MATERIAL DEVELOPMENT: A CORPUS-BASED APPROACH OF READING ENGLISH NEWSPAPER FOR EFL STUDENTS

[1]ANCHALEEVEERACHAISANTIKUL, [2]WARACHANSIN

[1]English for International Communication Department, the Faculty of Sciences and Liberal Arts,
Rajamangala University of Technology Isan, Nakhonratchasima, Thailand
[2]English Department, the Faculty of Humanities,Naresuan University, Phitsanulopk
E-mail: [1]anchalee.ve@rmuti.ac.th, [2]Chwara2180@hotmail.com

Abstract - Our study aimed at developing the reading English newspaper material by using corpus approach and finding out the effectiveness of the corpus-based material. The samplings were forty-three-third-year students majoring English and the corpus-based material were used for a period of 10 weeks. The data were collected through the pre-test and post-test which divided into two parts: multiple choices and translation test. The quantitative data were analyzed through the paired sample t-test. The results revealed that using corpus-based materials were effective and could help students to improve their vocabulary knowledge, perceive the words, comprehend the meaning of the words, and also apply the words properly in the news contexts.

Keywords - Material development, corpus-based approach, EFL students

I. INTRODUCTION

In the world today, reading is one of the most accepted and substantial ways of learning other languages. One primary kind of reading materials is newspaper. Newspaper is regarded as one of the most accepted sources of knowledge, forwarding news, entertainment and ideas for the readers. Besides, newspaper conveys news and information to a considerable number of readers in very short time because of its functions which speaks to their readers in their readers' language (Photi-on, 1997 and Macalister, 2001). Thus, people could apply them as a basic source for forwarding various sorts of skills and knowledge. Further, newspaper is much used in reading courses for promoting reading skills and expanding lexical knowledge (Hwang & Nation, 1989). Hwang and Nation (1989) also stated the three reasons for using newspaper in classroom that it is easy and cheap to access both in hard copy and online version. Moreover, it is an authentic material that is generally read by the native speakers and it also serves large alternatives for interesting topics that learners or teachers could select reading texts.

Lastly, it is not only a good way of brushing up old vocabulary but also a great way of learning new vocabulary from context. Although reading is emphasized in EFL courses for Thai learners, they still have a trouble with vocabulary that is essential to help them understand their text.

As suggested by Read (2000) that vocabulary is a word or a group of word combination in order to make a language that provides a grammatical and lexical function in both spoken and written texts. Thus, vocabulary is very meaningful to help EFL learners to comprehend the language (Krashen& Terrell, 1983; Krashen, 1989; Leki& Carson, 1994). In addition, a large vocabulary can assist learners in learning a language since the more vocabulary know, the better they understand what they read (Meara, 1996 and Nation, 2001). Moreover, high frequency words are a core of learner's thought in learning vocabulary as recommended by Nation and Newton (1997). Similarly, the competence of understanding words by achieving these words from encountering with them in particular contexts is still fruitful as it could be valuable for learners to accomplish their understanding as well as they could discover new words when they learn with new lessons. Thus, deciding which words should be taught and learners has been controversy in one of the recent trends is to accept. The corpus-based approach is applied as the fundamental part in vocabulary development. As suggested by many researchers that materials developed by using corpora are accepted to be more authentic and enabled researchers to determine with a high degree of particularity that appearing words and phrases a language learner would take advantage most from studying (Biber, Conrad, &Reppen, 1994; McCarthy & Carter, 1997; Simpson &Mendis, 2003; Schmitt et. al, 2004).

In this way, selecting the vocabulariesthat are most frequent and useful to guarantee the contexts in which words are used as authentic ones, based on the contexts that appear in corpora and most related to their EFL lives, is undeniably necessary. Consequently, the main reason for undertaking the currentstudy is to simplify the reading English newspaper comprehension by building up familiarity with English newspaper key words and develop EFL corpus-based material for reading English newspaper.

1.1 Research Purposes

The purposes of the current study are as follows:

1. To develop the reading English newspaper corpus-based materials for EFL students.
2. To determine the effectiveness of the reading English newspaper corpus-based materials.

II. PROCEDURE

2.1 The Samplings

Forty-three students majoring English who enrolled in the Reading English Newspaper coursein their third year were selected by the purposive sampling method.

2.2 The Source of Data for the Corpus

The source of data for the English Newspaper Corpus (ENC) analysis is the online news articles from the general, business, and sports section of the Nation Newspaper during the period of one month from April 1st to April 30 th, 2018. The total numbers of news articles were 422 files and the total number of tokens or running word were 198,742 words.As mentioned earlier, the current study aimed to investigate high frequency words and key words in order to use as data of developing the course material thus English news articles in an online version during the period of one month were chosen as a significant input for this study. The information of the current corpus is shown in Table 1.

Table 1: General Statistic of Each File

	General	Sports	Business	Total
No. of articles	150	151	121	422
No. of tokens	86,992	69,004	42,746	198,742
Word types	8,529	6,837	5,479	20,845

According to Table 1, the token running words of each file were 86,992 for general, 69,004 for sports, and 42,746 for business. The general file had the highest number of token words while the business file had the smallest number of tokens. As for word types, the general file also had the highest number of the different word types (8,529) whereas the business file had the smallest number of different word types.

2.3 Procedures of Corpus Construction

The research procedures of this study consisted of eight steps in order to profile the English Newspaper Corpus (ENC). The method of data analysis can be illustrated in Figure 1 which are relevant to how data were investigated and how to create a selection of texts in order to achieve the top 50 most frequently occurring word list.

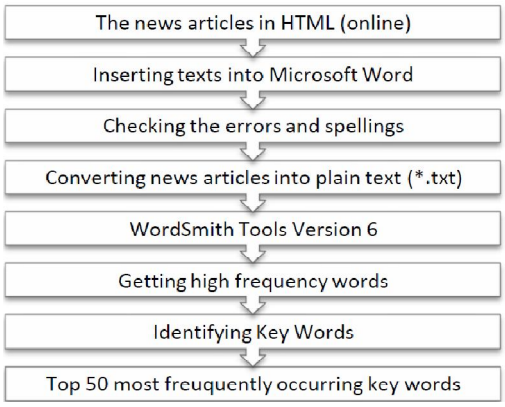

Fig. 1. Procedures of Corpus Construction

Step 1: The news articles in HTML (online)

To achieve the data, the news articles in the general, business, and sports section of the online version were applied as significant data for this study. The news articles were collected for a period of one month from April 1st to April 30 th, 2018.

Step 2: Inserting text into Microsoft Word

Thereafter gaining the data in HTML format, the online texts from the general, business, and sports news articles were inserted into "Microsoft Word" as a word document (*doc) and the total of news articles were 422 files.

Step 3: Checking the errors and spelling

The checking of the errors and spellings of the words analysis were checked before adopting texts into the analyzing program by this step.

Step 4: Converting news articles into plain text (*.txt)

After checking of the errors and spellings in word documents (*doc) then we saved all news articles as plain texts in order to make the analyzing program better.

Step 5: Word Smith Tools Version 6

All news articles were inserted into the "WordSmith Tools Version 6" (Scott, 2012) for analyzing the most frequent words in the English Newspaper Corpus (ENC).

Step 6: Getting High Frequency Words

After inserting all news articles into the "WordSmith Tools Version 6", the program was given the high frequency words in the text files.

Step 7: Classifying the Key Words in Newspaper

After getting the most frequently words of the English Newspaper Corpus (ENC), the frequency words list was eyed by applying the Oxford Advanced Learner's Dictionary manually to classify the key words. Later, the top 50 most frequently occurring key words were achieved. Moreover, these words occurring in the ENC at least 50 times.

2.4 Procedures of Making Lessons

The reading English newspaperlessons consisted of six activities. These activities composed of 50 words to be studied and each lesson comprised 10 words, which were presented in six vocabulary activities. The six vocabulary activities included:

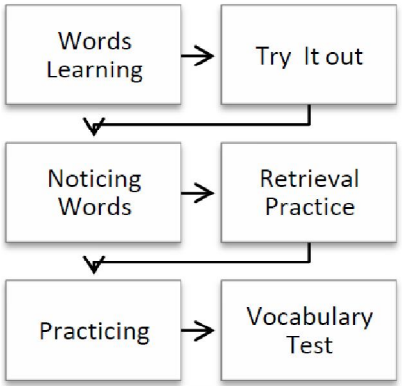

Fig. 2. Procedures of Making Lessons

1. Words Learning
This activity was designed to help the students to realize the 10 target words and introduced the target words to samplings that to be learned in each activity. The 10 words is a proper number of words per lesson as mentioned by Wallace (1984). Thus,the target words in each lessons were ordered alphabetically in order to improve the awareness of the meaning of words.

2. Try it Out
It is an activity that aimed to check which of the target words they know and do not know. After learning words, the students need to check out their vocabulary knowledge first for getting the picture of their learningimprovement (Read, 2000).

3. Noticing Words
This activity was created to remark the target words that students should pay attention to the unknown words.It was planned to emphasize learners to the target words (Nation, 2001).

4. Retrieval Practice
It is an activity which students can guess the meaning of target words from contexts. In this practice, students might perceive words and comprehend their meanings toward texts throughout the exercise,or the teacher's explanation or dictionary using. This way can help students to memory the words better when they recall those words during the task.

5. Practicing
This activity focused on practicing students with vocabulary by matching, gap filling, and word association. It is required students to remember the target words and at least of their meanings or requiring just partial knowledge of students.

6. Vocabulary Test
It is designed to rate the development of students in their vocabulary learning after accomplishing each lesson. As Nation (2001) stated that students should know their learning development in order to get an evaluation after learning.

2.5 Research Instruments
The research instruments used in the current study were the pre-test and post-test.
The pre-test was designed to set up the guideline to the samplings and the post-test was purposed to examine whether the students have accomplished their improvement or not after they have learnt the target words through the lessons. The pre-test and post-test were divided into two parts that are multiple choices and translation test.
The first part contained 30 multiple-choice questions with four alternatives. When they make multiple answers, they will receive one point for every correct answer and receive zero point for every wrong answer.
The second part is translation part comprised of 30 items to evaluate the students' understanding of perceiving the target words and comprehend their meanings. The translation test was designed by incidentallychoosing 30 out of 50 target words that students have learnt in the lessons. In this part, students have to translate the meanings of target words from English to Thai with the answer keys which was based on the dictionaries.

III. RESULTS

The purposes of the current studywereto develop the reading English newspaper corpus-based material and to determine the effectiveness of the corpus-based materials whether it improved the EFL students' vocabulary knowledge. The pre-testand post-test were used as the research instruments to obtain the findings of research purposes. The quantitative data were analyzed through the paired sample t-test.

Table 2: Pre-Test and Post-Test Scores

No.	Pre-Test	Post-Test	No.	Pre-Test	Post-Test
1	25	40	23	36	56
2	22	45	24	31	42
3	30	42	25	33	53
4	18	35	26	22	38
5	20	40	27	28	45
6	32	50	28	25	42
7	31	53	29	27	50
8	27	43	30	35	55
9	22	48	31	37	56
10	18	36	32	33	56
11	26	45	33	41	55
12	30	45	34	25	39
13	28	39	35	33	48
14	29	45	36	26	49
15	17	32	37	17	33
16	24	42	38	29	49
17	35	53	39	45	56
18	28	49	40	27	50
19	38	55	41	29	52
20	40	56	42	36	54
21	26	48	43	30	53
22	23	40			
			μ	28.69	46.79
			σ	6.56	6.95

Table 2 showed the mean score (μ)of the pre-test was 28.69 and the standard deviation (σ) was 6.56. It can be seen that in the post-test the mean score of all studentswas higher afterlearning throughthe lessons. The mean score (μ)of the post-test was 46.79 and the standard deviation (σ) was 6.95.Furthermore, the results of the pre-test revealed that there were 20 students who had the score below mean level which accounted for 46.51 percent. There were also 19 students who received the above level of mean scorevalue that accounted for 41.3 percent.
Besides, the results of post-test exhibited that there were 20 students who had the score below mean level.However, it can be seen that students have made good progress in their vocabulary learning through the lessonswhen looking at the pre-test score.

Table 3: The Maximum and Minimum Score of the Pre-Test and Post-Test

Test	Minimum	Maximum
Pre-test	17	45
Post-test	32	56

Table 3 displayed the minimum scores of the pre-test was 17 while the minimum score of the post-test was 32. For the maximum score of the pre-test 45 whereas the maximum score of the post-test was 56. This can be supposed that after learning through the lessons, students have made progress in vocabulary learning.

CONCLUSION

The current study was carried out with the following purposes: 1) to develop the reading English newspaper corpus-based materials for EFL students and 2) to determine the effectiveness of the reading English newspaper corpus-based materials. The two purposes were posed and the answers to these purposes are given as follows.

Research Purpose 1: To develop the reading English newspaper corpus-based materials for EFL students

To accomplish this purpose, we gathered 422 news articles from online version of the Nation Newspaper which is one of popular newspaper in Thailand during April 1st to April 30 th, 2018. We collected the online news articles from the general, business, and sports section then saved them into "Microsoft Word". Before the "WordSmith Tools Version 6" was used to classify the most frequent words and key words of English Newspaper Corpus (ENC), all files were converted into the plain text (*.txt) in order to make the analyzing program effectively. The concordancing software "Word Smith Tools Version 6" was applied. This concordancing software was used to see how words behave in texts. It was used in order to make the word frequency lists of news articles files by using the Wordlists Tool that offered both alphabetical and frequency order of the words in the texts files. The total corpus was 198,742 tokens or running words and the number of word types was 20,845. From the analysis of the study, the top 50most frequently occurring words of the English Newspaper Corpus (ENC) was obtained. Thereafter, the reading English newspaper lessons that consisted of six activities, were constructed. These activities composed of 50 words to be mastered and each lesson comprised 10 words. The six activities were Word Learning, Try It Out, Noticing Words, Retrieval Practice, Practicing, and Vocabulary Test. These lessons of the reading English newspaper corpus-based materials were used by the 43 students for a period of 10 weeks. Research Purpose 2:To determine the effectiveness of the reading English newspaper corpus-based materials. In order to determine the effectiveness of the English reading material, the pre-test and post-test were carried out. The pre-test and post-test were divided into two parts that are multiple choices and translation test. The multiple choices part contained 30 multiple-choice questions with four alternatives. When they make multiple answers, they will receive one point for every correct answer and receive zero point for every wrong answer. The second part is translation part comprised of 30 items which was designed by incidentally choosing 30 out of 50 target words that students have learnt in the lessons. The results revealed that the post-test mean score (μ)was increased which confirmed that the constructed lessons did support the 43 students to develop their vocabulary knowledge.The results also displayed that students can perceive the words, comprehend the meaning, andapply the words properly with the contexts. This can be concluded that the vocabulary exercises that designed for the lessons could help the students to enhance their vocabulary knowledge and help them to recognize the meaning of the target words. Additionally, it can be assumed that students could learn new words approximately 5 to 10 words each time and the number of times that words appear around 7 to 10 times could help them to recognize the words well.

REFERENCES

[1] Biber, D., S. Conrad and R. Reppen. (1994) "Corpus-based approaches to issues in Applied linguistics. Applied Linguistics, 15(2), pp. 169-189.

[2] Hwang, K. and Nation, P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. Reading in a Foreign Language, Vol. 6 (1), pp. 323-335.

[3] Krashen, S. and Terrell, D. (1983). The Natural Approach: Language Acquisition in the Classroom. New York: Prentice-Hall.

[4] Krashen, S. (1989). "We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis", The Modern Language Journal, 73 (4), pp. 440–464.

[5] Leki, I. and Carson, J. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. TESOL Quarterly 28 (1), pp. 81-101.

[6] Maclister, J. (2001). Introducing a New Zealand newspaper corpus. New Zeakand English Journal. Vol. 15. pp. 35-41.

[7] McCarthy, M. and R. Carter. (1997). The importance of corpus-based research for language teachers. Cambridge: Cambridge University Press.

[8] Meara, P. (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K. and Williams, J. (eds.) Performance and competence in second language acquisition. Cambridge: Cambridge University Press. pp. 35-53.

[9] Nation, P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press.

[10] Photi-on, C. (1997). Fundamentals of Public Relations. Bangkok: RajaphatChadrakasem Institute Press.

[11] Read, J. (2000). Assessing vocabulary. Cambridge: Cambridge University Press.

[12] Scott, M. (2012). WordSmith Tools version 6. Stroud: Lexical Analysis Software.

[13] Schmitt, N. et al. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? Formulaic sequence: acquisition, processing, and use. pp. 23-25.

[14] Simpson, R., and Mendis, D. (2003). A corpus-based study of idioms in academic speech. TESOL Quarterly, 37. pp. 419-441.

[15] Wallace, M.J. (1982). Teaching Vocabulary. London: Heineman Education Books.