Lexical Profiles of Thailand University Admission Tests

Wirun Cherngchawano
Natjiree Jaturapitakkul
King Mongkut's University of Technology Thonburi

Abstract

University Admission Tests in Thailand are important documents which reflect Thailand's education system. To study at a higher education level, all students generally need to take the University Admission Tests designed by the National Institute of Educational Testing Service (NIETS). For the English test, vocabulary and reading comprehension is one of the key elements. In order to prepare for and pass the test, students should learn and accumulate an adequate amount of vocabulary. The purpose of this research is to conduct a documentary study on the scattering and lexical profiles of Thailand University Admission Tests. Fifteen papers covering 55,161 running words were analyzed in a framework of two word lists: General Service List (GSL) and Academic Word List (AWL). The results showed that the coverage of GSL and AWL are 85.05% and 4.58%, respectively. A combination of the GSL and the

AWL covers 89.63% of the texts. For the coverage and reading comprehension, a 4,000-word level for reasonable comprehension covers 94.82% of the texts. It is suggested that both the GSL and the AWL could be good sources for students to learn so as to prepare for the test and to study at an advanced level in university.

Introduction

Vocabulary plays important roles in English language study, not only in learning but also in testing. Apparently, vocabulary is embedded in all parts of test. Even though there is no specific vocabulary section in most tests, students still need to understand and be able to use a large number of words to do well in all test sections. Somehow conventionally, some achievement tests in schools and some proficiency tests like the TU-GET (Thammasat University Graduate English Test) do provide a specific section that measures the vocabulary knowledge of the students. Furthermore, in the reading comprehension section of the test, students really need to know enough vocabulary in order to understand the text and pass the examination.

Thailand, where every student is required to take English tests as a part of university admission, the extent to which students are knowledgeable about vocabulary in the test seems to be genuinely important for and influential on test scores. Students then need to be very well-prepared and be ready to tackle vocabularies in the test. One way to assist students for successful and meaningful preparation is to provide them with the vocabulary that could appear on such a test. This study will

examine the profiles of vocabulary that appeared in the University Admission Tests in Thailand which could be useful to prepare students for the tests.

University Admission Systems in Thailand

Admission criteria for public universities have been changed over the past decade but generally have included secondary school grades, scores on ordinary tests and aptitude tests or admission examinations (in the past called the national entrance examination which had been in operation since 1962). The Central University Admissions System (CUAS) was launched in the 2006 academic year to replace the national entrance examination. It was recently implemented by 86 Thai universities in the 2014 academic year (Association of University Presidents of Thailand, 2014).

At present, students need to pass the CUAS which requires GPA (Grade Point Average), ONET (Ordinary National Education Testing), GAT (General Aptitude Test) which covers Thai reading passages and English communication skill, and PAT[1] (Professional and Academic Aptitude Test) which has a choice of seven subjects. At Mattayomsuksa 6 (equivalent to grade 12), all Thai students are required to take the ONET examination and English is one of the eight compulsory subjects that need to be tested. Moreover, if they want to study in a university, they have to attend the GAT in which an English test is one of the major components as well as in the PAT.

---

[1]The PAT or Professional and Academic Aptitude Test aims at assessing a test taker's fundamental knowledge in different professional and academic fields. Each student is required to take different sub-subject(s) in PAT depending on the field they would like to study. For example, if A wants to study Engineering, A has to take the PAT 3 (Engineering aptitude test). The percentage of PAT scores for admission varies from 0 to 40 which depends very much on the requirement of a particular faculty in a university).

Both students and teachers are well aware of the importance of English for gaining admission into a university as well as for future job opportunities. However, the proficiency in English of high school graduates is still much lower than the standard required on the national examination. According to the latest statistics from the National Institute of Educational Testing Service (NIETS), in 2014 the mean score of ONET for English subject was only 25.35% with a total of 414,688 students.

One of the main sections in both the ONET and the GAT is the vocabulary section. However, there have been a lot of complaints from students and teachers posting their opinions on educational websites in Thailand about what kinds of vocabulary are included in the admission tests. Some vocabularies are rarely seen in everyday life. This makes it hard for students to study and prepare for the test in this section. Furthermore, the vocabulary plays an important role in understanding the reading passages that appear in the other parts of the test. If students do not have sufficient knowledge of vocabulary, this can cause difficulties for students in tackling the tests.

Importance of Vocabulary in Language Learning and Testing

Vocabulary is central to English language learning because without sufficient vocabulary students cannot understand others or express their own ideas. According to Wilkins (1972), without grammar very little can be conveyed, but without vocabulary nothing can be conveyed. Particularly as students develop greater fluency and expression in English, it is significant for them to acquire more productive vocabulary knowledge and to develop their own personal vocabulary learning strategies.

Students often instinctively recognize the importance of vocabulary in their language learning. As Schmitt (2010) noted, "learners carry around dictionaries and not grammar books" (p.4). Teaching and learning vocabulary helps students understand and

communicate with others in English. Nation (2010) also stated that the most important jobs for English language learners are to make the most of opportunities to use the language, to deliberately learn vocabulary and to eventually take on responsibility for their own vocabulary learning.

In some tests, such as ONET, GAT or TOEIC, vocabulary is a main section of each test. It can be in the form of multiple choices, and filling-in the gaps whose meaning is based on the context, so as to check the knowledge of vocabulary of students. Moreover, vocabulary is one of the key success factors for passing reading comprehension tests. Students need to know enough vocabulary to understand the text in the test. Schmitt, Jiang and Grabe (2011) suggested that the 98% estimate of known words in a text is a reasonable coverage target for readers of academic texts. It is therefore important to assist students to boost up their corpus in order to make the most of text comprehension and be able to tackle the vocabularies in the tests.

To help facilitate students' learning and prepare them for the test, classifying vocabulary might help learners plan their learning and test preparation more effectively.

Classification of Vocabulary and Word List

In general, we can classify vocabulary into many categories depending on the criteria to be used such as by function word, content word or parts of speech. One criterion is the frequency of occurrence. Nation (2001) categorizes vocabulary into four groups according to their frequency of occurrence. The major reason for word classification is to give a basis for planning teaching and learning since different groups of vocabulary need different teaching and learning strategies. Here are the details of each group.

1. High-frequency words are basic English words which can be found in everyday conversation and every type of literature.

The General Service List of English words-GSL (West, 1953) is a standard list of high-frequency words containing 2,000 word families including function words and content words. Each of the 2,000 words is a headword representing a word family that is only loosely defined by West. Approximately 80% of running words in the text are high-frequency words.

2. The Academic Word List (AWL) was analyzed by Coxhead (2000).These words are commonly found in various kinds of academic texts but not in general English. They make up about 9% of the running words in an academic text. The list contains 570 word families that consist of head-words plus their inflected and derived forms. There are around 3,100 word-forms, altogether. The list was compiled following an analysis of over 3,500,000 words of text. The words selected for the AWL are words which occur frequently in a range of academic subjects, such as the arts (including history, psychology, sociology), commerce (including economics, marketing, management), law, and the sciences (including biology, computer science, mathematics).

3. Technical words are vocabulary used in a special area of study and are significantly different from field to field. As soon as we see them we know what topic is being dealt with. Normally, students obtain these words while they learn the specialized subject matter. Definitions of words in this group can be found in a technical dictionary of that specific discipline. Typically, technical words cover about 5% of the running words in an academic text.

4. Low frequency words are the words outside the above three groups. They include very technical words of other areas as well as words that are rarely found in everyday language. Proper nouns are included in this category. This group of words is likely to cover about 5% of the running words in an academic text.

Moreover, there still are many other word lists compiled and developed by many theorists, each for a specific purpose. All of them are normally intended for use as a basis for language teaching or for the preparation of teaching materials. Examples of these lists can be found in a study by Lessard-Claouston (2012). Notably, this study only focused on the use of the GSL and the AWL because of several reasons. As for the GSL, Nation (2001) claimed that the GSL covers 80% of various types of texts. In addition, Bauman (1995) stated that the GSL is used as the basis for many graded readers especially in a secondary level. Therefore, the GSL is important for students to learn and build a strong foundation of English vocabulary. For the AWL, it covers a wide range of academic texts across various disciplines which students will encounter when they study at the university level. For these reasons, it is interesting to investigate the lexical profiles of University Admission Tests which are supposed to cover a range of vocabulary from secondary education as well as those that might be encountered at the university level.

Frequency, Coverage and Reading Comprehension

Vocabulary knowledge is crucial not only in the vocabulary part of the tests, but it is also important in the reading part. Numerous studies indicate that vocabulary knowledge is an important factor for understanding the reading text. Students should know enough vocabulary to cover the main parts of the reading text.

In general, coverage means the percentage of running words on which we focus divided by the total number of running words in the text. For instance, 10% AWL coverage means that 10% of the AWL families appear in the text. Milton (2009) claimed that knowledge of 1,000 words in English should indicate that a learner would recognize and understand about three quarters of the words in a normal text. Knowledge of about 2000 words in

English should mean that 80% of words in a normal text would be understood. He also set up the rule of thumb that the most frequent 2,000 words in English are likely to be the most useful to a learner and that knowing these will enable the learner to recognize about 80% of any normal text.

For reading comprehension, Laufer (1989) originally came up with a 95% figure by exploring how much vocabulary was required for the participants to ensure 'reasonable' comprehension. Reasonable comprehension was assessed as the ability to achieve a score of 55% on a reading comprehension test, the minimum required for a pass in the Haifa university system. A later study by Hu and Nation (2000) reported that 98% coverage would be the threshold at which learners could understand enough of a text to be able to read it for pleasure. There do not need to be a contradiction between these two figures. Reading for pleasure may simply require different levels of knowledge. In addition, a follow-up study by Laufer (2010) suggested two thresholds: an optimal one, which is the knowledge of 8,000 word families yielding a coverage of 98% (including proper nouns) and a minimal one, which is 4,000-5000 word families, resulting in a coverage of 95% (including proper nouns).

Even though the later study argued that 98% coverage seems to be reasonable for reading comprehension, this study will use 95% coverage as a threshold for reading comprehension. As recommended by Milton (2009), with 95% coverage, most readers feel they can understand just about everything. This extensive coverage leaves only a negligible number of unknown words in a passage and most readers have the ability to skip over these and take the general meaning for the piece without needing to recognize or guess every single word. For understanding of a text, almost all the words, probably 95% or more, will need to be known.

Previous Studies

Many research studies have been conducted to analyze the profile of the GSL and the AWL in English texts. Different techniques and findings are shown in Table1.

Table 1: Summaries of the Previous Studies

Author (Year)	Studied Text	Findings

Poonpon (2002)	Intensive and extensive materials taken from English courses for the first and second year science students from Mahidol and Khonkaen universities in the 2001-2002 academic year	<u>Mahidol University</u>
		First Year Intensive Course
		    GSL covers 83.4% and AWL covers 5.6%
		First Year Extensive Course
		    GSL covers 88% and AWL covers 2.1%
		Second Year Intensive Course
		    GSL covers 83.2% and AWL covers 8%
		Second Year Extensive Course
		    GSL covers 78.1% and AWL covers 6.7%
		
		<u>Khonkaen University</u>
		First Year Intensive Course
		    GSL covers 89.9% and AWL covers 2%
		First Year Extensive Course
		    GSL covers 85.3% and AWL covers 3.5%
		Second Year Intensive Course
		    GSL covers 83.4% and AWL covers 7.1%
		Second Year Extensive Course
		    GSL covers 82.7% and AWL covers 6%

Para (2004)	A total of 136 research articles were used in this study: 68 from five Structural Engineering journals and 58 from Transportation Engineering journals published in 2002.	GSL covers 72.54% and AWL covers 12.46%
Boonyapapong (2007)	A corpus of 859,890 running words taken from The Nation – a local online newspaper.	AWL covers 2.09 % in the text
Chen and Ge (2007)	A corpus of 50 medical research articles written in English with 190,425 running words.	AWL covers 10.07%

Konstatakis (2007)	The corpora of Business Text (1 m words), General Fiction (2.5 m words) and Lord of the Rings (624,000 words).	AWL covers 11.15% in Business Text, 1.31% in General Fiction and 0.52% in Lord of the Rings
Thepwiwatjit (2008)	40 articles published in the Journal of Food Science from 2002 to 2007. The total running words is 121,308 words.	AWL covers approximately 8% in the text
Chanchanglek and Sriussadaporn (2009)	Textbook collected from universities running English for engineering courses: 1) Thammasat University 2) Rangsit University and 3) Rajamangala University of Technology	80% of every text comprised of words from the GSL and 5-6% from the AWL
Chung (2009)	The Newspaper Corpus which consists of 579,849 running words.	GSL covers 79.7%
Martinez, Beck, and Panza (2009)	The Agro Corpus – a 826,416-word corpus of research articles in the agricultural sciences.	GSL covers 67.53% and AWL covers 9.06%
Vongpumivitch, Huang, and Chang (2009)	The corpus consists of 200 research articles that have been published in five applied linguistics journals, namely, Applied Linguistics, Language Learning, The Modern Language Journal, Second Language Research and TESOL quarterly. It contains 1.5 millionwords.	AWL covers 11.17%
Li and Qian (2010)	The Hong Kong Financial Services Corpus (HKFSC) which consists of 25 text types (e.g., Annual Reports, Brochure, Fund Description, Ordinances, and Speeches) with 6.3 million running words.	GSL covers 72.63% and AWL covers 10.46%

From these studies, the researchers provided descriptive statistics on both the GSL and the AWL in different text types. These could give some practical guidelines to a researcher and provide some ideas for data analysis of this study. However, Thailand admission tests were another different text type used in this study.

Research Question

The focus of this study is to provide statistical data concerning the vocabulary profiles and coverage of university

admission tests which lead to the research question - What are the lexical profiles and vocabulary coverage of Thailand University Admission Tests?

Significance of the study

The study aims to establish whether the GSL and the AWL can be good references for word selection to help high school students prepare for and be aware in the admission test. The results of this study would significantly provide pedagogical implication for the following parties.

For students, they should know that frequency-based word lists can help them to expand their English vocabulary to handle admission tests. Focusing on the words that frequently appear in the examination is one of the vocabulary learning strategies. This is to confirm that the AWL is important for students because they will encounter these words in the admission test too.

For teachers, high frequency words, both in the GSL and the AWL, are good sources for teaching new vocabulary to high school students. Certainly, the teachers can prepare students not only for using English in everyday life, but also for testing.

For test organizers/designers (NIETS), the results of the study can reveal the validity of a test design particularly in the vocabulary that NIETS had been using and thus prove if those test papers contributed fairness to all test takers or not. If the results of the study show that the percentage of the GSL and the AWL in the test is significantly low, it will be interpreted that there are some problems in the scattering and variation of vocabulary in these tests. As a result, these tests would probably be too difficult for students.

Research Methodology

Data collection

This study included 15 papers of the Thailand university admission test from the years 2007-2011, which comprised 55,161 running words. These papers were designed by NIETS and were up-to-date at the time of this study. Each paper was of single use. Once it was used in a test, it was open to the public for a free download from NIETS's website. The details of 15 papers are shown in Table 2. All tests were typed in plain text to prepare for analysis.

Table 2: The 15 papers used in this study

No.	Tests	Year	Month
1	ONET	2007	February
2	ONET	2008	February
3	ONET	2009	February
4	ONET	2010	February
5	ONET	2011	February
6	ANET	2007	March
7	ANET	2008	March
8	BGAT	2008	October
9	GAT	2009	March
10	GAT	2009	July
11	GAT	2009	October
12	GAT	2010	March
13	GAT	2010	July
14	GAT	2010	October
15	GAT	2011	March

Remarks:
ONET: Ordinary National Education Testing
ANET: Advance National Education Testing (note: this test
       is obsolete)

BGAT: Beta General Aptitude Test (This test is a prototype of the GAT and was used for the Chulalongkorn University Admission Test in 2008.)

GAT: General Aptitude Test

Data analysis

The RANGE and FREQUENCY programs, designed by Nation and Coxhead and programmed by Heatley, Nation and Coxhead (2002), were the major tools used for analyzing the data in this study. The programs can count word frequencies, text length, compare different usages of a word, make indexes and word lists, analyze keywords, and find phrases and idioms. The RANGE program provides text coverage by certain word lists. This study compared the three word lists which included the 1st k GSL, the 2nd k GSL and the AWL. Each list contains the word families; the 1st k GSL (1000 word families), the 2nd k GSL (1000 word families) and the AWL (570 word families). However, Nation and Webb (2011) concluded that the RANGE program has several weaknesses which are as follows:

1. The RANGE program does not distinguish between homographs and homonyms. This is particularly noticeable when one of the members of the homographs is a proper noun, such as Bush, Green, Brown, or Nick.

2. Compound words are dealt with very inconsistently by the RANGE program. Should there be a space or hyphen between the words in compound nouns? Compounds can occur in a variety of forms such as website, web site or web-site. These different forms of the same compound would not all be counted as the same item.

3. Core idioms are not counted as single items, for example 'as well as', 'by and large' and 'such and such'. Each are counted as three separate items.

4. In the RANGE program, an apostrophe is treated as a word break. For example, the various uses of 's, which

can stand for *is*, the possessive, or a letter of the alphabet, are counted in the same family.

5. Some members of word families are very low-frequency items. These family members are usually transparently related to the headword.

In order to analyze data, the units of counting words, Tokens and Types, and Word Families were used in an analysis. According to Schmitt (2010), Tokens are the number of the running words in a text, while Types are the amount of different words. For example, the sentence 'Fat cats eat fat rats', contains five tokens but only four types. *Word Families* – A word family consists of a headword, its inflected forms, and its closely related derived form. This also includes affixes inflected words. For example, teach, taught, teaching, teaches, teacher, teachers, and teachable are in the same word family. In this study, these three are units of the counting words used to present findings.

The following are the steps of data analysis:

a) The total word tokens of the tests were typed in plain text.
b) The test lengths were counted in tokens by the RANGE program.
c) The RANGE program was also employed to find the profiles of the three word lists (i.e. 1K GSL, 2K GSL and AWL) appearing in the test and also the coverage of the three word lists.
d) The FREQUENCY program was employed to count the frequency of the words appearing in the test and to rearrange the order of words by frequency.

Findings and Discussions

In this section, test length, time allocation, lexical profiles, word family appearance, and vocabulary coverage and reading comprehension are presented and discussed.

Test length and time allocation

As a prelude to the findings, Table 3 below shows an overall picture of the Thailand university admission tests regarding their length and time allocation.

Table 3: Test Length and Time Allocation

No.	Exam.	Year	Month	Test Length -Tokens	Test Time - Min	Average -Tokens/Min
1	ONET	2007	February	3,476	120	28.97
2	ONET	2008	February	3,722	120	31.02
3	ONET	2009	February	3,745	120	31.21
4	ONET	2010	February	4,136	120	34.47
5	ONET	2011	February	4,492	120	37.43
6	ANET	2007	March	5,676	120	47.30
7	ANET	2008	March	4,860	120	40.50
8	BGAT	2008	October	3,114	90	34.60
9	GAT	2009	July	3,082	90	34.24
10	GAT	2009	October	2,887	90	32.08
11	GAT	2009	February	3,521	90	39.12
12	GAT	2010	July	2,943	90	32.70
13	GAT	2010	October	3,151	90	35.01
14	GAT	2010	February	3,019	90	33.54
15	GAT	2011	March	3,322	90	36.91
	Total			55,161	1560	
	Average			3,677	104	35.36

From Table 3, the average length of the tests is 3,677 tokens and the average token per minute is 35.36. This means that students are supposed to know at least 35 words per minute as the average speed of reading so that test taking can be done on time.

Comparing between the two major tests, ONET and GAT, the length of the ONET is from 3,476-4,492 tokens and the length of the GAT is from 2,887-3,322 tokens. In other words, there exist more tokens in the ONET than in the GAT. This can be understood by the longer time available in the ONET which can be accepted as reasonable.

Top 100 high frequency words

This section discusses the top 100 highest frequency words.

Table 4: Top 100 High Frequency Words

Word	Order	Frequency	%	Cum %	Word	Order	Frequency	%	Cum %
THE	**1**	**2795**	**5.07**	**5.07**	WOULD	51	126	0.23	39.44
TO	**2**	**1678**	**3.04**	**8.11**	NEW	52	125	0.23	39.67
A	3	1514	2.74	10.85	WHEN	53	122	0.22	39.89
OF	**4**	**1190**	**2.16**	**13.01**	HER	54	120	0.22	40.11
AND	5	969	1.76	14.77	SO	55	120	0.22	40.32
IN	6	966	1.75	16.52	ME	56	119	0.22	40.54
IS	7	809	1.47	17.99	WHO	57	115	0.21	40.75
YOU	**8**	**774**	**1.4**	**19.39**	MY	58	114	0.21	40.96
I	9	664	1.2	20.59	IF	59	113	0.2	41.16
THAT	10	601	1.09	21.68	M	60	111	0.2	41.36
IT	11	595	1.08	22.76	NO	61	111	0.2	41.56
FOR	12	558	1.01	23.77	THERE	62	110	0.2	41.76
S	13	509	0.92	24.7	SHOULD	63	109	0.2	41.96
ARE	14	450	0.82	25.51	**HIS**	**64**	**107**	**0.19**	**42.15**
BE	15	410	0.74	26.25	BEST	65	104	0.19	42.34
HAVE	**16**	**327**	**0.59**	**26.85**	PASSAGE	66	103	0.19	42.53
CAN	17	322	0.58	27.43	CHOOSE	67	101	0.18	42.71
WITH	18	314	0.57	28	OTHER	68	101	0.18	42.9
ON	19	276	0.5	28.5	THAN	69	101	0.18	43.08
NOT	20	270	0.49	28.99	GOOD	70	98	0.18	43.26
WHAT	21	268	0.49	29.48	DON	71	97	0.18	43.43
THIS	22	265	0.48	29.96	SOME	72	96	0.17	43.61
T	23	263	0.48	30.43	WORK	73	96	0.17	43.78
THEY	24	250	0.45	30.89	FOLLOWIN	74	92	0.17	43.95
AT	25	243	0.44	31.33	MAN	75	92	0.17	44.11
AS	26	239	0.43	31.76	MOST	76	92	0.17	44.28
B	27	212	0.38	32.14	OUT	77	92	0.17	44.45
WE	28	205	0.37	32.52	THEM	78	89	0.16	44.61
DO	29	204	0.37	32.89	MAY	79	88	0.16	44.77
THEIR	30	201	0.36	33.25	SHE	80	88	0.16	44.93
WILL	31	197	0.36	33.61	D	81	87	0.16	45.09
YOUR	**32**	**191**	**0.35**	**33.95**	PART	82	87	0.16	45.24
ONE	33	188	0.34	34.29	THINK	83	86	0.16	45.4
WHICH	34	186	0.34	34.63	UP	84	86	0.16	45.55
FROM	35	181	0.33	34.96	GO	85	85	0.15	45.71
AN	36	177	0.32	35.28	TWO	86	84	0.15	45.86
BY	37	174	0.32	35.6	BECAUSE	87	83	0.15	46.01
MORE	38	174	0.32	35.91	DOES	88	81	0.15	46.16
OR	39	173	0.31	36.23	GET	89	81	0.15	46.31
ABOUT	40	166	0.3	36.53	JUST	90	81	0.15	46.45
HOW	41	166	0.3	36.83	ONLY	91	81	0.15	46.6
HAS	42	164	0.3	37.12	BEEN	92	80	0.15	46.74
WAS	43	162	0.29	37.42	NOW	93	80	0.15	46.89
BUT	44	150	0.27	37.69	MANY	94	79	0.14	47.03
HE	45	147	0.27	37.96	WERE	95	79	0.14	47.18
PEOPLE	46	147	0.27	38.22	SEE	96	76	0.14	47.31
ALL	47	146	0.26	38.49	TAKE	97	75	0.14	47.45
ITEMS	48	142	0.26	38.75	HAD	98	74	0.13	47.58
LIKE	49	132	0.24	38.98	ITS	99	74	0.13	47.72
TIME	50	126	0.23	39.21	MUCH	100	72	0.13	47.85

Table 4 shows the top 100 high frequency words appearing in the Thailand University Admission Test. It is noticeable that the frequency of the word in the 1st, 2nd, 4th, 8th, 16th, 32nd and 64th ranks tends to follow Zipf's law. According to Milton (2009), Zipf's

law states that in a corpus of a natural language, the frequency of a word is roughly inversely proportional to its rank in the frequency table. The word which is ranked first in the table is likely to occur twice as often as the word ranked second. Similarly, the occurrence of the word ranked fourth is found twice as often as the word ranked eighth. For example, the frequency of these words '1-THE', '2-TO', '4-OF', '8-YOU', '16-HAVE' and '32-YOUR' are 2795, 1678, 1190, 774, 327 and 191, respectively. We can notice that the frequency of the previous ranked word is likely to occur about twice as often as the word ranked second.

Some letters frequently appear in this study such as A, B, C, D, M, S and T. It is noticed that these texts are found in tests with multiple choices. Then, when the program counts the frequency of words, the letters like a, b, c and d are significantly highly frequent. In addition, the RANGE program separates apostrophized words (_'_) into two words, so the letter 'M' from 'I'm' becomes a high frequency word. This is similar to the letters 'S' or 'T' as in 'it's', 'he's', 'don't' or 'doesn't'.

One more important thing is that from the first 100 high frequency words, it is apparent that most of them are function words or grammar words. Function words, also known as structure words, are words that have a grammatical (or syntactic) role in a sentence or clause as opposed to a lexical meaning. Function words include determiners (such as this, that, some...), prepositions (such as in, on, at), conjunctions (such as or, and, because), interjections, and auxiliary verbs (such as can, had, would, may). By their nature, it is likely that function words could be easily encountered in all kinds of texts, in contrast to content words which are less applicable and can be seen only in a specific text or topic. Therefore, students should focus on studying function words first.

Lexical profiles of Thailand University Admission Test

The lexical profiles of the University Admission Test can be classified by the word frequency list. The General Service List – GSL (both 1K and 2K) and the Academic World List –AWL were used to analyze their profiles appearing in the test.

Table 5: Lexical Profiles of University Admission Test

No.	Exam.	Year	Month	Tokens	1K-GSL	2K-GSL	AWL	Others	Total	GSL	GSL+AWL
1	ONET	2007	February	Tokens	2,755	269	102	350	3,476	3,024	3,126
				%	79.26	7.74	2.93	10.07		87.00	89.93
2	ONET	2008	February	Tokens	2,944	335	98	345	3,722	3,279	3,377
				%	79.10	9.00	2.63	9.27		88.10	90.73
3	ONET	2009	February	Tokens	2,943	298	115	389	3,745	3,241	3,356
				%	78.58	7.96	3.07	10.39		86.54	89.61
4	ONET	2010	February	Tokens	3,323	274	152	387	4,136	3,597	3,749
				%	80.34	6.62	3.68	9.36		86.96	90.64
5	ONET	2011	February	Tokens	3,590	395	145	362	4,492	3,985	4,130
				%	79.92	8.79	3.23	8.06		88.71	91.94
6	ANET	2007	March	Tokens	4,400	363	245	668	5,676	4,763	5,008
				%	77.52	6.40	4.32	11.77		83.92	88.24
7	ANET	2008	March	Tokens	3,820	343	211	486	4,860	4,163	4,374
				%	78.60	7.06	4.34	10.00		85.66	90.00
8	BGAT	2008	October	Tokens	2,301	214	209	390	3,114	2,515	2,724
				%	73.89	6.87	6.71	12.52		80.76	87.47
9	GAT	2009	July	Tokens	2,374	173	156	379	3,082	2,547	2,703
				%	77.03	5.61	5.06	12.30		82.64	87.70
10	GAT	2009	October	Tokens	2,187	228	176	296	2,887	2,415	2,591
				%	75.75	7.90	6.10	10.25		83.65	89.75
11	GAT	2009	February	Tokens	2,669	217	230	405	3,521	2,886	3,116
				%	75.80	6.16	6.53	11.50		81.96	88.49
12	GAT	2010	July	Tokens	2,259	199	141	344	2,943	2,458	2,599
				%	76.76	6.76	4.79	11.69		83.52	88.31
13	GAT	2010	October	Tokens	2,434	244	167	306	3,151	2,678	2,845
				%	77.25	7.74	5.30	9.71		84.99	90.29
14	GAT	2010	February	Tokens	2,323	209	186	301	3,019	2,532	2,718
				%	76.95	6.92	6.16	9.97		83.87	90.03
15	GAT	2011	March	Tokens	2,600	235	195	292	3,322	2,835	3,030
				%	78.27	7.07	5.87	8.79		85.34	91.21
	Total			Tokens	42,922	3,996	2,528	5,715	55,161	46,918	49,446
				%	77.81	7.24	4.58	10.36		85.05	89.63

As shown in Table 5, the GSL generally covers 85.05% of texts and the AWL covers 4.58% of texts. A combination of GSL and AWL covers 89.63% of the texts. Other words which are excluded from the GSL and the AWL cover 10.37% of texts. This group contains low frequency words and proper nouns, for example, agony, Airlanga, Bigfoot, blemish, Bob, Canada, Cocaine, Edward, festive etc.

Focusing on the AWL and comparing to previous studies which focused on academic texts, such as the study of engineering articles by Para (2004) – AWL 12.46%, the study of medical articles by Chen and Ge (2007) – AWL 10.46 %, or the study of applied linguistics journal articles by Vongpumivitch, et al. (2009) – AWL 11.17%, the percentage of the AWL that appeared in the Thailand University Admission Test – 4.58%, is less than half of those found in those studies. On the other hand, the studies about general texts, such as the study of online news by Boonyapapong (2007) – AWL 2.09 % or the studies of general fictions and The Lord of the Rings by Konstatakis (2007) – AWL 1.31% and 0.52%, respectively, show that the averages of AWL coverage are lower than 4.58%. We can see that the AWL coverage in Thailand University Admission Tests is in between general texts and academic texts.

When the tests were divided into 3 groups, ONET, ANET and GAT, and focused only on the AWL, it was found that the percentage of coverage of the AWL was different. For the ONET, the range was from 2.63-3.68%, the ANET's range was from 4.32-4.34% and the GAT indicated a range from 4.79-6.71%. The percentage of coverage of the AWL in the GAT seemed to be double that of the percentage of coverage of the AWL in the ONET. One of the reasons is that the ONET is the test that measures the knowledge of high school students and is designed based on high school curricula. It was also designed by high school teachers. On the other hand, by its different nature, the GAT is the proficiency test that is used to measure student's overall ability to use language. This test is not related to a high school curriculum and is designed by professors from universities. Professors are likely to use more advanced vocabulary or they designed the test based on what students are believed to know in order to achieve academic success at the tertiary level.

Word families appearing in the University Admission Test

This section discusses the percentage of word families from the GSL and the AWL that appear in the tests.

Table 6: Percentage of Word Families Appearing in the University Admission Test

No.	Exam.	Year	Month	1K-GSL	%	2K-GSL	%	AWL (570)	%
1	ONET	2007	February	473	47.3%	144	14.4%	56	9.8%
2	ONET	2008	February	470	47.0%	140	14.0%	58	10.2%
3	ONET	2009	February	468	46.8%	139	13.9%	65	11.4%
4	ONET	2010	February	428	42.8%	106	10.6%	70	12.3%
5	ONET	2011	February	431	43.1%	124	12.4%	58	10.2%
6	ANET	2007	March	556	55.6%	174	17.4%	124	21.8%
7	ANET	2008	March	535	53.5%	179	17.9%	116	20.4%
8	BGAT	2008	October	456	45.6%	146	14.6%	124	21.8%
9	GAT	2009	July	463	46.3%	120	12.0%	91	16.0%
10	GAT	2009	October	458	45.8%	127	12.7%	89	15.6%
11	GAT	2009	February	496	49.6%	143	14.3%	119	20.9%
12	GAT	2010	July	429	42.9%	117	11.7%	94	16.5%
13	GAT	2010	October	464	46.4%	117	11.7%	107	18.8%
14	GAT	2010	February	436	43.6%	103	10.3%	96	16.8%
15	GAT	2011	March	469	46.9%	129	12.9%	98	17.2%
	Total			**943**	**94.3%**	**659**	**65.9%**	**411**	**72.1%**

Taking into consideration the following word lists, it can be summarized that there are 943 families (94.3%) from 1,000 families of the 1K-GSL appearing in the text, 659 families (65.9%) from 1,000 families of the 2K-GSL appearing in the text and 411 families (72.1%) from 570 families of the AWL appearing in the text.

Frequency, coverage and reading comprehension

This section discusses the vocabulary size that needs to meet 95% coverage for reasonable comprehension in reading.

Table 7: Frequency and Text Coverage

Vocab Size (Fanilies)	Percentage of Coverage	Cumulative Percent
0-1000	77.18	77.18
1001-2000	9.40	86.58
2001-3000	4.93	91.51
3001-4000	3.31	94.82
4001-5000	1.81	96.63
5001-6000	1.81	98.44
6001-7000	1.56	100.00

Table 7 illustrates that the high frequency words (0-1000) greatly contribute to text coverage (77.18 %) and the low frequency words contribute the least. Nation (2001) and Milton (2009) claimed that the first 2,000 word families would cover 80% of text. Comparing to this study, the first 2,000 word families cover 86.58% of text.

However, in order to meet 95% coverage of text for reasonable comprehension as recommended by Milton (2009), students need to know approximately 4,000 word families. With the vocabulary size of 4,000 the accumulative percentage is 94.82% which is close enough to 95%. In other words, the 4,000-5,000 word level is the vocabulary size that students need to know to meet 95% coverage for reasonable comprehension.

Conclusions

This paper focuses on the lexical profiles of vocabulary in Thailand University Admission Tests. It is found that the GSL generally covers 85.05% of texts and the AWL covers 4.58%. We can conclude that these tests are sufficient in terms of vocabulary scatter and diversification. A combination of the GSL and the AWL covers 89.63% of the texts. It is useful for students to learn these words and accumulate an adequate stock of vocabulary to handle the tests. The others are low frequency words and proper nouns. If

students are trained to guess meaning from contexts and the suitable contexts are given in the test, they can guess the meaning and handle the test.

Pedagogy and testing implication

This study can prove that both the GSL and the AWL comprise high portions of the tests and are important for students to learn them, especially the GSL. It is impossible to understand English in the tests without knowing these words. GSL words are worth spending time to learn because they are found repeatedly in tests. By learning the combination of the GSL and the AWL, students will gain enough proportion of vocabulary required for handling the tests. Because the GSL and AWL are considered high frequency word lists, students are encouraged to study both. High frequency words are encountered in a wide range of language uses, including testing. In other words, teachers and students should not spend so much time on low frequency words which are rarely present in both testing and everyday life.

Some teachers might feel it difficult to find material to be used in the classroom. Teachers can adapt the reading passages that contain both the GSL and the AWL from the tests to teach in classrooms. The teacher can prepare both the GSL and the AWL in the pre-teaching stage and let students read through the lesson's passage later. This also can help students to be familiar with passage readings that are at the same level as the tests. Teachers should consider developing vocabulary knowledge in the preparation stage of teaching reading, especially high frequency words.

the same time, NIEST, as a test designer, should give careful attention to word selection in terms of actual word use and coverage when designing tests. According to Hughes (2007), vocabulary plays different roles depending on the type of test.

Testing vocabulary proficiency is still useful in most of the tests. One reason for this must be the ease with which large numbers of items can be administered and scored within a short period of time. That is why the GAT still contains a vocabulary section. Whereas in the placement test, all we would be looking for is some general indication of the adequacy of the students' vocabularies. It can be said that the GSL and the AWL scattering in tests would be one indication to show the quality and level of tests.

Recommendation for future research

An additional amount of tests could be added to the data pool to increase validity. Thailand has had standard university entrance exams since 1962, so a longitudinal study can be done in order to compare tests periodically. For example, we can compare the differences in lexical profiles every ten years.

Not only the exams from the central system can be studied, but also the quota exams from regional universities such as Chiangmai University, Khonkaen University and Prince of Songkla University could be added to a future study.

This study was conducted in 2013 and it was based on two word lists, the GSL in 1953 and the AWL in 2000. Consequent to that, many researchers have tried to compile and create new word lists based on current corpora. For example, Browne, Culligan and Phillips (2014) have created a New General Service List (NGSL) of core vocabulary for second language learners. The words in the NGSL represent the most important high frequency words of the English language for second language learners and are a major update of Michael West's 1953 GSL (Browne, 2014). The other high frequency word list is a New Academic Vocabulary List (NAVL) proposed by Gardner and Davies (2014). The NAVL is derived from a 120million-word academic sub-corpus of the 425 million-word Corpus of Contemporary American English (COCA). They claimed that the NAVL used more texts and more coverage

than the AWL. Other researchers are recommended to make use of these potential world lists, the NGSL and the NAVL (2014), as a framework for their future studies in order to analyze the Thailand University Admission Tests.

In addition, the results from this study can be compared with the studies of actual vocabulary knowledge or vocabulary level of high school students by using a vocabulary level test and its effects on students' comprehension. We can compare the lexical profiles of Thailand University Admission Tests with the level of vocabulary knowledge of high school students. We can also compare this study to the vocabulary input of the ESL materials or course books employed by high schools, so as to ascertain their pedagogical suitability.

The Authors

Wirun Cherngchawano is a graduate student in Applied Linguistics (English Language Teaching) at King Mongkut's University of Technology Thonburi (KMUTT). He is working as a business development manager in a multinational firm and spends his free time teaching English to high school students. His interests include the studies of English vocabulary learning and teaching, and corpus linguistics.

Natjiree Jaturapitakkul, Ph.D. is an English lecturer in the Department of Language Studies, School of Liberal Arts (SoLA), King Mongkut's University of Technology Thonburi, Thailand. She is currently the Head of Testing Unit of the SoLA. Her research interests include English language teaching, language assessment and evaluation, test development, and ESP testing.

References

Association of University Presidents of Thailand. (2014). *Regulation of University Admission*

Academic Year 2014. Retrieved September 1, 2014 from http://www.cuas.or.th/admissions.html

Bauman, J. (1995). *About the General Service List.* Retrieved January 10, 2014, from http://jbauman.com/aboutgsl.html

Boonyapapong, K. (2007). *An analysis of academic vocabulary in business news: A corpus-based study.* Unpublished master's thesis, Srinakharinwirot University, Bangkok, Thailand.

Browne, C. (2014). A new general service list: The better mousetrap we've been looking for*? Vocabulary Learning and Instruction, 3*(1), 1-10. Doi: 10.7820/vli.v03.1.browne

Browne, C., Culligan, B. & Phillips, J. (2014). *A New General Service List (1.01).*Retrieved September 11, 2014 from http://www.newgeneralservicelist.org

Chanchanglek, S. & Sriussadaporn, N. (2009). A corpus analysis of English vocabulary input in course materials used for engineering students. *Humanities Journal, 18*(1), 141-149.

Chen, Q. & Ge, G. (2007).A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes, 26*, 502-504.

Chung, M. (2009). The newspaper word list: A specialised vocabulary for reading newspapers. *JALT Journal, 31*, 159-182.

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

Gardner, D. & Davies, M. (2014). *A New Academic Vocabulary List.* Retrieved September 11, 2014 from http://applij.oxfordjournals.org/content/early/2013/08/02/applin.amt015

Heatley, A., Nation, P. & Coxhead, A. (2002). *RANGE and FREQUENCY programs.* Retrieved December 10, 2013, fromhttp://www.victoria.ac.nz/lals/resources/range

Hu, M. & Nation, P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language*, *23*, 403-430.

Hughes, A. (2007). *Testing for language teachers (2nd ed.).* Cambridge*:* Cambridge University Press.

Konstatakis, N. (2007). Creating a business word list for teaching business English, *Estudios de Linguistica Inglesa Aplicada (ELIA), 7,* 103-126.

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

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

Lessard-Claouston, M. (2012). Word lists for vocabulary learning and teaching. *The CATESOL Journal, 24*(1), 287-304.

Li, Y. & Qian, D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *SYSTEM, 38*, 402-411.

Martinez, A., Beck, C. & Panza, B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*, 183-198.

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

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

Nation, P. (2010). *Teaching vocabulary strategies and techniques.* Heinle Cengage Learning.

Nation, P. & Webb, S. (2011). *Researching and analyzing vocabulary.* Heinle Cengage Learning.

Para, C. (2004). *A corpus study of high-frequency words in civil engineering research articles: Sub disciplinary differences between structure and transportation.* Unpublished master's thesis, Mahidol University, Nakhon Pathom, Thailand.

Poonpon, K. (2002). *Vocabulary input in English for Science course: A corpus analysis of intensive and extensive course materials.* Unpublished master's thesis, Mahidol University, Nakhon Pathom, Thailand.

Schmitt, N. (2010). *Researching vocabulary – A vocabulary research manual.* Palgrave Macmillan.

Schmitt, N., Jiang, X. & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26-43.

Thepwiwatjit, N. (2008). *Frequency of academic words in food science articles.* Unpublished master's thesis, King Mongkut's University of Technology Thonburi, Bangkok, Thailand.

Vongpumivitch, V. Huang J. & Chang Y. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes, 28*, 33-41.

West, M. (1953). *A general service list of English words.* London: Longman, Green and Company.

Wilkins, A. (1972). *Linguistics and language teaching.* London: Edward Arnold.