LEXICAL LOADS OF MATHEMATICA DISCOURSE FOR YOUNG LEARNERS: A STEP TOWARDS VOCABULARY EVALUATION OF MULTI-SEMIOTIC DISCOURSE

Ton Nu My Nhat*

Quy Nhon University

Received: 20/09/2017; Revised: 26/10/2017; Accepted: 27/12/2017

Abstract: This study was an attempt to examine the vocabulary demands of the mathematic discourse (MD) written for young learners (YLs). The data for this research were two sets of books - one for the Vietnamese learners of English as a foreign language (L2) and one for the Singaporean learners of English as a first language (L1). To find out more about the lexical profiles of this genre, a total of 1,729 mathematic problems from two series of four books for primary school children, consisting of 15,545 running words, were analyzed to determine the vocabulary size necessary for comprehension and the potential to learn vocabulary incidentally through doing mathematics in English. The article concludes with a discussion of pedagogical implications of this study for material designers and teachers of MD for YLs.

Key words: lexical coverage, mathematic discourse, multisemiotic discourse, vocabulary size, word frequency

1. Introduction

Comprehension research has shown that besides the other factors that may have an impact on reading and listening comprehension such as background knowledge, syntactic structures, and/or discourse structure, vocabulary proves to be the most influential (Laufer & Sim 1985, Webb & Rodgers 2009). Studies of lexical coverage have indicated that there is typically a positive correlation between vocabulary size and degree of comprehension (Laufer, 1989, 1992; Laufer & Ravenhorst, 2010): comprehension is likely to increase as the proportion of known words in a text rises. The justification for this is that the fewer words within a text there are, the fewer comprehension gaps follow and the better understanding is achieved (Webb & Paribakht, 2014).

Surprisingly, whereas the strong correlation between coverage and comprehension is extensively documented by corpus-driven research, there has been little research focusing on the lexical profile of multimodal texts, of which language constitutes only one component. Our present study aims to shed light on this issue. The aim of this study is to examine the lexical profiles of mathematics discourse for young learners (MDYL). Specifically, it determines the vocabulary demands of MDYL and investigates the potential for incidental vocabulary learning through doing mathematics in English. By doing this, the present research may shed light on the target vocabulary size necessary for adequate comprehension of MDYLs, providing useful data for researchers, material designers, teachers, and learners who are concerned about a vocabulary threshold and goal for English-medium mathematics courses for YLs. Knowing how often words are encountered provides some indication of the potential for incidental vocabulary learning through engaging in this genre, which may lead to cumulative growth in vocabulary knowledge.

* Email: tnmynhat70@gmail.com

2. Background

2.1. Vocabulary size and comprehension

Over the last thirty years, corpus-driven research examining lexical profiles and coverage has painted a clearer picture of the lexical demands of a wide range of genres. Lexical coverage refers to the percentage of words in a text covered by items from a particular word list (Nation & Waring, 1997). It thus reveals the proportion of words in the discourse a reader needs to know with reference to a word list for adequate comprehension to occur (Dang & Webb, 2016). Studies have examined the number of words necessary for comprehension of both spoken discourse, such as movies (Nation, 2006; Webb & Rodgers, 2009) and television programs (Rodgers & Webb, 2011) and written discourse such as novels (Hirsh & Nation, 1992), comic books (Meara, 1993), and graded readers (Nation, 2006; Wodinsky & Nation, 1988). Research has provided evidence of a comparatively linear relationship between the two variables. However, L2 studies seem slightly varied in the amount of text coverage that is need for comprehension to occur.

The coverage necessary for comprehension are likely to vary between discourse types and the degree of understanding accounted - poor, adequate, or reasonable. Employing word lists developed from the British National Corpus (BNC), Nation (2006) investigated the vocabulary size necessary for reading several different types of discourse. Nation found that learners would need to know 6000-7000 word families to have the ideal 98% coverage of spoken texts, and 8000-9000 word families for written texts. Nation (2006) also states that for both native and non-native learners, high-frequency and wide-range words are largely learned before lower-frequency and narrower-range words. However, Webb & Paribakht (2014)'s study indicates that lexical profiles are likely to differ from text to text: an average knowledge of 2000 word families and proper nouns were sufficient to reach 96% coverage for one movie, but 3000 word families was sufficient to reach 95% coverage of movies; however for another, 4000 word families and proper nouns were required to reach this level of coverage. Laufer (1989) found that language learners had poor comprehension of an L2 academic text at 90% coverage of the words in the text, but reasonable understanding of that text at 95% coverage. Hu and Nation's (2000) study indicated that comprehension was poor when coverage ranged from 80% to 90% and understanding improved at 95% coverage; 98% coverage may provide an acceptable level of unassisted reading comprehension. Other studies of different types of written text have indicated that knowledge of the most frequent 3000 word families was required to reach 98% coverage of graded readers (Webb & Macalister, 2013), of 5000 word families to reach 97-98% coverage of three short novels aimed at teenage or younger readers (Hirsh & Nation, 1992).

Results of the studies of spoken discourse have been relatively consistent, indicating that knowledge of the most frequent 3000 word families was necessary to reach 95% coverage of television programs and movies (Rodgers & Webb, 2011; Webb & Rodgers, 2009a, 2009b).

Studies have also looked at text coverage and listening comprehension correlation (Bonk, 2000; Schmitt, 2008; Stæhr, 2009), and findings are generally varied. Bonk (2000)'s study of the listening comprehension of 59 Japanese university students of varying English proficiency levels suggests that a majority of the participants achieved 'good' comprehension at 90%. Schmitt (2008) found that the lexical knowledge may be close to 95% for an acceptable level

listening comprehension. Stæhr (2009) looked at the advanced listening comprehension of more than one hundred Danish learners of English and found that 98% coverage was required for comprehension of a listening test from the Cambridge certificate of proficiency in English (CPE).

Gardner (2004)'s and Macalister (1999)'s studies provide data on the coverage of the vocabulary in text written for children. Macalister found that approximately 85% coverage of texts written for children may be gained with a knowledge of the 2,000 most frequent word families. Similarly, Gardner indicated that the most frequent 3,000 word families accounted for around 89% of coverage. The findings further indicate that a rather large vocabulary size is necessary for children to reach 98% coverage. Regarding age and vocabulary correlation, in a research on texts written for children, Webb & Macalister (2013) found that, contrary to what might be expected, texts targeted at L1 children had a similar vocabulary demand to that of texts written for adults, and that graded readers exhibited a much higher percentage of high frequency words than both of the other types.

In terms of L2 learning, Cobb (2010) argues that although words are likely to be acquired in order of frequency in first language development, this assumption is not always the case when it comes to a second language. His own research with several groups of both school and adult learners in Quebec has provided evidence that these learners know as many words at a medium-frequency level (3k, 5k) as at a higher frequency level (1k, 2k).

Research investigating the lexical profiles of scientific texts is limited. Coxhead (2000) showed that academic learners will reach about 90% or more coverage of their subject-specific texts at a mere set of 570 mainly Greco-Latin word families of post-2,000 level frequency, in addition to knowledge of some technical items in their field.

Research has also revealed that although vocabulary might have the greatest impact on comprehension (Laufer & Ravenhorst, 2010; Laufer & Sim, 1985), comprehension of a text depends on many other factors, such as background knowledge (Leeser, 2007; Pulido, 2004), individual competences in reading skills (Mezynski, 1983), the ability to infer word meanings from context (Hulstijn, 1993), the lexical weight of unknown items in a text (Hulstijn, 1993), or the amount of circumstantial information in the text (Kameenui, Carnine & Freschi, 1982).

Within a genre, as Webb & Paribakht (2014) caution, "learners with a vocabulary size that is sufficient to understand one text may not have the same degree of comprehension of another text from the same discourse type".

2.2. Incidental vocabulary learning

Research has studied the potential for incidental vocabulary learning in various text types. Incidental vocabulary learning may be defined as "learning words without deliberate decision to commit information to memory" (Laufer & Hulstijn, 2001, p. 11). Findings have indicated that the potential for incidental learning of words is likely to increase as the number of encounters with them increases (Horst et al., 1998; Jenkins, Stein, & Wysocki, 1984; Rott, 1999; Waring & Takaki, 2003; Webb, 2007). However, the average number of encounters necessary for reliable retention of a new word is varied among studies, ranging from six (Rott, 1999) to twenty (Waring & Takaki, 2003). A single encounter rarely leads to learning. Factors that may affect retention are the spacing between encounters, the surrounding contexts, the proficiency of the learners (Webb, 2008; Webb & Macalister, 2013; Zahar, Cobb & Spada, 2001). Webb &

Macalister (2013) maintain frequent repetition of topic-related vocabulary may benefit young L1 and L2 learners who are likely to be typically involved in age-specific reading in lexical growth.

The potential for vocabulary learning from text written for children in English is not well-researched. Gardner (2004) compared the frequency of words in expository and narrative texts written for children. He found a greater rate of word repetition in the former, indicating that informative texts may lead to higher potential for incidental vocabulary learning. By contrast, Macalister (1999) found that imaginative rather than informative texts provided greater opportunity for incidental vocabulary learning. Webb & Macalister (2013) accounted the difference between the findings in the two studies for the characteristics of the corpora. They also indicate that as a large amount of children's reading material is aimed to promote vocabulary growth, repetition of less frequent words might be common. They then conclude that text written for children might be more beneficial as a source of incidental vocabulary learning than text written for older groups.

Research questions

The current study seeks to address the following research questions:

(1) What vocabulary size is necessary to reach 95% and 98% coverage of MDYLs?

(2) Do the two sets targeted at L1 and L2 learners have similar or different vocabulary profiles?

(3) How frequently are the word families of the texts analyzed encountered in each book and what is the recycling index across two grades within each set?

3. Methodology

3.1. Data

The books which served as a data of the present study comprise two sets targeted at primary school children. The first set consisted of two books published by Vietnam Education Publishing House - *Math ViOlympic 4* (Đang Minh Tuan & Nguyen Thi Hai, 2016; hereafter *MV 4*) and *Math ViOlympic 5* (Đang Minh Tuan & Nguyen Thi Bich Phuong, 2016; hereafter *MV 5*)*;* the second was two books published by Singapore Asia Publishers - *Learning Maths 1B* (Tan, 2016a; hereafter *LM 1B*) and *Learning Maths 2A* (Tan, 2016b; hereafter *LM 2A*). *MV 4* and *MV 5* are the only two published by a reliable publisher in Vietnam so far in this realm. From many series published by foreign publishers, these two books were chosen for analysis as these two are for the children of the same age groups as those targeted at in the first set. The numbers of problems and running words of the verbal texts in each book are shown in Table 1.

Table 1. Number of problems and words in individual books analyzed

Book	No. of Problems	Running words
Learning Maths 1B	381	3488
Learning Maths 2A	393	1589
Math ViOlympic 4	555	5578
Math ViOlympic 5	400	5141
Total	1,729	15,796

3.2. Data analysis

To achieve the aims, in the absence of the electronic versions of these books, the lexical components of all these 1,729 problems were typed and computerized. The raw data were manually processed to omit the proper nouns. This is because many researchers have taken the approach that proper nouns have a minimal learning burden and may be easily understood by readers (Nation, 2006); how proper nouns are handled makes a big difference to an output profile (Cobb, 2010). The symbolic components and numbers, which are inherent and pervasive of this multimodal genre, were also removed. The sets of data were then analyzed using *Compleat Lexical Tutor* developed by Tom Cobb (available at http://www.lextutor), using the BNC-20 wordlist. *VocabProfile* broke each corpora into its frequency levels according to the thousand-levels scheme, Academic and off-list words, indicated by colors and gives all the information regarding vocabularies of the data - the number of type, token, word families, type-token ratio, function and content words. *Frequency* extracted frequency lists from the corpora. *TextLexCompare* was used to tract the amount of vocabulary repetition across the books within each set.

4. Findings and discussion

It might seem that MDYLs might contain easy vocabulary, yet the analysis reveals otherwise. Table 2 and 3 summarize the data in terms of tokens, types, and families of the two corpora, *Learning Maths* and *Math ViOlympic,* respectively; the cumulative coverage for each book is shown in Table 4.

In answer to the first RQ (*What is the vocabulary necessary to reach 95% and 98% coverage of MDYLs*), the findings indicate that 95% and 98% is reached at around 3,000 - 4,000 and 6,000 - 7,000 respectively. This suggests that the vocabulary found in MDYLs is likely to be challenging for most language learners. Research indicates that a reasonable proportion of L2 learners in different contexts fail to learn the most frequent 2,000 and even the most frequent 1,000 after many years of formal instruction (Dang & Webb, 2016). The common finding is that many ESL learners tend to plateau with usable knowledge of about 2000 words families or less (Cobb, 2007).

Tables 2 and 3 show that the tokens are spread over the 20 most frequent 1,000 word families of the BNC. The importance of knowing the most frequent 1,000 word families is clearly demonstrated in the first rows of these three tables. The first 1,000 word families from the BNC account for up to approximately four-fifths of tokens in the problems in all these books - 76.29%, 84.02%, 84.06%, and 81.13%. For example, regarding *Math ViOlympic 4*, the first row indicates that 424 different word forms (types) are the source of these 4689 tokens. These 424 types reduce to 303 word-families. Similarly, as for *Learning Maths 2A*, the first 1,000 word families account for 1335 of the tokens, 223 of the types, and 173 of the families. It is useful to consider the output in terms of word families because similarity in forms and meanings for tokens from the same family may facilitate understanding and retention. It is also clear that after the second 1,000 word-families, the decreasing rate of the tokens tend to be approximately the same across the four books. From the third - 1,000 onwards, the word families thin out rapidly, which suggests that the number of low-frequency words is few and far between.

As shown in Table 5, it is also important to note that of these huge coverages of the first 1,000 word-families, the number of the function words tends to double that of the content words throughout the data.

The findings suggest that only a small vocabulary is needed for young learners to comprehend these mathematic problems. The number of word-families a learner would meet when s/he finished MV 4, MV 5, LM 1B, and LM 2A is 434^{+}, 343^{+}, 415^{+}, and 230^{+}, respectively.

The corpus was shown to contain not only a small number of word-families but also a high frequency rate of encounter of each word, which is strikingly similar across the two series. A small number of these word families are met from as high as 592 to six times (64.32%, 86.94%, 76.28%, and 70.35%). The overall and unexpected finding from a close analysis of the lists of frequency indicates that these soaring high percentages are typically represented by function words and technical words.

Table 2. Tokens, types, and families at each level in Learning Maths 1B and 2A

	LM 1B			***LM 2A***		
Word list (1,000)	**Tokens (%)**	**Types (%)**	**Families**	**Tokens (%)**	**Types (%)**	**Families**
1	2661 (76.29)	303 (57.71)	231 (55.66)	1335 (84.02)	223 (76.63)	173 (75.22)
2	415 (11.90)	101 (19.24)	80 (19.28)	153 (9.63)	37 (12.71)	31 (13.48)
3	35 (1.00)	19 (3.62)	16 (3.86)	29 (1.83)	7 (2.41)	6 (2.61)
4	161 (4.62)	32 (6.10)	27 (6.51)	16 (1.01)	7 (2.41)	6 (2.61)
5	75 (2.15)	20 (3.81)	18 (4.38)	8 (0.50)	5 (1.72)	5 (2.17)
6	59 (1.69)	14 (2.67)	12 (2.89)	33 (2.08)	3 (1.03)	2 (0.87)
7	40 (1.15)	14 (2.67)	14 (3.37)	6 (0.38)	3 (1.03)	2 (0.87)
8	4 (0.11)	4 (0.76)	4 (0.96)			
9	4 (0.11)	2 (0.38)	2 (0.48)	4 (0.25)	2 (0.69)	2 (0.87)
10	6 (0.17)	3 (0.57)	2 (0.48)			
11	6 (0.17)	3 (0.57)	3 (0.72)	4 (0.25)	3 (1.03)	3 (1.30)
12	1 (0.03)	1 (0.19)	1 (0.24)			
13	1 (0.03)	1 (0.19)	1 (0.24)			
14	2 (0.06)	1 (0.19)	1 (0.24)			
15						
16						
17	4 (0.11)	1(0.19)	1 (0.24)			
18						
19	4 (0.11)	2. (0.38)	2 (0.48)			
20						
Off-List	10 (0.29)	4. (0.76)	??	1 (0.06)	1 (0.34)	??
Total	3488 (100)	525 (100)	415+?	1589 (100)	291 (100)	230+?

Table 3. Tokens, types, and families at each level in Math ViOlympic 4 and 5

	MV 4			*MV 5*		
Word list (1,000)	**Tokens (%)**	**Types (%)**	**Families**	**Tokens (%)**	**Types (%)**	**Families**
1	4689 (84.06)	424 (70.78)	303 (69.82)	4171 (81.13)	290 (67.29)	226 (65.89)
2	482 (8.64)	92 (15.36)	72 (16.59)	529 (10.29)	77 (17.87)	65 (18.95)
3	109 (1.95)	23 (3.84)	22 (5.07)	105 (2.04)	17 (3.94)	15 (4.37)
4	51 (0.91)	17 (2.84)	11 (2.53)	120 (2.33)	16 (3.71)	11 (3.21)
5	78 (1.40)	11 (1.84)	8 (1.84)	51 (0.99)	11 (2.55)	9 (2.62)
6	86 (1.54)	6 (1.00)	4 (0.92)	72 (1.40)	6 (1.39)	5 (1.46)
7	5 (0.09)	4 (0.67)	2 (0.46)			
8				1 (0.02)	1 (0.23)	1 (0.29)
9	11 (0.20)	5 (0.83)	5 (1.15)	63 (1.23)	4 (0.93)	3 (0.87)
10	3 (0.05)	1 (0.17)	1 (0.23)	5 (0.10)	3 (0.70)	3 (0.87)
11	43 (0.77)	3 (0.50)	3 (0.69)	8 (0.16)	2 (0.46)	2 (0.58)
12						
13						
14						
15	1 (0.02)	1 (0.17)	1 (0.23)	8 (0.16)	1(0.23)	1 (0.29)
16				7 (0.14)	2(0.46)	2 (0.58)
17	1 (0.02)	1 (0.17)	1 (0.23)			
18	1 (0.02)	1 (0.17)	1 (0.23)			
19						
20						
Off-List	18 (0.32)	10 (1.67)	??	1 (0.02)	1 (0.23)	??
Total	5578 (100)	599 (100)	434+?	5141 (100)	431 (100)	343+?

Table 4. Cumulative coverage (%) for each book

Word list	***LM 1B***	***LM 2A***	***MV 4***	***MV 5***
1,000	76.29	84.02	84.06	81.13
2,000	88.19	93.65	92.70	91.42
3,000	89.19	**95.48**	94.65	93.46
4,000	93.81	96.49	**95.56**	**95.76**
5,000	**95.96**	96.99	96.96	96.78
6,000	97.65	**99.07**	**98.50**	**98.18**
7,000	**98.80**	99.45	98.59	
8,000	98.91			98.20
9,000	99.02	99.70	98.79	99.43
10,000	99.19		98.84	99.53
11,000	99.36	99.95	99.61	99.69
12,000	99.39			
13,000	99.42			
14,000	99.48			
15,000			99.63	99.85
16,000				99.99
17,000	99.59		99.65	
18,000			99.67	
19,000	99.70			
20,000				
Off-List	99.99	100.00	99.99	100.00
Tokens	≈100.00	≈100.00	≈100.00	≈100.00

By contrast, a substantial majority occur merely once or twice in each book (Table 6). Beyond the fifth 1,000 level, there are only a few words that occur in both sets. It should also be noticed tokens from this low-frequency group typically lie with everyday common vocabulary related to children's world, namely family, school, animals, and fruits. Therefore, it is possible to deduce from the findings that the chance for vocabulary growth of age-appropriate items via doing ME is minimal.

Table 5. K-1 sub-analysis in terms of content and function words for individual books

K1 Words	***MV 4***	***MV5***	***LM 1B***	***LM 2A***
Function words	59.27%	52.69%	46.40%	50.16%
Content words	27.54%	31.24%	31.17%	34.36%

Table 6. Number and percentage of encounters with word families (WF) in each book

	MV4		**MV5**		**LM 1B**		**LM 2A**	
	%	WF	%	WF	%	WF	%	WF
6 times & >	64.32	165	86.94	153	76.28	146	70.35	64
5-3 times	26.75	111	7.9	108	12.44	121	14.95	67
2-1 times	8.93	70	5.15	214	11.28	299	14.7	167

Table 7. Recyclying index over each set

	Math ViOlympic 4 & Math ViOlympic 5	***Learning Maths 1B &Learning Maths 2A***
Token	84.84%	74.94 %
Types	55.46%	49.47%

A further analysis by means of TextLexCompare yields the percentage of recycled vocabulary in each set of data, summarized in Table 7. The output shows that the recycling index does not go above 85% and 75% for Math ViOlympic and Learning Maths, respectively. This means that most words throughout the two successive books of each set are being met in density environments of around 3-4 unknown words in 10, which is much larger than the density that learners can handle. This result significantly supports the finding that there may be very little incidental vocabulary learning from doing ME for primary school children.

5. Conclusions

The data we have examined in this article yield the following conclusions. The cumulative coverage figures highlight the challenges that YLs may face when engaging in this text type due to insufficient vocabulary knowledge, which is a common case to EFL learners in various contexts. This study finds that the vocabulary sizes of 3,000-4,000 and 10,000-11,000 word needed to cover 95% and 98% of the verbal component, respectively, holds over the four books analyzed. The results of the present study also suggest that there is the potential for incidental vocabulary learning of the first 1,000, however, due to the small number of encounters with age-specific items, very few words are likely to be learned incidentally through doing mathematics in English.

Limitations and implications for further research

The first limitation lies with the treatment of proper names. Without proper names, the

texts are made to be likely more lexically challenging than they can really be. A further study to calculate proper nouns as a separate category and classify them as first 1,000 items needs to be considered to arrive at a more accurate description of the vocabulary loads of MDYLs.

The second limitation of this study is the size of the corpus. A larger corpus that includes materials currently being used in other English-speaking countries would certainly shed more light on the picture of vocabulary loads of this genre.

Thirdly, the findings indicate the vocabulary sizes necessary to reach 95% and 98% cumulative coverage were approximately 4000 and 11,000 word families, respectively. This analysis was, however, based on the most lexical-burdened conditions possible, with the symbolic elements excluded. My hunch is that these consistently high figure may be due to the multisemiotic nature of MD. When these problems also contain the integral components, namely symbolism and imagery, they may, in some cases, be comprehensible irrespective of a limited vocabulary knowledge. Coverage may be the most important factor in determining comprehension, but it is one of the many factors that are involved in comprehension (Webb & Macalister, 2013; Webb & Rodgers, 2009). MD depends on both intrasemiosis and intersemiosis. As the types of meaning made by each semiotic are fundamentally different, and thus the three semiotic resources fulfil individual functions, the success of mathematics depends on utilizing and combining the unique meaning potentials of language, symbolism and visual display in such a way that the semantic expansion is greater than the sum of meanings derived from each of the three resources. (Halloran, 2004, p. 16). A corpus-based approach including symbolism and/or visual images in the analysis to determine the degree to which these two semiotics may have an impact on comprehension would be a useful follow-up to this study. An experimental study comparing degree of understanding MDs that include symbolism and/or visual images and the same texts with the symbolism and/or visual images removed may also provide data on the effects of these factors. The results of this study of this multisemiotic discourse support the view that although coverage may be a very important factor in comprehension, it is only one of a number of factors that need to be considered in studies of comprehension. (Webb & Rodgers, 2009)

Research has provided evidence for non-linear profiles, especially in the early stage of learning a foreign language (Cobb, 2010). Cobb (2010)'s study suggests that the mixed profilers perform better with technical texts than with easy texts or conversations. His argument is that "if the goal is to read in a professional domain, then technical lexis is probably the shortest route to higher coverage" (Cobb, 2010). It is my hope that this study may serve as a starting point for developing a word list for EFL learners of mathematics. A word list of a manageable size, which could be glossed and/or pre-taught, would certainly pave the way for independent, unassisted engagement in MD. Amid the oceans of lexis that EFL learners may face, a specialized word list of items which are both frequent and useful may be of great value in helping the learners meet the initial challenge in content-language integrated learning that MD may present.

References

Bonk, W.J. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening, 14*(1), 14-31.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, *11*(3), 38-63.

Cobb, T. (2010). Learning about language and learners from computer programs. *Reading in a Foreign Language*, *22*(1), 181-200.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Đặng Minh Tuấn & Nguyễn Thị Bích Phượng (2016). *Math ViOlympic 5*. Hanoi: Vietnam Education Publishing House.

Đặng Minh Tuấn & Nguyễn Thị Hải (2016). *Math ViOlympic 4*. Hanoi: Vietnam Education Publishing House.

Dang Thi Ngoc Yen & Webb, S. (2016). Evaluating lists of high-frequency words. *ITL - International Journal of Applied Linguistics, 167*(2), 132-158.

Gardner, D. (2004). Vocabulary input through extensive reading: A comparison of words found in children's narrative and expository reading materials. *Applied Linguistics*, *25*(1), 1-37.

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.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*(2), 207-223.

Hu, M., & Nation, I.S.P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Hulstijn, J.H. (1993). When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables. *Modern Language Journal*, *77*(2), 139-147.

Jenkins, J.R., Stein, M.L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal*, *21*(4), 767-787.

Kameenui, E.J., Carnine, D.C., & Freschi, R. (1982). Effects of text construction and instructional procedures for teaching word meanings on comprehension and recall. *Reading Research Quarterly*, *17*(3), 367-388.

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

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint, & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). Basingstoke & London: Macmillan.

Laufer, B., & Ravenhorst-Kalovski, G.C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15-30.

Laufer, B., & Sim, D.D. (1985). Taking the easy way out: Non-use and misuse of clues in EFL reading. *English Teaching Forum*, *23*(2), 7-10.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26.

Leeser, M.J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning*, *57*(2), 229-270.

Macalister, J. (1999). School Journals and TESOL: An evaluation of the reading difficulty of School Journals for second and foreign language learners. *New Zealand Studies in Applied Linguistics*, *5*, 61-85.

Meara, P.M. (1993). Tintin and the World service: A look at lexical environments. *IATEFL: Annual Conference Report*, 32-37.

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, *53*(2), 253-279.

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

Nation, I.S.P. (2001) How many high frequency words are there in English? In M. Gill, A.W. Johnson, L.M. Koski, R.D. Sell & B. Wårvik (Eds.), *Language, learning and literature: Studies presented to Håkan Ringbom English Department Publications 4* (pp. 167-181). Åbo: Åbo Akademi University.

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

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

O'Halloran, K.L. (2004). *Mathematical discourse - language, symbolism and visual images.* London: Continuum.

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *54*(3), 469-523.

Rodgers, M.P.H., & Webb, S. (2011). Narrow viewing: The vocabulary in related television pro-grams. *TESOL Quarterly*, *45*(4), 689-717.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition*, *21*(1), 589-619.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329-363.

Stæhr, L.S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31*(4), 577-607.

Tan, A. (2016a). *Learning Maths - 1B*. (Bilingual version). Singapore Asia Publishers.

Tan, A. (2016b). *Learning Maths - 2A*. (Bilingual version). Singapore Asia Publishers.

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), 130-163.

Webb, S. & Paribakht, T.S. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes, 38*(1), 34-43.

Webb, S. (2007). The effect of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46-65.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(2), 232-245.

Webb, S. (2010). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics*, *15*(4), 497-519.

Webb, S., & Macalister, J. (2013). Is text written for children appropriate for L2 extensive reading? *TESOL Quarterly*, *47*(2), 300-322.

Webb, S., & Rodgers, M.P.H. (2009a). The lexical coverage of movies. *Applied Linguistics*, *30*(3), 407-427.

Webb, S., & Rodgers, M.P.H. (2009b). Vocabulary demands of television programs. *Language Learning*, *59*(2), 335-366.

Wodinsky, M., & Nation, I.S.P. (1988). Learning from graded readers. *Reading in a Foreign Language*, *5*(1), 155-161.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, *57*(4), 541-573.

ĐỘ KHÓ TỪ VỰNG TRONG THỂ LOẠI TOÁN DÀNH CHO LỨA TUỔI TIỂU HỌC: BƯỚC TIẾP CẬN ĐÁNH GIÁ TỪ VỰNG TRONG DIỄN NGÔN ĐA THỨC

Tóm tắt: Mục đích của công trình nghiên cứu này là đánh giá những yêu cầu về từ vựng trong thể loại diễn ngôn toán học dành cho lứa tuổi tiểu học. Dữ liệu phân tích là hai bộ sách Toán bằng tiếng Anh: một bộ viết cho trẻ em Việt Nam học tiếng Anh như một ngoại ngữ và một bộ viết cho trẻ em Singapore học tiếng Anh như ngôn ngữ chính. Khối liệu bao gồm tất cả 1.729 bài toán, với tổng số 15.545 từ, được phân tích để xác định khối lượng từ vựng cần thiết để hiểu các bài toán và tiềm năng hỗ trợ phát triển từ vựng khi trẻ em học làm toán bằng tiếng Anh. Bài báo kết thúc với một số ý nghĩa ứng dụng đối với việc soạn sách và dạy toán tiếng Anh cho lứa tuổi tiểu học.

Từ khóa: kiến thức từ vựng, diễn ngôn đa thức, diễn ngôn toán học, tần số sử dụng của từ vựng