

English Core Word List for Grade 4-6 Thai Students

- Nakarin Suwadit
- Achara Wongsorthorn

Abstract: The purposes of this research were: 1) to study and develop a list of high-frequency 2,000 core words from course books, O-Net, E-graded readers, Aesop's fables and cartoons for grades 4-6 students; and 2) to develop the Core Word List as a guideline for Thai students learning the English language. The instrument used was the Range program developed by Paul Nation and Alex Heatley (2002), the statistical procedures employed were frequency and percentage. A total of 352,289 running words from the initial corpus were lemmatized to extract all base forms. The word frequency and the Thai students' environment were used as criteria for selecting 2,000 words in the construction of the English core word list. The first 2,000 words of this study were categorized according to the themes presented in Basic Education Curriculum for upper elementary students (2008); the words filled all 10 of these themes plus the researcher's collective theme including loan words. Three English language teaching experts confirmed the appropriateness of the words for the target age levels, and the 2,000 word category. Finally, the high frequency 2,000-word family list was compared with the General Service List (GSL) to analyze whether there was any overlap. The results indicated that the most frequently occurring 1,000 words attained a cumulative coverage of 83.60 %, and the 2,000 words attained a cumulative coverage of 91.20 % in the corpus. The 2,000 words could fill all of the themes, and the percentage of overlap is approximately 74.69 % with the GSL. This study could serve as a reference source of the core vocabulary for the upper elementary grades in addition to the core content areas of learning foreign languages (The Basic Education Core Curriculum).

Keywords: english core word , thai students, grades 4-6

Introduction

Statement and Significance of the Problem

A number of researchers have conducted studies focusing on the problem of inadequate English vocabulary knowledge of Thai learners. The results show that most Thai students fail to meet the basic requirements in the English syllabus during those six years. A lot of Thai students still have problems in English in terms of their learning outcome. One main problem concerning English learning in Thailand is due to inadequate vocabulary knowledge. Kamol and Sripetpun (2011) studied the ability of Thai students using English and found that inadequate vocabulary knowledge at the elementary level as shown in the result of the National Test is significant evidence of the English proficiency problem. Chinnawongs (2001) showed that students still have difficulty in language production because they do not have adequate knowledge of vocabulary to convey their meaning accurately. Therefore, according to these researchers, vocabulary is the major factor that affects students' English development. It might be the case that many elementary school students in fact do not have

the appropriate vocabulary size. Thus, schools need improvement in vocabulary instruction since the vocabulary gap gets worse as students get older and most failures in school are linked to deficient vocabularies.

Second language vocabulary research is principally interested in a core vocabulary word list comprised of high frequency words. The most well-known general service vocabulary is West's (1953) General Service List of English Words (GSL). Nation and Waring (1977) pointed out that the beginners of English learning should focus on the first 2,000 most frequently-occurring word families of English in the GSL. Even though, in Thailand, English for communication has been promoted in school, the outcomes of the English learning and teaching at the primary school level through secondary school level have continually gone down. As stipulated in the English syllabus (Ministry of Education, 2008), 6th grade Thai students should meet the basic requirements, where their vocabulary size should reach 1,050-1,200 words. This study reflects two existing situations. First, no detailed word list for teaching vocabulary to 4th- 6th grade is available in the Thai English curriculum. Second, the extensive study on a core vocabulary word list for primary school children has never been done before.

The goal of the study is to use corpus methodology to create a word list that is pedagogically useful to elementary level students in Thailand as a step towards addressing this gap in vocabulary knowledge. That is to say, a 2,000 word family may serve as a reasonable vocabulary learning goal for elementary school students in Thailand, with a view to investigate the vocabulary load of the school texts and authentic sources. If students are expected to have 2,000 new word families in three years, teaching only three or four words a day will significantly increase the number of new words they learn per year. The contributions of this study not only would help students improve their English skills but also provide a glossary of core English vocabulary words for Thai learners. This study focused on only basic high-frequency words for young learner of Grades 4-6. This area was selected for the study for two reasons: first, during the upper elementary grades, learners should take more responsibility for learning the needed meanings. In addition to root word meanings, directly teaching about the use of prefixes and suffixes to extend vocabulary is needed in the upper elementary grades. Second, by the end of grade six, it will be much more effective and beneficial for the learners' future language learning for the transition from the elementary level to the secondary level.

Research Objectives

- 1) To study and make a list of 2,000 core vocabulary most frequently used in course books, the Ordinary National Education Tests (O-NET), graded readers, story books, an Aesop's Fables book, and cartoon strips used for children at the primary level.
- 2) To develop a 2,000 High Frequency Word List as a guideline for grades 4-6 Thai students learning the English language.

Research Questions

- 1) What are the first 2,000 high-frequency words found in course books, the Ordinary National Education Test (O-NET), e-graded readers, story books, an Aesop's Fables book, and cartoon strips?

2) What percentage do these 2,000 high frequency words overlap with the General Service List of West (1953)

Literature Review

In learning English, vocabulary is one of the important language components and is of critical importance to foreign language learners. Researchers have been able to ascertain that the size of one's vocabulary seems to be a determining factor for second language learning (Meara 1996). They confirm that learning vocabulary is an essential part of mastering a second language (Schmitt, 2008). This is based on the belief that learners need a certain amount of vocabulary in order to be able to operate independently in the target language (Alderson & Banerjee, 2001). The foundation of vocabulary knowledge, thus, must be in place early if non-native learners are going to perform successfully in school (National Reading Panel/NRP, 2000).

There are a huge number of vocabulary words in English language, so vocabulary should be taught through an appropriate approach to achieve the goal. For EFL teaching, one of the fundamental issues that course designers and language teachers have to consider is what vocabulary to teach and how large the vocabulary should be. There have been many attempts to define the vocabulary size for non-native learners. Many studies show that learning root words helps students learn vocabulary is a shorter amount of time. With reference to word counts, Nation (2001) holds that knowing a word involves knowing the members of its word family and the number of members of the word family will increase as proficiency develops. Besides, it has been widely accepted that teaching vocabulary needs to be done on the basis of word frequency and text coverage (Meara 1993). Nation (1990, 2001) reported that frequency-based studies have shown that a small group of very frequent words cover a very large proportion of the running words in any spoken or written text and occur in all kinds of uses of language. High-frequency words refer to those basic general service English words which constitute the majority of all the running words in all types of writing. The most well-known general service vocabulary is West's (1953) General Service List of English Words (GSL). The GSL containing the most frequently-occurring 2,000 word families of English (3,372 word types) accounts for approximately 80% of the running words in most texts (Coxhead and Nation, 2001; Cobb and Horst 2001; Nation and Waring, 1997). Meara (1995) claimed that these words are so essential for any real language users and that it might be a good idea to teach them right at the beginning of the language course. Therefore, students should learn vocabulary as much as they can by starting from basic vocabulary in elementary or secondary school, and then to specific vocabulary at university.

Research Methods

English materials and instruments

In this study, there were three instruments: English materials, Range Software, and a validating form.

English materials

First, English materials were collected as the initial word lists. To examine the vocabulary of English resources students may read or encounter during elementary school

(Grades 4-6), 6 main sources of materials were printed and grouped into 6 word lists to form a basis of comparison, which are:

- 1) Course books for Grades 4-6
- 2) Ordinary National Education Tests (O-NET)
- 3) E-graded readers for young readers 9-11 years
- 4) Story books
- 5) An Aesop's Fables book
- 6) Cartoon strips

GSL RANGE software

The researcher chose GSL RANGE software (Heatley and Nation, 2002) as it is a smaller corpus than BNC and the results would be more accurate. It included two programs: RANGE and FREQUENCY programs which incorporates the General Service List of English Words (GSL), Academic Word List (AWL) and British National Corpus High Frequency Word List (BNC HFWL) based on English words' occurring frequency, range and dispersion figures. The Range and Frequency programs for Windows based PCs are vocabulary profilers developed by Heatley and Nation (2002) and are freely downloadable.

Validating form

After the 2,000-word list was created, the validation forms were sent to three English teaching experts to obtain their opinions concerning the words that they agree to keep or to change. The 23-page validation form included 20 pages of a 2000-word list and 3 pages of a 234-word list. The 2,000 words were for validation and the extra 234 words for substitution if necessary.

Refinement and processing of the instrument

After compiling all materials, the process of generating and preparing word lists for analysis was made three steps described hereafter.

Inputting and organizing vocabulary

The first step of this phase consisted of inputting all English vocabulary that occurred in each material, and the contents of six materials were typed manually into the computer. The computer files of each material were compiled into a single computer file. All vocabulary files were combined into a corpus and then the corpus was transferred into one text file. There were 352,289 running words of materials in the corpus.

Developing the first word list by using GSL Range Software

The first word list with ranks, frequency number results and a list of word types including the ranks and frequency digits was derived using Range software.

Lemmatizing the extracted word list to a core list of the 2,000 high-frequency words

In order to have the same word family building basis for comparison as the GSL 1st and 2nd 1000 word lists in the RANGE program, the high frequency word list of 2,000 headwords was lemmatized into word families (Nation, 2001). The criteria used in the RANGE program to make word families were based on Bauer and Nation (1993), and include all the affixes, inflected and derived. A word family consists of a base word and all the derived and inflected forms that can be understood by a learner without having to learn each form separately. The derivational affixes at Level 3 are -able, -er, -ish, -less, -ly, -ness, -th, -y, non, and un-, were used because all the inflected forms and a small group of

high-frequency, regular, and productive derived forms were included. This level seemed appropriate for the proficiency of elementary school learners (Nation, 2001).

Analysis and statistical treatment

Classifying 2,000 high frequency word families

The 2,000 high frequency word families were classified into 10 themes defined by the Basic Education Curriculum for upper elementary level (2008) plus two additional themes (collective theme of story, graded readers, Aesop's fables, cartoon strips, and loan words).. Because the corpus was collected from two types of input: school resources (student's books and the O-Net) and children's resources (graded reader, story, Aesop's Fable, and cartoon strip).

Validation of the obtained Word List

An additional way to identify the validity of this word list for the present study was an experts' analysis. Three experts were asked based on their teaching experience in an elementary school to confirm the appropriateness of the words selected by the researcher for the Grades 4-6 students. If inappropriate, some words might be deleted or others might be added.

Comaring the 2,000-word List with the controlled word List

In order to tackle research question 2, the obtained 2,000-word list had to be compared with a well-known word list, the coverage percentage of the overlapping words of the word lists, and the number of absent words from the corpus. The General Service List (GSL) of West (1953) was used as the well-known word list with a base list of the GSL, and one category of the word list was downloaded from the Range Software.

By running the Range software, the overlap percentage of the 2,000-word list and the GSL were found. If the percentage of overlapping was high the word list in the study was reliable and valid. The comparisons of another study reveal missed words that were not found in the GSL.

Designing a 2,000-word bank for Grades 4-6 students

After the target word list was created, they were used in designing an English core word list with examples of pictures and sentences for Grades 4-6 students.

Results of the Study

Words lemmatization

To extract all base forms, the mastery word list was lemmatized, and proper nouns and numerals were excluded from the list. For this research a word family was defined as the base form. The idea behind a word family is that inflected and regularly derived forms of a known base word can also be considered as known words if the learners are familiar with the affixes (Bauer & Nation, 1993), i.e., inflectional forms such as cat-cats and go-goes-went-gone-going were listed under the base word forms of cat and go. Proper nouns such as 'Ben', 'Mickey', 'Dan', 'Jack', and 'Donald', and numerals such as '12', '2011', were manually excluded from each list (Coxhead, 2000; JACET, 2003; West, 1953), since "they are of high frequency in particular texts but not in others, ... and they could not be sensibly pre-taught because their use in the text reveals their meaning" (Nation 2001: 19-20).

However, this study did not exclude days of the week, months of the year, numerals in words, and prepositional phrases. This is mainly due to the fact that all the items mentioned are included as parts of the English learning goals for primary education (Nation, 2001).

Results of the first 2,000 words selected

Due to the purposes of this study and the validating of the results if a word appeared fewer than 6 times on the initial word list, it was deleted according to Gardner (2008). The total number of words that have 6+ repetitions is considered to be the threshold at which vocabulary is acquired. Consequently, 2,234 words formed the initial word list and were kept for the candidate word list.

It was found that the top 1,000 word families on the initial word list which occurred more than or equal to 33 times are the words that occurred most frequently in the corpus and the low coverage of the types not in the first 1,000 words. Whereas the first 1,000 types cover approximately 80 % of the running words on the initial word list, the second 1,000 words outside the first 1,000 cover only 7%. So, the 1,000 word families that occurred in the corpus were considered to be the most essential 1,000 words learners should know according to Adolphs and Schmitt (2003; 2004). These top 1,000 words were kept and formed the English Core Word List (ECWL).

Next, the remaining word families were analyzed and only 1,000 from 1,234 word families were selected to add into the ECWL. The analysis of words beyond the 1,000 high frequency words would not produce a normal frequency list as would be obtained from an analysis of the corpus. Because of this, the criteria of frequency of occurrences often used in many studies were not applicable (Nation, 2001) and the remaining 1,234 word families provided coverage of only 7% of the initial word list. Furthermore, it was found that some words, for example, 'illustrate', 'immigrate' and 'research', are not suitable for a beginner's vocabulary came in the second 1,000 words of the initial list. It is impossible to use all the items on the list, for a number of reasons. They are difficult to learn until students have more English or learn at intermediate or advanced level where learners usually study English for academic purposes.

Results of theme-related words on the ECWL

In addition to providing a listing of the ECWL, the 2,000 words were classified into 10 themes defined by the Basic Education Curriculum for upper elementary learners (2008) plus two additional themes (collective theme of stories, graded readers, Aesop's fables, cartoon strips, and loan words).

Results from comparison between the first 2,000-word list and the General Service List

Compared with the GSL's two base lists, the most words are found in the GSL's base list one (BASEWRD1) and the percentage is 42.49 %, follow by 32.20 % in the GSL's base list two (BASEWRD2), and 21.67 % not in the GSL's two base lists. The ECWL resulted from the research and through comparisons of the GSL. This indicated that the ECWL has a satisfyingly high percentage of overlap with the controlled word list. The reliability of the ECWL has been proved to be consistent with similar word list.

2,000-English Core Word List

The 2,000-English Core Word List (ECWL) resulted after a step-by-step research procedure to develop the 2,000 high-frequency words as a guideline for grades 4-6 Thai students learning the English language. It is a high frequency list of all the different words that occur in six different English materials. They were recorded on a CD and are representative of words for Grades 4-6 students. In the ECWL, some words are presented with pictures.

Conclusions and Recommendations

Conclusions

In the present study, the role of high frequency words in children's English sources in the development of learning materials has been explored. The criteria of frequency of occurrence, range, validation, a controlled word list comparison, and Thai students' environment were used in this study for selecting words. Most of the analyses done in the study involved the use of a computer program called RANGE which was developed by Nation and Heatley (2002). Six English material were complied as the corpus creation, course books and the O-NET set under the Basic Education in curriculum B.E. 2551 (Ministry of Education, 2008), the graded reader e-books, story books, an Aesop Fable book and cartoon strips set under children sources. 352,289 running words were complied onto a corpus.

The findings of the present study are helpful for students, English teachers, and English learning and teaching as a second or a foreign language at the primary level. Since the 2,000 word families in ECWL are based on frequency and range, the occurrences of a word across subsections of a corpus were used in order not to overload learners with too many target words. A word family refers to a grouping containing a headword, its inflections, and its closest derivations (Nation, 2001). Awareness of word family relationships can greatly decrease the learning burden of derived words containing known base forms (Nation, 2001).

The Basic Education Curriculum for the elementary level under the Ministry of Education offers English courses and provides outlines for contents and words taught in the nation. The learning objectives of foreign languages are aimed at enabling learners to acquire a favorable attitude towards the English language, the ability to use English for communicating in various situations, seeking knowledge, engaging in a livelihood and pursuing further education at higher levels. After graduation, learners have to use English (listening, speaking, reading and writing) to communicate about themselves, their families, schools, their environment, foods, beverages, free time and recreation, health and welfare, selling and buying, and climate (Ministry of Education, 2008). The present study establishes those words found in the 6 materials (course books, Ordinary National Education Tests, e-graded readers, story books, an Aesop's Fables book, and cartoon strips) into the themes. An investigation of the ECWL's word frequencies show that semantic networks are well developed. This is determined by the word association integration. The content words on the ECWL were completely classified into the 10 themes plus two additional themes (collective theme of stories, graded readers, Aesop's fables, cartoon strips, and loan words). Thus, the ECWL provides a clear vocabulary goal for teachers and learners to be able to cope

effectively with specific goals in language use as it shows the word as well as the number of words in the text that need to be learned.

It is crucial to understand the roots of English problems of Thai students as well as to investigate the learning activities to enhance the human capital in order to achieve in this competitive era. Thai students are required to improve their English skills in order to compete with other ASEAN countries and achieve on an international stage. The findings from this study show the importance of the ECWL for young Thai learners at the elementary level. It is, therefore, reasonable to conclude that theoretically the 2,000 words should be the priority for young second language or foreign language learners. Attaining these 2000-words is an essential goal for second language learners. The researcher hopes that the course book will enable the students to attain at least the 1,000 word level by the end of their elementary education.

Recommendations

1) Teachers could use the core words as a word bank in the classroom to help Grades 4-6 students build the foundation vocabulary in an effective way. The English core word list could also serve as an example for both teachers and learners. Teachers could provide the target words in a Word Bank with learners supplying the definitions. This word list is a generative one; it can be used again and again and even updated if learners find over time that there are key words which are not included in the developed word list.

2) Based on the findings of the present study, textbook writers and EFL teachers could improve vocabulary teaching by paying closer attention to systematic vocabulary selection.

3) Based on the curriculum, teachers could apply the list to measuring the students' vocabulary performance and plan how to monitor the learning and teaching progress.

4) Teachers could use semantic networks as part of their learner's vocabulary learning, and use it as a reference to make informed decisions about the proper words in graded classes.

5) The results could help a teacher or a school to decide to retain or reselect the textbooks or redesign the teaching materials because the study of corpora of textbooks can show what vocabulary is presently being used in an English class.

6) The RANGE program is very useful for vocabulary learning. The teachers may use this program to highlight some words in the GSL. The design of the word list should be such that words appearing in the GSL and textbooks are included in the target lexis in lessons and given due emphasis to ensure that learners know them. Furthermore, they should be included in formative and summative assessments.

Recommendations for further research of these six types of materials will help language teachers and learners gain a much clearer understanding of how to choose texts and also how the conditions for vocabulary learning and reading comprehension can be improved. The present researcher suggests:

1) researchers should investigate the vocabulary issues by conducting more widely across all levels of elementary school to see if the word list is still consistent;

2) researchers should conduct a comparative study of the words in a corpus of texts in textbooks in schools (Thai program) and English textbooks (international program) which completely cover a large scale of daily life content; and

3) researchers should conduct comparative research on the GSL in a spoken corpus for children's sources. This would also serve as an alternative source for learners and teachers.

Finally, the researcher hopes that this present study will contribute valuable information on vocabulary teaching and learning for EFL/ESL learners, teachers, and text and/or curriculum designers and developers. The ECWL can provide guidelines for the development of a core vocabulary word list of the Basic Education Curriculum so that those learners whose English language learning is limited to the EFL learning contexts can seek ways to expand their vocabulary knowledge. Since a research-based word list in curriculum instruction for primary school grades 4-6 is not yet available, research in this area is urgently needed. Further research of this nature should study in detail about the words particularly for the exact elementary school level, i.e., which words for which grade. This in-depth classification should be more appropriate for course book development leading to more effective learning and teaching.

References

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(3), 425-438.
- Adolphs, S., & Schmitt, N. (2004). Vocabulary coverage according to spoken discourse context. *Vocabulary in a Second Language: Selection, Acquisition, and Testing*, 10, 39-49.
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part 1). *State-of-the-Art Review, Language Testing*, 18, 213-236.
- Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253-279.
- Chinnawongs, S. (2001). In search of an optimal writing feedback strategy. *PASAA*, 31, 27-43.
- Cobb, T., & Horst, M. (2001). *Reading academic English: Carrying learners across the lexical threshold*. London: Cambridge University Press.
- Coxhead, A. (2000). *The Academic word list*. Retrieved from <http://www.vuw.ac.nz/lals/research/awl/index.html>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead, A., & Nation, I. S. P. (2001). The specialized vocabulary of English for academic purposes. *Research Perspectives on English for Academic Purposes*, London: Cambridge University Press.
- Gardner, D. (2008). Vocabulary recycling in children's authentic reading materials: A corpus-based investigation of narrow reading. *Reading in a Foreign Language*, 20, 1-34.
- Heathley, A., & Nation, I. S. P. (2002). *Range and frequency programs*. Retrieved from <http://www.vuw.ac.nz/lals/staff/paul-nation/RANGE32.ZIP>
- JACET. (2003). *JACET List of 8000 Basic words*. Tokyo: JACET.
- Kamol, T., & Sripetpun, W. (2011). Vocabulary learning strategies employed by undergraduate students and its relationship to their vocabulary knowledge. *Proceedings of the 3rd International Conference on Humanities and Social Sciences*, (pp. 187-205). Songkla, Thailand: Prince of Songkla University.
- Meara, P. (1993). Tintin and the world service: A look at lexical environments. *IATEFL Annual Conference Report*, 32-37.

- Meara, P. (1995). The importance of early emphasis on L2 vocabulary. *The Language Teacher*, 19(2), 8-11.
- Meara, P. (1996). *The dimensions of lexical competence*. Cambridge: Cambridge University Press.
- Ministry of Education. (2008). *The basic education core curriculum B.E. 2551*. Bangkok: Express Transportation Organization of Thailand.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Heinle and Heinle.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt, & McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*.
- Retrieved from <http://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf>
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329-363.
- West, M. (1953). *A general service list of English words*. London: Green & Co.