documented over the past few decades. However, few studies, if any, have used a corpus analysis approach to analyze the vocabulary coverage within a single-author story series, its repetition of vocabulary, and the incidental and intentional vocabulary learning opportunities therein. The current study analyzes 48 publications in the Magic Tree House series and compares (1) the vocabulary coverage across the series with the three word lists developed for middle school and high school in Taiwan; (2) the word-type repetition patterns observed in the corpus across the three word lists; and (3) the prospects of incidental vocabulary learning, word list learning, and intensive new word learning during extensive reading of the series. The rich Magic Tree House corpus makes repeated use of vocabulary from all three word lists, and the findings indicate that extensive reading can support word list learning, especially of the Middle School Basic 1,200 list. Research finding indicates that acquiring words from word lists prior to reading is not as effective as focusing on new word types as they appear in the series; as the number of readings increases, so too do the opportunities for incidental vocabulary acquisition.

INTRODUCTION

One of the most globally recognized problems in second language learning is learners' reluctance to read in the target language independently (Arnold, 2009; Day & Bamford, 2002; Lin, 2014). As Grabe and Stoller (2002) have pointed out, most second language learners lack enough exposure to material in order to develop fluent textual processing and to build a large vocabulary. The short articles in typical textbooks can hardly provide sufficient input to achieve the desirable learning outcome (Horst, 2005). To overcome these difficulties, reading a large quantity of target language materials can not only aid learners in acquiring new words, but incidentally strengthen previously learned ones (Horst, 2005).

Definition of extensive reading

Many researchers consider extensive reading (ER) to be key to holistic language proficiency (Day & Bamford, 1998; Horst, 2005). The theoretical underpinning of extensive

reading lies in the concept of input hypothesis (Krashen, 1982), which argues that sufficient comprehensible input ($i + 1$) can facilitate language acquisition across various language skills, such as reading, writing, vocabulary, grammar, and spelling (Yamashita, 2008). Extensive reading is a pedagogy with the following seven characteristics: (1) learners are encouraged to read a large quantity of materials (Richards & Schmidt, 2002; Rodrigo et al., 2007; Yamashita, 2008); (2) learners read for pleasure, information, or general comprehension, instead of learning specific language forms (Day & Bamford, 1998; Rodrigo et al., 2007; Stoeckel, Reagan, & Hann, 2012); (3) learners usually have the autonomy of freely choosing reading materials relevant to their interests (Grabe & Stoller, 2002), and read at a fast pace (Grabe & Stoller, 2002; Lin, 2014); (4) there is a focus on holistic language proficiency—for example, vocabulary, grammar, and reading comprehension skills (Iwahori, 2008; Mason & Krashen, 1997); (5) the ultimate goal for learners is to develop a sustained, life-long reading habit (Logan & Johnston, 2009; Richards & Schmidt, 2002); (6) the materials are within learners' linguistic competence and are easily comprehensible (Grabe & Stoller, 2002; Stoeckel, Reagan, & Hann, 2012); and (7) learners read independently (Stoeckel, Reagan, & Hann, 2012). In short, extensive reading aims to provide enjoyable reading and learning opportunities for language learners and thus differs from traditional intensive reading approaches.

Benefits of extensive reading

The benefits of extensive reading are well documented in the literature. A large body of research has reported beneficial effects in the following areas: (1) Reading fluency: extensive reading can develop fluency by providing repeated access to known words, leading to improvement in reading comprehension (Hafiz & Tudor, 1989; Hitosugi & Day, 2004; Mason & Krashen, 1997; Nakanishi & Ueda, 2011; Robb & Susser, 1989; Yamashita, 2008), reading strategies (Nishino, 2007), and reading rates or speed in the target language (Beglar, Hunt, & Kite, 2012; Day & Bamford, 1998; Iwahori, 2008; Lao & Krashen, 2000; Mason & Krashen, 1997; Nuttall, 2005; Stoeckel, Reagan, & Hann, 2012). (2) Vocabulary acquisition: through extensive reading, learners can acquire new vocabulary by encountering words repeatedly, enlarging their vocabulary size (Grabe & Stoller, 1997; Horst, 2005; Pigada & Schmitt, 2006; Shin, 2003), improving their spelling (Krashen, 1993), and enhancing their knowledge of contextualized vocabulary (Nakanishi, 2014). (3) Grammatical competence (Krashen, 1982; Yang, 2001): extensive reading can foster learners' awareness of syntactic structure, (4) Writing proficiency (Hafiz & Tudor, 1989; Tsang, 1996): extensive reading can develop learners' ability in idea expression, language usage, and writing skills, (5) Affective promotion: extensive reading can help learners develop a positive attitude toward language learning (Al-Homoud & Schmitt, 2009; Asraf & Ahmad, 2003; Hitosugi & Day, 2004), making it a source of enjoyment (Day & Bamford, 1998; Nakanishi, 2014), and (6) Information: reading can be a source of meaning-focused input (Nakanishi, 2014).

In summary, several studies have demonstrated that an extensive reading approach outperforms traditional approaches (Mason & Krashen, 1997; Nakanishi, 2014) and can greatly contribute to learners' overall language proficiency (Iwahori, 2008; Mason & Krashen, 1997). Furthermore, extensive reading can improve learners' performance on tests, such as the TOEFL (Constantino, Lee, Cho, & Krashen, 1997) and TOEIC tests (Storey, Gibson, & Williamson, 2006), especially for low-motivation learners (Stoeckel, Reagan, & Hann, 2012).

However, researchers also caution that the delayed observable benefits of extensive reading make it challenging for shorter periods of engagement. As Krashen (1993) notes, “Short-term projects are not as consistently successful” (p. 73) and at least one school year of extensive reading is necessary. Yamashita (2008) also points out that “extensive reading is often dropped from the curriculum before L2 teachers have a chance to recognize its real impact” (p. 663).

Prior vocabulary knowledge in extensive reading

Readers need to have a basic working vocabulary before engaging in extensive reading (Bryan, 2013; Coady, 1997; Laufer, 1997). Several studies have further noted that learners need to build a sufficient initial repertoire of vocabulary to develop their competency in listening, speaking, reading, and writing (Coady & Huckin, 1997; Krashen, 1989). In addition, Collins (2010) notes that learners with insufficient initial vocabulary need to receive more intensive exposure and instruction. Coady (1997) even argues that extensive reading is only effective for intermediate and advanced learners with basic vocabulary. He refers to the requirement as the “beginner’s paradox,” in which learners need to first learn enough words before they can incidentally learn vocabulary through extensive reading (p. 230). Nation (2001) studied the relationship between known-word densities in text and reading comprehension and noted that 5% unknown words (one word in every twenty) is the threshold for learners to comprehend and infer meaning from unfamiliar words they meet in reading.

Incidental vocabulary learning in extensive reading

Studies have shown that both receptive and productive language skills can be improved incidentally through extensive reading (e.g., Eckerth & Tavakoli, 2012). Furthermore, extensive reading can consolidate partially learned words (Brown, 2009) and improve learners’ knowledge of the contextual usage of vocabulary (Nakanishi, 2014) and serves as “a quantitatively and qualitatively rich context and resource for lifelong lexical development” (Eckerth & Tavakoli, 2012, p. 229).

Research findings have indicated that incidental vocabulary learning in extensive reading is a “gradual process in which gains are made in small increments with repeated encounters needed to gain full knowledge of a word” (Webb, 2007, p. 47). Krashen (1989) also notes that incidental vocabulary learning can occur through meaning-focused extensive reading, and graded reading is effective in developing vocabulary knowledge (Chun, Choi, & Kim, 2012; Nakanishi, 2014; Nation, 2009; Schmitt, 2008; Waring & Takaki, 2003). Extensive reading over a period of time can compensate for the limited vocabulary gains from single texts and reduce the effect of knowledge decay (Laufer & Rozovski-Roitblat, 2014).

Regarding the number of repetitions required to “pick up” vocabulary, previous studies have shown that no set number of repetitions will guarantee success (Laufer & Rozovski-Roitblat, 2014). Several other mediating variables explain the lack of agreement—for example, the context in which the target vocabulary occurs and the relevant clues inferred in context (Hulstijn et al., 1996; Peters, 2014; Rott, 1999). In addition, learners’ cognitive skills, inference skills, and vocabulary size also have a bearing on their ability to acquire target vocabulary (Horst et al., 1998; Peters, 2014; Zahar et al., 2001).

Narrow reading

Studies have shown that narrow reading—reading several books by the same author—improves the likelihood of incidental vocabulary learning (Cho, Ahn, & Krashen 2005; Gardner 2008; Maxim II, 2002; Pulido, 2004), as authors tend to possess a certain repertoire of vocabulary, have observable preferences of word choice, and tend to use the same phrases repeatedly. In addition, the familiar characters and background of a series make for an easier reading experience (Krashen, 1993).

Gardner (2008) states that books by the same author tend to exhibit similar collocation, allowing for more effective vocabulary recycling than those written by multiple authors. The familiar “style and contextual background of the stories” (Bryan, 2013, p. 117) affords higher repetition rates and leads to more effective incidental vocabulary acquisition. Furthermore, Gardner (2008) shows that narrative writing tends to be more effective for vocabulary learning when it is written by the same author, whereas expository writing is more effective if it is thematically linked. Bryan (2013) also states that authentic children’s narrative stories written by the same author tend to have significantly more specialized words recycled than those by different authors.

Purpose of the study

In spite of its many advocates in education, extensive reading is still not commonly practiced in EFL settings in Taiwan. One factor that hinders the adoption of extensive reading pedagogy is that the culture is rooted in teaching for high-stakes high school and college entrance examinations, meaning that teachers, students, and parents perceive extensive reading as impractical due to the lack of immediate measurable gains in test scores. Another reason is the lack of reading materials brought about by budget shortages, meaning departments are unable to purchase enough books for learners to select reading materials based on their varying interests and proficiency. Most students will focus on memorizing vocabulary from word lists, intensively read short articles in the textbooks, or spend a lot of time working on practice tests, mostly multiple-choice questions on grammar, vocabulary, and reading comprehension.

The benefits of story series reading are listed as follows: (1) Economics: as in many EFL contexts, access to story books can be very limited due to budget constraints in schools or the lack of librarian resources to manage book circulation. An abundant choice of books is a luxury and thus the concept of voluntary extensive reading is pragmatically difficult to implement. A more feasible alternative is to adopt extensive reading with a fixed but large set of popular stories, such as the Magic Tree House series, which has over fifty published books. In addition, even though the books in the Magic Tree House series are not graded, there is an acceleration in difficulty in terms of vocabulary size and book length between the first 28 books and those after. So, providing teachers with one popular book series costs less than providing them with a wide variety of books. In addition, following the same series is more cost-effective, especially for large-scale implementations of extensive reading across a whole school or even school district, (2) Motivation: readers will become familiar with the characters in the story series and will be more likely to stay with the books, (3) Convenience: by following the same book series, it saves the time and energy when finding the next book to read, and (4) Vocabulary recycling: research has shown that books by the same author tend to exhibit more recycling than books by different authors, and thus benefit learners’ vocabulary gains due to the higher number of repetitions.

To date, few studies, if any, focus on the coverage of vocabulary in word lists compared to that in a book series. The current study analyzes the corpus of a selected series (Magic Tree House, books 1 to 48) and compares the vocabulary used in the series with three word lists published by the Ministry of Education in Taiwan to show how reading extensively on a particular story series can support vocabulary acquisition on word lists. In addition, it should be of interest to educators how extensive reading of a series can foster incidental vocabulary learning due to the greater repetition of vocabulary. Finally, the study sketches three approaches to vocabulary learning (incidental vocabulary learning, word list learning, and intensive learning of new vocabulary in reading materials) and uses corpus analysis to highlight the potential pedagogic opportunities for each. The findings of the study can hopefully provide a prospective vision of how extensive reading of a narrow, single-author story series can benefit vocabulary learning and thus boost confidence and interest in the adoption of extensive reading in EFL settings in Taiwan.

Research questions:

1. How does the vocabulary coverage in the Magic Tree House series compare to the three word lists developed for middle school and high school in Taiwan?
2. Which are the word-type repetition patterns observed in the Magic Tree House corpus across the three word lists examined in the study?
3. What are the prospects of incidental vocabulary learning, word list learning, and intensive new word learning following extensive reading of the Magic Tree House series?

METHODOLOGY

Materials

The Magic Tree House series (books 1–48) by American writer Mary Pope Osborne was chosen as the target corpus for analysis due to its popularity among native speakers of English as well as its worldwide acceptance (the series has been translated into more than 30 languages, including Chinese, and more than 100 million copies have been sold worldwide; https://en.wikipedia.org/wiki/Mary_Pope_Osborne). The amount of material in the series makes it an ideal selection for long-term extensive reading, taking into account budget constraints and convenience. Furthermore, due to the highly competitive nature of examination, Taiwan's 6–12 age group is more likely to engage in extensive reading than older age groups, thus making the findings of the study more applicable to wider educational contexts.

Word lists

The Ministry of Education in Taiwan has published three word lists, *Middle School Basic 1,200*, *Middle School Advanced 800*, and *High School 4,700*, as guidelines for textbook developers to follow. That is, the vocabulary coverage in the textbooks should contain only words found on those lists. The aim is to ensure common core English-language materials are followed nationwide, so that test developers for the entrance exams are confined to the vocabulary range specified by the ministry.

Procedure and tool development

First, in preparation for corpus analysis, a scanned file for each of the 48 Magic Tree House books was created. Then, a computerized tool was developed by Visual Studio C# to count each word type in the corpus. The tool automatically splits each sentence in the corpus into single words and calculates the frequency of each word. The tool also matches and compares words from the corpus with the three word lists mentioned above. Thus, a list of word types and running words appearing in each story was retrieved and a frequency profile for each word type across the whole scanned corpus was calculated and analyzed. The data were then exported to an Excel file for further analysis.

RESULTS

3.1 Research question 1: How does the vocabulary coverage in the Magic Tree House series compare to the three word lists developed for middle school and high school in Taiwan?

For books 1 to 28, the range of running words is from 4,500 to 6,564 and the mean is 5,584 words, whereas for books 29–48, the range of running words is from 7,613 to 13,170 and the mean is 11,278. Regarding the total number of word types, the range of the first 28 books falls between 857 and 1,438, with a mean of 1,111; whereas the range of word types for books 29–48 falls between 1,452 and 1,991, with a mean of 1,754. It is obvious then that in terms of the length and variety of word types, the second part of the series is both longer and more diverse.

In terms of the Magic Tree House's vocabulary coverage of the three word lists, for *Middle School Basic 1,200* words, the range of vocabulary coverage is from 356 to 639 words, with a mean of 476 words (38%); for *Middle School Advanced 800* words, the range of coverage is from 44 to 140, with a mean of 91 (12%); for *High School 4,700* words, the coverage range is from 61 to 247, with a mean of 143 (3%).

Table 1 presents the cumulated percentage of coverage if readers read from book 1 to 48. The results show that the percentage of cumulated coverage increases gradually. Reading books 1–5 will cover 50% of *Middle School Basic 1,200* words, reading books 1–15 will cover over 70% of *Basic 1,200* words, reading books 1–25 will cover 80% of *Basic 1,200* words, and reading books 1–44 will cover 90% of *Basic 1,200* words. For *Middle School Advanced 800* words, reading books 1–24 will cover 50% of the words on the list and reading books 1–44 will cover 70%. For *High School 4,700* words, the coverage rates decrease drastically and reading books 1–48 will only cover 36% of vocabulary on the list.

Table 1. Cumulated Percentage of Word-type Coverage Across the Three Word Lists

Book	Running Words	Total Word Types	Cumulated Coverage of MSB 1,200 Words (%)	Cumulated Coverage of MSA 800 Words (%)	Cumulated Coverage of HS 4,700 Words (%)
1.	4,775	857	28.5	6.8	1.3
2.	5,272	1,042	38.2	11.0	3.0
3.	5,114	955	43.8	14.7	4.2
4.	5,356	1,011	47.4	19.0	5.3
5.	5,328	904	50.0	21.2	6.0
6.	4,500	860	51.8	22.2	6.6

7.	4,731	955	54.7	22.8	7.2
8.	5,689	1,066	57.3	25.1	7.9
9.	4,764	1,009	59.7	28.5	8.9
10.	6,351	1,168	62.3	31.8	10.1
11.	5,332	1,085	64.3	33.2	10.8
12.	5,674	1,118	65.8	34.6	11.5
13.	6,049	1,247	68.1	35.9	12.5
14.	5,422	1,108	69.5	36.9	13.1
15.	5,793	1,181	71.0	38.5	13.7
16.	5,511	1,088	71.7	40.4	14.4
17.	5,641	1,438	73.9	43.5	15.8
18.	5,645	1,116	74.7	44.0	16.2
19.	5,482	1,134	75.2	44.7	16.6
20.	5,669	1,141	76.1	45.6	17.1
21.	6,153	1,206	77.7	47.1	17.9
22.	5,119	1,069	78.5	48.3	18.4
23.	5,502	1,181	79.0	49.7	18.8
24.	5,956	1,181	79.7	50.6	19.3
25.	6,564	1,279	80.4	52.2	20.1
26.	6,250	1,200	80.7	53.1	20.8
27.	6,419	1,258	81.9	54.0	21.5
28.	6,279	1,238	82.2	54.8	21.7
29.	11,063	1,643	82.7	55.8	22.6
30.	10,670	1,697	83.4	56.7	23.5
31.	10,702	1,610	83.8	57.3	24.2
32.	11,047	1,633	84.0	58.6	24.8
33.	8,849	1,572	84.3	59.3	25.7
34.	11,593	1,854	84.9	60.7	26.6
35.	11,455	1,763	86.0	62.6	27.5
36.	7,613	1,452	86.5	63.6	28.2
37.	9,777	1,708	86.8	64.4	28.8
38.	12,392	1,905	87.4	65.8	29.8
39.	11,067	1,781	87.6	66.3	30.7
40.	10,891	1,700	87.8	67.7	31.4
41.	11,427	1,673	88.1	68.3	32.3
42.	12,796	1,925	88.9	69.5	33.0
43.	13,170	1,958	89.6	69.8	33.7
44.	12,952	1,991	90.2	71.1	34.8
45.	12,907	1,933	90.8	71.7	35.1
46.	13,099	1,831	91.0	72.5	35.5
47.	12,641	1,794	91.1	73.0	36.1
48.	9,443	1,659	91.4	73.8	36.4

The findings reveal that extensive reading of the Magic Tree House series can support word list learning, although there is more coverage in the series of the *Middle School Basic 1,200* word list than there is of the other two word lists. Reading the first 48 books of the series covers 91.4% of the *Basic 1,200* word list, 36.4% of *Advanced 800*, and only 36.4% of *High School 4,700*.

3.2 Research question 2: Which are the word-type repetition patterns observed in the Magic Tree House corpus across the three word lists examined in the study?

Middle School Basic 1,200 words

Table 2 presents the number of vocabulary repetitions for *Middle School Basic 1,200* words. The ten most repeated words from the list are *the* (25,565 times), *and* (10,633 times), *to* (8,289 times), *a* (7,986 times), *he* (7,548 times), *of* (6,028 times), *in* (4,084 times), *his* (3,884 times), *was* (3,654 times), and *you* (3,496). Of the total 1,250 words on the list, 951 (76.8%) have six or more repetitions throughout the series, 845 (67.6%) have eleven or more repetitions, and 700 (56%) of them reach over twenty-one repetitions.

Table 2. Number of Word-type Repetitions of *Middle School Basic 1,200* Words

Number of Occurrences	Number of Word Types	Cumulated Word Types	Percentage (%)	Cumulated Percentage (%)
Over 1,000	44	44	3.52	3.52
500–1,000	41	85	3.28	6.80
100–500	230	315	18.40	25.20
51–100	169	484	13.52	38.72
41–50	57	541	4.56	43.28
31–40	70	611	5.60	48.88
21–30	89	700	7.12	56.00
11–20	145	845	11.60	67.60
6–10	106	951	8.48	76.08
5	33	984	2.64	78.72
4	25	1,009	2.00	80.72
3	37	1,046	2.96	83.68
2	41	1,087	3.28	86.96
1	55	1,142	4.40	91.36
0	108	1,250	8.64	100.00

Middle School Advanced 800 words

Table 3 summarizes the vocabulary repetition of *Middle School Advanced 800* words. The ten most repeated words are *through* (830 times), *toward* (410 times), *thought* (398 times), *woods* (310 times), *backpack* (225 times), *deep* (222 times), *forest* (165 times), *against* (146 times), *stone* (133 times), and *aloud* (131). Of the 789 words on the list, 317 (40.18%) have six or more repetitions throughout the series, 229 (29.02%) have eleven or more repetitions, and 152 (19.26%) reach over twenty-one repetitions.

Table 3. Number of Word-type Repetitions of *Middle School Advanced 800* Words

Number of Occurrences	Number of Word Types	Cumulated Word Types	Percentage (%)	Cumulated Percentage (%)
Over 1,000	0	0	0	0
500–1,000	1	1	0.13	0.13
100–500	21	22	2.663	2.79
51–100	38	60	4.82	7.60
41–50	21	81	2.66	10.27
31–40	27	108	3.42	13.69
21–30	44	152	5.58	19.26
11–20	77	229	9.76	29.02
6–10	88	317	11.15	40.18
5	32	349	4.06	44.23
4	34	383	4.31	48.54
3	48	431	6.08	54.63
2	64	495	8.11	62.74

1	87	582	11.03	73.76
0	207	789	26.24	100.00

High School 4,700 words

Table 4 highlights the vocabulary repetition of *High School 4,700* words. The ten most repeated words are *Jack* (10,609 times), *saw* (385 times), *okay* (341 times), *ladder* (303 times), *found* (298 times), *chapter* (278 times), *creek* (248 times), *onto* (236 times), *breath* (216 times), and *huge* (205 times). Of the 4,531 words on the list, 595 (13.13%) have six or more repetitions throughout the series, 359 (7.92%) have eleven or more repetitions, and 193 (4.26%) reach over twenty-one repetitions.

Table 4. Number of Word-type Repetitions of *High School 4,700* Words

Number of Occurrences	Number of Word Types	Cumulated Word Types	Percentage (%)	Cumulated Percentage (%)
Over 1,000	1	1	0.02	0.02
500–1,000	0	1	0	0.02
100–500	21	22	0.46	0.49
51–100	40	62	0.88	1.37
41–50	25	87	0.55	1.92
31–40	35	122	0.77	2.69
21–30	71	193	1.57	4.26
11–20	166	359	3.66	7.92
6–10	236	595	5.21	13.13
5	76	671	1.68	14.81
4	121	792	2.67	17.48
3	158	950	3.49	20.97
2	222	1,172	4.90	25.87
1	478	1,650	10.55	36.42
0	2,881	4,531	63.58	100.00

Summary of vocabulary from the three word lists covered in the Magic Tree House series

The total number of word types covered in the three word lists is 6,570. Table 5 illustrates the distribution of repetitions covered in the 48 books of the Magic Tree House series across the three word lists. Of the 6,570 words listed in the three word lists, 1,863 (28.36%) have six or more repetitions throughout the 48 books, 1,433 (21.81%) have eleven or more repetitions, and 1,045 (15.91%) reach twenty-one repetitions.

Table 5. Number of Vocabulary Repetitions in the Three Word Lists throughout the 48 Books of the Magic Tree House Series

Number of Repeated Occurrences	Total Word Types	Cumulated Total Word Types	Percentage (%)	Cumulated Percentage (%)
Over 20,000	1	1	0.02	0.02
Over 10,000	2	3	0.03	0.05
Over 1,000	42	45	0.64	0.68
500–1,000	42	87	0.64	1.32
101–500	273	360	4.16	5.48
91–100	31	391	0.47	5.95

81-90	35	426	0.53	6.48
71-80	57	483	0.87	7.35
61-70	47	530	0.72	8.07
51-60	74	604	1.13	9.19
41-50	105	709	1.60	10.79
31-40	132	841	2.01	12.80
21-30	204	1,045	3.11	15.91
11-20	388	1,433	5.91	21.81
6-10	430	1,863	6.54	28.36
5	141	2,004	2.15	30.50
4	180	2,184	2.74	33.24
3	243	2,427	3.70	36.94
2	327	2,754	4.98	41.92
1	620	3,374	9.44	51.35
0	3,196	6,570	48.65	100.00

3.3 Research question 3: What are the prospects of incidental vocabulary learning, word list learning, and intensive new word learning following extensive reading of the Magic Tree House series?

Model 1: Learning vocabulary through extensive reading only

The total number of word types presented in the Magic Tree House corpus is 10,750. Table 6 illustrates the distribution of repetitions of these words. Among them, 3,809 (35.43%) words have six or more repetitions throughout the series, 2,579 words (23.99%) have eleven or more repetitions, and 1,677 word types have twenty-one (15.6%) or more repetitions.

Table 6. Distribution of Vocabulary Repetition for All Word Types throughout the 48 Books of the Magic Tree House Series

Number of Occurrences	Total Word Types	Cumulated Word Types	Percentage (%)	Cumulated Percentage (%)
Over 20,000	1	1	0.01	0.01
Over 10,000	2	3	0.02	0.03
Over 1,000	44	47	0.41	0.44
500-1,000	51	98	0.47	0.91
101-500	374	472	3.48	4.39
91-100	46	518	0.43	4.82
81-90	53	571	0.49	5.31
71-80	84	655	0.78	6.09
61-70	67	722	0.62	6.72
51-60	123	845	1.14	7.86
41-50	182	1,027	1.69	9.55
31-40	226	1,253	2.10	11.66
21-30	424	1,677	3.94	15.60
11-20	902	2,579	8.39	23.99
6-10	1,230	3,809	11.44	35.43
5	413	4,222	3.84	39.27
4	600	4,822	5.58	44.86
3	855	5,677	7.95	52.81
2	1,305	6,982	12.14	64.95
1	3,768	10,750	35.05	100.00

Even after reading only the first 28 books, the total number of word types covered is 6,444 words. Table 7 illustrates the distribution of repetition of these words. Among them, 2,106 words (32.68%) have six or more repetitions, 1,406 words (21.82%) have eleven or more repetitions, and 864 words (13.41%) have twenty-one or more repetitions.

Table 7. Distribution of Vocabulary Repetition throughout Books 1–28 of the Magic Tree House Series

Number of Occurrences	Total Word Types	Cumulated Word Types	Percentage (%)	Cumulated Percentage (%)
Over 20,000	1	1	0.02	0.02
Over 10,000	1	2	0.02	0.03
Over 1,000	15	17	0.23	0.26
500–1,000	24	41	0.37	0.64
101–500	164	205	2.55	3.18
91–100	18	223	0.28	3.46
81–90	29	252	0.45	3.91
71–80	42	294	0.65	4.56
61–70	51	345	0.79	5.35
51–60	64	409	0.99	6.35
41–50	89	498	1.38	7.73
31–40	128	626	1.99	9.71
21–30	238	864	3.69	13.41
11–20	542	1,406	8.41	21.82
6–10	700	2,106	10.86	32.68
5	279	2,385	4.33	37.01
4	342	2,727	5.31	42.32
3	527	3,254	8.18	50.50
2	883	4,137	13.70	64.20
1	2,307	6,444	35.80	100.00

Regarding the findings show that for each new reading of the series, repetitions increase naturally; thus, incidental vocabulary acquisition can be expected accordingly. By the end of the series, 836 words fall in the category of repeating 11–20 times, 388 words repeat 21–30 times, 300 words repeat 31–40 times, 189 words repeat 41–50 times, and 853 words repeat over 51 times.

Thus, for vocabulary that requires 6–10 repetitions to acquire, learners reading the 48 books of the Magic Tree House series can acquire 3,767 words; for vocabulary that requires 11–20 repetitions, 2,566 words will be acquired; for 21–30 repetitions, 1,730 words will be acquired; for 31–40 repetitions, 1,342 words will be acquired; and for vocabulary that requires over 51 repetitions, 1,042 words will be acquired.

Model 2: Learning word lists prior to engaging in extensive reading

Table 8 summarizes the mean and percentage of unknown words from the three words lists learned prior to reading Magic Tree House. If learners master the *Middle School Basic 1,200* words prior to reading the series, the percentage of unknown word types is 37.69% and unknown running words is 64.63%. However, if learners focus on the *Advanced 800* words or *High School 4,700* word lists, the percentage of unknown word types actually increases drastically to 96.73% and 92.64% respectively. The findings show that it would be more effective if learners focus on learning words that are of a similar difficulty level to the books. In this case, the *Middle School Basic 1,200* word list is more relevant and helpful for learners.

Table 8. Mean and Percentage of Unknown Words from the Three Word Lists Learned prior to Reading Magic Tree House

Various Vocabulary Learning approaches	Unknown Word Types and Running Words	Unknown Words for Series 1–28		Unknown Words for Series 29–48		Unknown Words for Series 1–48	
		Mean	%	Mean	%	Mean	%
After learning MSB 1,200 words	Word types	729	37.83	1,200	37.48	902	37.69
	Running words	2,262	61.99	4,226	68.34	2,992	64.63
After learning MSA 800 words	Word types	1,095	96.78	1,633	96.66	1,287	96.73
	Running words	5,786	93.84	10,904	93.12	7,695	93.54
After learning HS 4,700 words	Word types	1,059	92.53	1,555	92.80	1,235	92.64
	Running words	5,534	90.89	10,469	88.67	7,377	89.97
After learning MSB 1,200 and MSA 800 words	Word types	657	34.60	1,079	34.14	811	34.41
	Running words	2,070	55.83	3,852	61.44	2,731	58.17
After learning all three lists	Word types	549	27.14	880	26.93	669	27.05
	Running words	1,625	46.72	3,044	50.13	2,152	48.14

Model 3: Intensive vocabulary learning for each newly encountered word

Table 9 summarizes the number and percentage of new word types encountered in reading each book while intensively acquiring vocabulary for each newly encountered word. The results show that after reading 13 books, the new vocabulary percentage (words that have not appeared before) in each reading is less than 20%. After reading 16 books, the new vocabulary percentage is less than 15%. After reading 45 books, the new vocabulary percentage drops to less than 10%.

Table 9. Number and Percentage of New Word Types Encountered in the Series

Book Number	Total Number of Word Types	Old Words	New Words	New Words %
1	806	0	806	100.0%
2	898	456	442	49.2%
4	932	620	312	33.5%
13	1,164	877	287	24.7%
16	1,030	875	155	15.0%
45	1,826	1,677	149	8.2%
48	1,571	1,456	115	7.3%

In terms of running words, after reading the first four books and acquire all the word types in those four books, readers basically encounter less than 10% new running words (except for books 9 and 10) and there are thirteen books that have less than 5% new running words. The findings also indicate that after mastering the first four books with guided and intensive vocabulary learning, most learners will encounter less than 10% unknown running words in the subsequent reading. Thus, reading alone with dictionary support is a feasible approach. After the first 28 books, reading for pleasure (i.e., without needing to look up unknown words) is feasible for most learners. Figure 1 illustrates the drastically declining curve of new running words across the series.

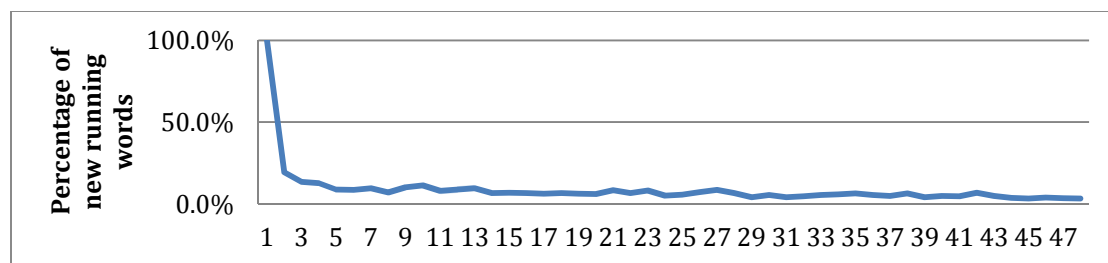


Figure 1. Distribution of new running words across books 1–48 of the Magic Tree House series

DISCUSSION

The findings indicate that extensive reading of the Magic Tree House series can support word list learning due to its wide coverage of vocabulary from the three lists, especially the *Middle School Basic 1,200* list. The 48 books of the series cover 91.4% of *Basic 1,200* words, 36.4% of *Advanced 800* words, and 36.4% of *High School 4,700* words. In addition, the rich Magic Tree House corpus provides ample repetition of vocabulary from all three word lists. In particular, the findings indicate that the more basic the vocabulary list is, the higher the percentage of repetition. Therefore, extensive reading of only one story series will not only expose learners to a large portion of words from the lists, but many of them repeatedly.

For incidental vocabulary learning through extensive reading, regardless of how many repetitions are required, the results of the quantitative corpus analysis of the Magic Tree House series reveal that as the number of readings increases, so too does the likelihood of incidental vocabulary acquisition. In addition, for every book read, new vocabulary is acquired due to the cumulated number of repetitions. So, the overall findings of the study indicate that incidental vocabulary learning through extensive reading is very probable and should be expected. What is more encouraging is that readers will not only incidentally acquire many words from the official word lists, but also acquire many from outside them

The findings of the study also indicate that even for learners mastering all the words from the three word lists prior to extensively reading the story series, the percentage of unknown words is still over 25%. In addition, since the word types occurring in the Magic Tree House series cover more words from the *Basic 1,200* word list, it is much more effective to acquire *Basic 1,200* words beforehand in order to increase the percentage of known words during reading, rather than focusing on *Advanced 800* or *High School 4,700* words. The last approach is to focus only on the words as they appear in each book. The findings of the corpus analysis reveal that by mastering all of the first five books' word types (4,445 words), the percentage of unknown word types drops to 23.3% while reading the sixth book, whereas by mastering all the words on the three lists, the percentage of unknown word types is still 27.05%. The results show that acquiring words from word lists prior to extensive reading is not as effective as focusing on the word types as they appear in the series.

CONCLUSION

Pedagogical implications

Research findings have indicated that motivation is the prime factor for the success of extensive reading in L1 and L2 contexts (Al-Homoud & Schmitt, 2009; Hitosugi & Day, 2004) and the availability of reading materials is another determining factor for the adoption and sustainment of extensive reading, especially in EFL contexts. Thus, utilizing popular book series is a simple solution for the adoption of extensive reading in class, due to their relative accessibility, popularity, and use of repeated vocabulary.

The findings of the study provide evidence of the Magic Tree House's institutionally approved vocabulary coverage and should encourage the adoption of extensive reading in a highly test-driven, word-list-guided education context. The demonstrable relationship between texts presented in extensive reading and word lists specified by the Ministry of Education can serve as a convincing reason for reluctant parents, teachers, and students to accept the non-test-oriented learning approach. The findings of the study strengthen our beliefs in the possibility of acquiring vocabulary incidentally through extensive reading. The findings also show that the word-list approach works best when the list corresponds with the vocabulary level of the target reading materials. In addition, mastering the word types in a story series appears to be more effective in reducing the percentage of unknown words than mastering vocabulary from word lists prior to reading extensively.

Limitations and recommendations for future study

The current study employs corpus analysis to investigate the repetition of vocabulary in an authentic story series and sketch three possible vocabulary teaching/learning paths based on the findings. However, it lacks empirical evidence from learners and actual learning to support the hypothetical arguments. Future research using an experimental or quasi-experimental approach to examine learners' vocabulary acquisition could provide empirical verification of the hypothetical predictions presented in the study—that is, to what extent repetition of words can lead to vocabulary acquisition in the short and long term. Furthermore, future research comparing learners with different vocabulary sizes and their acquisition of new vocabulary could shed new light on Krashen's comprehension input hypothesis on extensive reading and the subsequent selection of reading materials. In other words, do reading materials at a relatively more difficult level lead to more or less effective vocabulary acquisition? In addition, the current study is only limited to one book series, the Magic Tree House. Future research using simplified, graded reading material could further explore the relationship between vocabulary size and frequency of repetition in the material selected for extensive reading.

Since some popular story series publish Chinese and even bilingual versions, future research comparing the effectiveness of reading English and Chinese/bilingual versions could clarify whether familiarizing learners with story content in their first language will facilitate their vocabulary acquisition in extensive reading in English. Also, future research using technology that fosters a ubiquitous learning environment (i.e., mobile devices such as smartphones and tablets) could be useful to see how technology can facilitate extensive reading. A comparison of vocabulary acquisition through natural repetition in different contexts and repetition by re-reading the same texts could also shed new light on the effectiveness of different reading approaches. Finally, qualitative studies exploring the factors involved in not only successful extensive-reading programs but also not-so-successful ones might help educators gain more in-depth understanding of the complexities involved in implementing extensive reading in EFL contexts.



Yu-Chih Sun is a professor in the Institute of TESOL at National Chiao Tung University, Taiwan. Her research interests include computer-assisted language learning, academic writing, and extensive reading.

Email: sunyc@mail.nctu.edu.tw

REFERENCES

- Al-Homoud, F., Schmitt, N. (2009). Extensive reading in a challenging environment: a comparison of extensive and intensive reading approaches in Saudi Arabia, *Language Teaching Research*, 13(4), 383–401.
- Arnold, N. (2009). Online extensive reading for advanced foreign language learners: an evaluation study. *Foreign Language Annals*, 42(2), 340-366.
- Asraf, R. M., & Ahmad, I. S. (2003). Promoting English language development and the reading habit among students in rural schools through the Guided Extensive Reading program. *Reading in a Foreign Language*, 15(2), 83-102.
- Beglar, D., Hunt, A., & Kite, Y. (2012). The effect of pleasure reading on Japanese EFL learners' reading rates. *Language Learning*, 62, 665–703.
- Brown, D. (2009). Why and how textbooks should encourage extensive reading, *ELT Journal*, 63(3), 238-245.
- Bryan, S. (2013). Extensive Reading, Narrow Reading and second language learners: implications for libraries, *The Australian Library Journal*, 60(2), 113-122.
- Cho, K. S., Kyoung, O. A., and Krashen S. (2005). The effects of narrow reading of authentic texts on interest and reading ability in English as a Foreign Language. *Reading Improvement*, 42(1): 58-63.
- Chun, E., Choi, S., Kim, J. (2012). The effect of extensive reading and paired-associate learning on long-term vocabulary retention: An event-related potential study, *Neuroscience Letters*, 521, 125-129.
- Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In *Second language vocabulary acquisition: A rationale for pedagogy*, edited by Coady, J. and Huckin, T. Cambridge: Cambridge University Press.
- Coady, J, Huckin T. N. (1997). *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, Cambridge University Press.
- Collins, M. F. (2010). ELL preschoolers' English vocabulary acquisition from storybook reading, *Early Childhood Research Quarterly*, 25, 84-97.
- Constantino, R., Lee, S., Cho, K. S., & Krashen, S. (1997). Free voluntary reading as a predictor of TOEFL scores. *Applied Language Learning*, 8, 111–118.
- Day, R. R., & Bamford, J. (1998). *Extensive reading in the second language classroom*. Cambridge, England: Cambridge University Press.
- Day, R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, 14(2), 136-141

- Eckerth, J., Tavakoli P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading, *Language Teaching Research*, 16(2), 227–252
- Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpus-based investigation of narrow reading. *Reading in a Foreign Language*, 20(1), 92-122.
- Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge, England: Cambridge University Press.
- Grabe, W., & Stoller, F. L. (2002). *Teaching and researching: Reading*. Pearson Education.
- Hafiz, F. and Tudor, I. (1989). Extensive reading and the development of language skills. *English Language Teaching Journal*, 43, 4--13.
- Hafiz, F. M., & Tudor, I. (1990). Graded readers as an input medium in L2 learning. *System*, 18, 31–42.
- Hitosugi, C. I., & Day, R. R. (2004). Extensive reading in Japanese. *Reading in a Foreign Language*, 16(1), 20-39.
- Horst, M. (2005). Learning L2 vocabulary through extensive reading: a measurement study. *The Canadian Modern Language Review*, 61, 355–382.
- Horst, M., Cobb, P. Meara, (1998). Beyond a Clockwork Orange: acquiring second language vocabulary through reading, *Reading in a Foreign Language*, 11. 207–233.
- Hulstijn, J.H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80, 327–339.
- Iwahori, Y. (2008). Developing reading fluency: a study of extensive reading in EFL. *Reading in a Foreign Language*, 20(1), 70-91.
- Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Pergamon Press, Oxford.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: additional evidence for the input hypothesis, *Modern Language Journal*, 73. 440–464.
- Krashen, S. (1993). *The Power of Reading: Insights from the Research*. Libraries Unlimited, Englewood, CO.
- Lao, C. Y., & Krashen, S. (2000). The impact of popular literature study on literacy development in EFL: More evidence for the power of reading. *System*, 28, 261–70.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In *Second language vocabulary acquisition: A rationale for pedagogy*, edited by Coady, J. and Huckin, T. Cambridge: Cambridge University Press.
- Laufer, B., Rozovski-Roitblat, B. (2014). Retention of new words: Quantity of encounters, quality of task, and degree of knowledge, *Language Teaching Research, Special Issue: Vocabulary research and pedagogy*, 1–25
- Lin, C. C. (2014). Learning English reading in a mobile-assisted extensive reading program, *Computers & Education*, 8, 48-59.
- Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading*, 32(2), 199-214.
- Mason, B., & Krashen, S. (1997). Extensive reading in English as a foreign language. *System*, 25, 91–102.

- Maxim II, Hiram H. (2002). A study into the feasibility and effects of reading extended authentic discourse in the beginning German language classroom *The Modern Language Journal*, 86(1):20-35.
- Nakanishi, T. (2014). A Meta-Analysis of Extensive Reading Research, *TESOL Quarterly*, 49(1), 6-37
- Nakanishi, T., & Ueda, A. (2011). Extensive reading and the effect of shadowing. *Reading in a Foreign Language*, 23(1), 1-16.
- Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. New York, NY: Routledge.
- Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- 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
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18, 1–28.
- Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language Learning*, 53, 233–84.
- Richards, J.C., Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics* (3rd ed). Pearson Education, London.
- Robb, T. N. and Susser, B. (1989). Extensive reading vs. skills building in an EFL context. *Reading in a Foreign Language*, 5, 239–251.
- Rodrigo, V., Greenberg, D., Burke, V., Hall, R., Berry, A., Brinck, T., Joseph, H., & Oby, M. (2007). Implementing an extensive reading program and library for adult literacy learners. *Reading in a Foreign Language*, 19(2), 106-119.
- 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, 589–619.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–63.
- Shin, K. (2003). Vocabulary acquisition through extensive reading in EFL class. *Modern English Education*, 4(2), 3-16.
- Stoeckel, T., Reagan, N., Hann, F. (2012). Extensive Reading Quizzes and Reading Attitudes, *TESOL Quarterly*, 46, 187–198.
- Storey, C., Gibson, K., & Williamson, R. (2006). Can extensive reading boost TOEIC scores? In K. Bradford-Watts, C. Ikeguchi, & M. Swanson (Eds.), JALT 2005 conference proceedings, 1004–1018. Tokyo, Japan: JALT.
- 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, 130–63.
- Webb, S. (2007). The Effects of Repetition on Vocabulary Knowledge, *Applied Linguistics*, 28(1). 46–65
- Yamashita, J. (2008). Extensive reading and development of different aspects of L2 proficiency. *System*, 36(4), 661-672.
- Yang, A. (2001). Reading and the non-academic learner: a mystery solved. *System*, 29, 451–466.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness, *Canadian Modern Language Review*, 57, 541–572.