

Textbook Influence on EFL Learners' Vocabulary Development

SAKATA Naoki*

Abstract

Textbook vocabulary has been shown to be quite influential on learners' vocabulary development in EFL environments because of its irreplaceable role. Conversely, the selection of words in textbooks has been criticized due to apparent arbitrariness, leading to insufficient coverage of generally frequent words. Concurrently, it has also been suggested that textbook influence on vocabulary learning is limited since other sources for vocabulary learning are available as well. To disentangle the seemingly complex textbook effect on vocabulary learning, this study investigated how much textbook vocabulary affects learners' knowledge, comparing the frequency effects of textbooks and a general corpus. The results showed that university students were rather affected by textbook frequencies than general frequencies; the students were found to be less familiar with words that were generally frequent but did not often appear in the textbooks.

1. Introduction

Textbooks play an important role in foreign language learning. Learners do not usually know what should be learned and textbooks often lead them to points to learn among many complicated linguistic elements in accordance with the learning stage. With little doubt, this notion applies to vocabulary selection in textbooks. Textbooks are supposed to guide learners to useful words and learners are likely to think that the words in them are worth learning. They rarely meet words that do not appear in them. Hence, textbook words should be useful and usable for learners. Namely, the words to be learned should be more compatible with learners' communication purposes (in spoken or written language) and easy enough to comprehend or use for the learners. Conversely, textbooks have not been found to be ideal in vocabulary selection.

1.1 Textbook Vocabulary

Regarding EFL textbooks in general, Hsu (2009) investigated words occurring in 36 English university textbooks and found that the textbooks for similar proficiency

*Associate Professor, Institute of Foreign Language Education, Kurume University
(Part-time Lecturer, Fukuoka University)

learners used words with similar frequency bands (e.g., up to 4,000 words for low-intermediate learners) in the British National Corpus (BNC). However, some lacked the proper amount of Academic Word List (AWL) (Coxhead, 2000) words, and the number of new words to learn in each textbook was quite diverse. In one textbook, students only met 162 new words, whereas they encountered as many as 2,001 new words in another. These facts show that the textbooks are not properly refined for university use for students to learn proper academic and new words.

Mármol (2011) analyzed vocabulary used in an EFL course book for Spanish elementary school students with the ranking of words in the BNC. He found that 28% of the words were not within the 2,000 most frequent words in the BNC. Furthermore, words facilitating communicative activities in the book were more prioritized than those that were more frequent in general. Those activities are no doubt important for learning practical English, but it may need to be recognized that meaningful communication is not possible without a suitable amount of basic vocabulary. Catalan and Francisco (2008) compared two EFL textbooks for Spanish secondary education and found that only 18 words were shared in the 50 most frequent words of both textbooks. Furthermore, some words did not occur frequently or were repeated at improper intervals. Koosha and Akbari (2010) analyzed seven English textbooks for Iranian junior and senior high school students by the RANGE (Range software download, 2018), which shows the BNC frequencies of words in a text, and found that words appearing in the textbooks covered only 15.4% of the word families of the BNC's first 3,000 words, and 7% of them were outside the first 3,000 words. Alsaif and Milton (2012) found that 15% of the most frequent 2,000 words in the RANGE were not covered in the collated vocabulary of 22 secondary-school EFL textbooks prescribed by the Ministry of Education of Saudi Arabia. Furthermore, only 2,800 out of the most frequent 5,000 words in the RANGE appeared in the textbooks, whereas 1,000 words not included in them were used. They also noted that new words were not introduced evenly among learning stages. The students met 931 new words in the intermediate stage but 263 at the second (advanced) stage. These three studies showed that the words in the English textbooks were not chosen carefully in terms of frequency.

Regarding EFL textbooks for Japanese high school students, which are targeted in the present study, Chujo (2004) showed that the 3,000 highest frequency words in the lemmatized BNC word list covered 95% of the token of a series of English textbooks for the secondary, six-year education in Japan, indicating that the textbooks used relatively high-frequency words. On the other hand, the token of 34,026 words only included 2,443

types out of the top 3,000 in the BNC (i.e., 557 types, about 19%, of them did not appear), showing that the textbooks may need to contain more generally frequent words. Hasegawa, Chujo, and Nishigaki (2008) estimated that between 2,454 and 3,250 lemmas are learnable throughout the six-year secondary education with the vocabularies found in the most used English textbook series in junior high school and one of the five upper-level English textbook series used extensively in high school. These numbers, however, are not sufficient for listening or reading TOEIC, TOEFL, VOA, or TIME. It was found that they did not cover 95% of any token of those. For example, they covered, on average, only 92.0% of the 6,000 words that appeared in the listening sections of two TOEFL tests and 80.7% of the 7,500 words (five 1,500-word extracts) in TIME. This indicates that the textbooks should have included more practical words.

1.2 Textbook Influence on Vocabulary Learning

Although teaching vocabulary directly in the classroom was not found to have much influence on the retention in some studies (e.g., Harris & Snow, 2004), and it has been suggested by Ellis (1994) that most second language vocabulary may be acquired incidentally and from oral input, textbook input does not seem to be negligible in the EFL environment since the students do not have much opportunity to be exposed to the language outside the classroom. Indeed, several studies have shown that textbooks or reading activities have at least some effect on vocabulary uptake. Milton and Vassiliu (2000) showed that Greek learners of English acquired much of the textbook vocabulary they were presented with. Vidal (2011) found that low-proficiency learners gained and retained more vocabulary knowledge through reading than listening.

To further examine textbooks' influence on vocabulary learning, this study compared the frequency effects of textbooks and a general corpus on learners' knowledge. Words in a textbook occur with different frequencies, in contrast to those in a word list, and studies have shown that lexical frequency affects learners' initial vocabulary learning (Eckerth & Tavakoli, 2012) and has an impact with word-focused exercises (Laufer & Rozovski-Roitblat, 2011). The number of encounters necessary for the acquisition of word meaning has been suggested as eight (Horst, Cobb, & Meara, 1998), ten, or more (Saragi, Nation, & Meister, 1978; Webb, 2007) and a study has found that more occurrences are needed for learners with smaller vocabulary sizes (Zahar, Cobb, & Spada, 2001). Consequently, words more frequent in textbooks seem to be more acquirable. This process is fine as long as words frequently occurring in textbooks are worth learning, but some textbook-specific words, which are not generally useful, may also be learned

unexpectedly over words generally frequent but infrequent in textbooks. Concurrently, it is also probable that learners may utilize other sources such as word books or lists to complement what lacks in textbooks and textbook influence is limited. In such case, words useful but infrequently occurring in textbooks should be learned in other material and textbook influence may be weakened. Hence, comparing the frequency effects of textbooks and a general corpus may measure how much textbooks affect vocabulary learning.

2. Method

2.1 Material

Among the variety of foreign language textbooks, this study focused on high school textbooks published for Japanese EFL learners. A vocabulary test was created with a corpus compiled from 24 government-authorized textbooks published in 2013 for Japanese high schools. Their compatibilities with the government's curriculum guidelines (MEXT, 2014) were checked by the Ministry of Education, Culture, Sports, and Technology of Japan (MEXT). Three-year (senior) high school education is part of the six-year secondary education in Japan. The 24 textbooks from the three publishers consisted of eight "English I" textbooks for the first year of high school, nine "English II" textbooks for the second year, and seven "English Reading" textbooks for the third year. All books were top sellers and variant in targeted students' proficiency so as to correspond to high school students' actual diversity in English proficiency. As most students used "English I," "English II," and "English Reading" or "English I" and "English II" in the curriculum, the corpus was supposed to represent the input for the high school students at the time. Speaking and writing textbooks, which the corpus did not use as its source, were used less often.

The textbooks were scanned and converted into PDFs by a company specializing in the field, from which text files were extracted by the author. A corpus and its lemmatized word list were compiled from the text files with Antconc (Anthony, 2014) and E-Lemma (Someya, 2014). Although some studies cited in the introduction used word-family, lemma was used as the unit of counting words in this study since English learners using textbooks as the main source of input are not usually very proficient and cannot easily associate members of a word family (Gardner, 2007). Using word-family instead could have overestimated the usefulness of textbook vocabulary: an occurrence of a word belonging to a word-family may not lead to learning other words in the same family. As the raw list of 36,912 lemmas contained noise, the top 60,000 lemma list of COCA (Word

frequency data, 2014) was used to identify 7,678 lemmas in the list, which accounted for 59.9% of the token of the textbook corpus (see Table 1). Refinement of the list through the analysis of the other components was considered unnecessary because the purpose of making the list was to know the overall dispersion of textbook words in terms of frequency; grasping accurate frequencies of the words was not the goal. Moreover, among the rest of the components, which accounted for around 40% of the whole token, one-letter items and items occurring only once composed 21.8% and 2.6%, respectively, both of which were not probably words that would interfere with the list above. Although some of the remaining 8,469 types initially counted as noise, constituting a token of 122,277 words, and could be identified as words and alter the list mentioned above slightly with more elaborative analysis, overall tendency of vocabulary usage in the textbooks would not be so different, the author thought, considering the ratio of 6 (initially identified as lemmas in COCA) and 1.6 (initially analyzed as noise).

Table 1.

Proportions of Identified Lemmas and Other Items in the Textbook Corpus

	Types	Tokens	Tokens in %
Lemma in COCA top 60,000	7,678	470,240	59.9
One-letter	21	171,135	21.8
One occurrence	20,744	20,744	2.6
Others	8,469	122,277	15.6
Total	36,912	784,396	100.0

Based on the frequency counts and those of the COCA, a multiple-choice test of 48 words, formatted for the test taker to choose the correct definition of an English word out of four choices written in Japanese, was compiled (see Appendix C for the tested words and their frequencies in the corpora and Appendix D for the format). The 48 words consisted of four groups: (a) frequent both in the textbooks and in the COCA; (b) frequent

in the textbooks but not in the COCA; (c) not frequent in the textbooks but frequent in the COCA; (d) not frequent either in the textbooks or in the COCA. Each group comprised 12 words: three nouns, three verbs, three adjectives, and three adverbs. The average frequency rankings in the textbooks and in the COCA were 793 and 763 in condition (a), 1,151 and 3,637 in (b), 4,027 and 1,176 in (c), and 4,042 and 3,996 in condition (d), respectively. By an ANOVA and a post hoc Bonferroni test, any difference between an average frequency ranking called “frequent” and one called “infrequent” was significant, $F(7, 88) = 295.3, ps < .05$, but no difference was found ($ps > .05$) among the average frequencies that belonged to the same frequency group (frequent or infrequent), regardless of the kinds of corpus the rankings were from.

2.2 Participants and Procedure

Ninety-four students of low-intermediate to intermediate proficiencies from two universities in Japan participated in the survey. The participation was not financially compensated. The data of one participant was excluded because she did not notice the double-sided nature of the test sheet and did not complete the latter half of the materials. The test was administered during regular class time and took about 15 minutes. The test sheets were scanned and scored with OMR software.

2.3 Analysis

A Generalized Linear Mixed Model (GLMM)—which can include both fixed effects of independent variables as traditional methods (e.g., ANOVA) and random-effects yielded by sampling participants or items so that errors to skew results can be reduced (see Quené & van den Bergh, 2008, for details)—analysis with Laplace approximation was conducted to see how the frequency bands (high and low) in the two corpora affected vocabulary knowledge. The score of each item was binary, so a logit link function was used for addressing the skewness of binomial distribution. The scores were assigned as the dependent variable, the frequency band categories as fixed effect, and the participants by university and the items by part of speech as random effects.

3. Results

The frequency band categories were found to significantly affect the vocabulary knowledge of participants, $F(3, 44) = 10.33, p < .01$. Post hoc multiple comparisons by Bonferroni tests showed that the lemmas more frequently appearing in the textbook corpus were better known by participants than the lemmas less frequently occurring in

the corpus, regardless of the frequency in the COCA (i.e., (a) (b) > (c) (d) in the grouping above; see Table 2).

Table 2.

Accuracy Rates and t Values of Multiple Comparisons Adjusted with Bonferroni

	Accuracy Rate	t Value		
		(b)	(c)	(d)
(a)	0.89	0.87	4.18**	4.47**
(b)	0.83		3.31*	3.60**
(c)	0.59			0.29
(d)	0.57			

** $p < .01$, * $p < .05$

4. Discussion

The purposes of this study was to investigate textbook vocabulary's influence on students. The test results showed that vocabulary in the textbooks actually affected the knowledge of students. At least in an EFL environment, where participants of this study learned the language, the frequencies of textbook words had a substantial influence. The results are also in line with the study by Zahar et al. (2001), in which less proficient learners were more affected by the frequencies of words they learned. Less proficient students as the participants of this study do not usually know which words are useful, so they often rely on textbooks for reference and work diligently to learn the words there.

Textbook writers may believe that learners acquire vocabulary not only from textbooks and they can acquire words from other sources, such as word learning books. However, the results showed that textbooks are actually very influential on acquisition. Many high school students learn English not only for improving their language skills but also for preparing for entrance exams of universities. The exams are often created based on high school curriculum and tend to use words more frequent in high school textbooks because using unfamiliar words will not lead to an accurate measurement of students' English proficiencies. In effect, word learning books targeted at entrance examinations

are apt to select words frequently used in the textbooks. This way, vocabulary learning of high school students may be closely connected to words frequently used in textbooks.

Since textbook vocabulary was found to affect rather more significantly than general English usage as represented by the COCA frequencies, textbook writers may need to use objective methods to select words for use; their selection of words may be substantial for constituting learners' vocabulary. Nowadays, plenty of data such as corpora and word lists, which reflect general usage of English, are available. Frequency counts of basic words are becoming easier to obtain on the Internet. These methods are especially useful when they rewrite texts originally written for L1 readers. Observing objective data, they can replace infrequent words with more useful, frequent, and educationally important ones. Knowing what kinds of words are used in a text is also available with text analysis methods, using computer software such as Antconc (Anthony, 2014) or the RANGE (Range software download, 2018). This way, the nature of textbook words may be able to improve in the future.

5. Conclusion

Alsaif and Milton (2012) suggested that although EFL students have many opportunities to be exposed to English materials through the Internet these days, textbooks may still remain the main source for the input of English. This assumption may be strongly supported when teaching materials are heavily recycled for certain purposes, such as preparing for entrance examinations. This study found that textbook vocabulary was indeed very influential; the vocabulary knowledge of participants of the present study was affected by the frequencies of textbooks, rather than by those of a general corpus. These results may apply to other EFL settings, especially where textbook vocabulary tends to be recycled repeatedly for certain reasons (e.g., entrance examinations).

Acknowledgement

This paper is based on the survey conducted by a research group of JACET English Vocabulary SIG. The author greatly appreciates contributions by the other members, Prof. Kenji Tagashira (Tokyo Kasei University), and Prof. Masamichi Mochizuki (Reitaku University). The study was supported by JSPS KAKENHI Grant Number JP25770214.

References

- Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal*, 40, 21–33.
doi:10.1080/09571736.2012.658221
- Anthony, L. (2014, July 25). AntConc (Version 3.4.4) [Windows]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/>
- Catalán, R. M. J., & Francisco, R. M. (2008). Vocabulary input in EFL textbooks. *Revista española de lingüística aplicada*, 21, 147–166.
- Chujo, K. (2004). Measuring vocabulary levels of English textbooks and tests using a BNC lemmatised high frequency word list. *Language and Computers*, 51, 231–249.
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34, 213–238.
- 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, 227–252.
doi:10.1177/1362168811431377
- Ellis, R. (1994). *The study of second language acquisition*. Oxford, England: Oxford University Press.
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28, 241–265.
doi:10.1093/applin/amm010
- Harris, V., & Snow, S. (2004). *Doing it for themselves: focus on learning strategies and vocabulary building*. London, England: CILT, the National Centre for Languages.
- Hasegawa, S., Chujo, K., & Nishigaki, C. (2008). Examining the utility of junior and senior high school English textbook vocabulary. *Journal of the College of Industrial Technology, Nihon University*, 41, 49–56.
- 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, 126–149.
- 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.
- Koosha, M., & Akbari, G. (2010). An evaluation of the vocabulary used in Iranian EFL secondary and high school textbooks based on the BNC first three 1000 high

- frequency word lists. *Research in Curriculum Planning: A Quarterly Journal of Science and Research*, 1, 157–186.
- Laufer, B., & Rozovski-Roitblat, B. (2011). Incidental vocabulary acquisition: the effects of task type, word occurrence and their combination. *Language Teaching Research*, 15, 391–411. doi:10.1177/1362168811412019
- Mármol, G. A. (2011). Vocabulary input in classroom materials: two EFL coursebooks used in Spanish schools. *Revista Española de Lingüística Aplicada*, 24, 9–28.
- MEXT. (2014, June 14). Koutougakkou gakushu-sidou youryou [The curriculum guidelines for high school]. Retrieved from http://www.mext.go.jp/a_menu/shotou/cs/1320221.htm
- Milton, J., & Vassiliu, P. (2000). Frequency and the lexis of low-level EFL texts. In *Proceedings of the 13th International Symposium on Theoretical and Applied Linguistics* (pp. 444–55). Thessaloniki, Greece: Aristotle University.
- Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59, 413–425. doi: 10.1016/j.jml.2008.02.002
- Range software download. (2018, April 25). Retrieved from <https://www.victoria.ac.nz/lals/resources/range>
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6(2), 72–78. doi:10.1016/0346-251x(78)90027-1
- Someya, Y. (2014, July 25). E-Lemma. Retrieved from http://www.lexically.net/downloads/e_lemma.zip
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61, 219–258. doi:10.1111/j.1467-9922.2010.00593.x
- Word frequency data. (2014, August 23). Retrieved from <https://www.wordfrequency.info/purchase1.asp?i=w60a>
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: effects of frequency and contextual richness. *Canadian Modern Language Review*, 57, 541–572. doi:10.3138/cmlr.57.4.541

Appendix A

The Words Tested

	Ranking in the Textbook Corpus	Ranking in the COCA
Group (a)		
weight	980	955
difference	548	496
design	830	835
cause	510	601
enter	728	709
contain	966	932
common	648	710
similar	889	720
huge	1056	912
rather	807	861
quickly	673	679
exactly	886	748
Group (b)		
prize	732	3714
temple	761	3642
clerk	918	3907
lend	956	3574
injure	1153	3749
praise	1294	3747
instant	998	4493
loud	1319	3113
artificial	1470	4034
abroad	1236	3385
mainly	1400	2822

greatly	1572	3459
Group (c)		
decade	3654	730
management	4389	907
firm	3683	903
commit	4652	1398
reveal	3801	928
extend	3442	1449
financial	3682	828
broad	4230	1408
previous	4445	1280
relatively	4988	1466
otherwise	4067	1532
obviously	3285	1282
Group (d)		
hallway	3982	3585
possession	4442	3619
instinct	3716	4274
modify	3750	3713
resume	3800	3810
punish	4092	3892
cooperative	4261	4285
influential	4355	4327
precise	4086	3856
hopefully	3710	4043
desperately	4282	4490
lately	4023	4060

Appendix B

The Test Format with the Translations of Choices (abridged)

Words		Choices			
injure	1 拒む (reject)	2 傷つける (injure)	3 発表する (present)	4 組み立てる (build)	
modify	1 修正する (modify)	2 扱う (treat)	3 達成する (achieve)	4 あざける (mock)	
instant	1 頭の良い (clever)	2 軍隊の (military)	3 激しい (fierce)	4 即時の (instant)	
punish	1 見落とす (overlook)	2 罰する (punish)	3 感謝する (thank)	4 打ち明ける (confess)	
weight	1 郊外 (suburb)	2 依頼 (request)	3 締め切り (deadline)	4 重さ (weight)	
previous	1 中央の (central)	2 残酷な (cruel)	3 前の (previous)	4 鈍い (dull)	
clerk	1 科学 (science)	2 事務員 (clerk)	3 屈服 (submission)	4 荷物 (luggage)	
mainly	1 頻繁に (frequently)	2 厳しく (strictly)	3 主に (mainly)	4 何とか (somehow)	
management	1 土地 (land)	2 苦悩 (distress)	3 聴衆 (audience)	4 管理 (management)	
financial	1 活発な (active)	2 財政上の (financial)	3 真剣な (serious)	4 無駄な (wasteful)	
...					