

# Corpus-Based Vocabulary Analysis of Rock Guitar Lessons and a Sample Dictionary

Nantakarn Impong and Jirapa Vitayapirak

**Abstract**—This study explored in-depth to the language of rock guitar lessons. The aims of this study are to find out terminologies used in rock guitar lessons and to design a sample bilingual dictionary for Thais. The user needs analysis was done to receive information about reading materials to compile the Rock Guitar Lessons Corpus (RGL). The RGL Corpus consists of 1,356,029 words (tokens). The concordance software “WordSmith Tools”, and “RANGE\_GAL\_AWL Programs” were used.

The word lists were classified into three groups of general, academic, and technical vocabularies. The function words, content words, collocation, and abbreviations were also analyzed. The total word types in the corpus were 40,542 and the types and tokens ratio was 1:33.44. The results showed that general vocabulary occurred at the highest frequency. Some general and academic words were used as technical terms. Many multi-word terms were frequently found when using KWIC concordance. These terms were formed by combination of general and academic vocabularies. Symbols and abbreviations were significantly found in the corpus, i.e. clippings, initials, acronyms, contractions, and substitutions. The word frequencies, word combination and typical usage were used to create sample English-Thai dictionary entries of rock guitar.

**Index Terms**—Corpus, rock guitar, dictionary, vocabulary, terminology, KWIC concordance.

## I. INTRODUCTION

Rock music is a kind of music which is popular in the world and Thai society. It started in Thai society in the period of Vietnam War (1964-1972) due to the United States of America sending troops to Vietnam using Thailand as the military base in 1964. There were many US army camps around Thailand. One of the famous entertainment during the war period for the American soldiers was string music bands in the bar and night club. Western popular music was usually played to serve the American soldiers and started to be popular among the Thai musical audience. The electric guitar is the major instrument in rock music and it has been popular from then onto the present.

At present, rock music is very popular in Thailand. There are many multimedia like teaching guitar through the videos, guitar books and also various English music notes. They were launched to whoever was interested in playing rock music [1]. Many academic institutions in Thailand such as

College of Music, Mahidol University, Silpakorn University and Rangsit University include the electric guitar into their curriculum. According to Thai university library statistics in the year 2012, there are many guitar references or dictionaries of music written in English. However, very few references about rock music were written in Thai. It is difficult for Thai musicians to clearly understand the rock guitar terms. Therefore, this study aims to investigate the important vocabularies and compile a bilingual (English-Thai) rock guitar dictionary for Thai students who love playing rock guitar. As [2] stated that the dictionary is a part of vocabulary learning. For foreign learners of English like Thais, they should be encouraged to avail themselves of the substantial information contained in their dictionary.

The rock guitar terms are a specialized form of natural language used to describe a limited subject matter or a sublanguage, generally employed by a group of specialists among guitar players dealing with rock music. According to [3], the set of sentences in a sublanguage is a subset of the sentences in a whole language [4]. Most guitar players always use the sublanguages in terms of specific terminology such as *ARPEGGIO*, *PENTATONIC SCALE*, *MELODIC MINOR SCALE*, and so on in their community.

According to [5], the English vocabulary in academic text can be divided into three categories: General Service List (GSL) or high frequency words, academic vocabulary (AWL), and Off-list words (including technical and low frequency words). On average, the GSL and AWL vocabulary represent approximately 90% of running words in academic text. The technical vocabulary covers around 5% and the other low frequency words such as proper name, abbreviation, and numbers represents around 5% of the total words in the text.

In Thailand, there have been no previous studies about the vocabulary levels of rock guitar, especially the technical vocabulary. This research thus focuses on compiling corpus to find out rock guitar terminologies. A corpus is a large body of machine — readable texts and a collection of linguistic data, either compiled as written texts or as a transcription of recorded speech [6], [7] found that corpus-based study is another way to use computer readable forms in doing the linguistic research with less effort. It has many advantages in using computer readable such as finding, classifying, retrieving, and calculating data with high speed. The corpus-based analysis has also never been used for analyzing rock guitar lessons. Therefore, this study used corpus-based analysis to study rock guitar lesson texts and lead to design a sample dictionary of rock guitar lessons.

## II. OBJECTIVES

- 1) To analyze technical vocabulary used in rock guitar

Manuscript received September 1, 2014; revised January 23, 2015.

Nantakarn Impong is with the Dept. of Applied Linguistics (English for Science and Technology), King Mongkut's Institute of Technology, Ladkrabang, Thailand (e-mail: Modrockguitar@gmail.com).

Jirapa Vitayapirak is with the Dept. of Applied Arts (English), Faculty of Industrial Education, King Mongkut's Institute of Technology, Ladkrabang, Thailand (e-mail: kvjirapa@kmitl.ac.th).

lessons.

- 2) To identify abbreviations and symbols found in rock guitar lessons.
- 3) To examine collocations found in rock guitar lessons.
- 4) To design a sample dictionary entries based on the technical vocabularies found in the corpus.

### III. MATERIALS AND METHODS

#### A. Data Collection

The surveys of rock guitar lessons in the top ten leading Universities in music education were used to elicit information about the English guitar lessons that they encountered and these selected lessons were used to compile the RGL corpus.

Three hundred rock guitar lessons from guitar magazines and guitar instruction books during the year 2009-2013 were collected as a sample in this study. In other words, the 150 rock guitar lessons from guitar magazines and 150 rock guitar lessons from guitar instruction books were selected. The weighting of two samples is equal because both of them are equally used among guitar players.

#### B. Research Instruments

In this study selected three instruments: Wordsmith Tools Version 6.0 [8], RANGE\_GAL\_AWL Programs by [9] and AntConc 3.4.3 by [10]. The first program used was Wordsmith Tools Version 6.0. It was used to analyze basic statistics of word analysis in terms of types, tokens, word frequency and concordance. The second program used was RANGE\_GAL\_AWL Programs. It was used to calculate the vocabulary levels: General vocabulary, Academic vocabulary, Off-list words refer to Technical vocabulary, Abbreviation, and Symbols. The last program used was AntConc 3.4.3. It was used to calculate compounds and multiword terms.

#### C. Data Analysis

In this study, the Wordlist, cluster and concordance tools were used to generate lists in frequency order for lexical comparison of texts. In data processing, the frequency and distribution of word types and tokens in the RGL Corpus were first determined. Since the focus of this research was on the terminologies. In order to find out important terms, the three word types, i.e. general (GSL), academic (AWL), and off-list words were analyzed by using RANGE\_GAL\_AWL Programs. Then, the abbreviations, symbols, and collocation were identified.

### IV. RESULTS AND DISCUSSION

#### A. Statistical Analysis of the RGL Corpus

Table I below shows the statistics or the output of the lexicon extraction program, showing the size of the lexicon produced from the RGL Corpus.

Statistics can provide various kinds of summary of the contents of the RGL corpus. They can show the Type/Token ratio of the whole vocabulary, which is computed by dividing the number of the tokens by the number of types. It indicates the relative concentration or dispersion of the vocabulary, and offers a measure of its diversity. The lower

of type-token ratio, the greater the diversity of words in the corpus.

Text file	Overall
file size	14,465,223
tokens (running words) in text	1,356,029
tokens used for word list	1,158,603
sum of entries	
types (distinct words)	40,542
type/token ratio (TTR)	3.50
standardized TTR	31.76
STTR std.dev.	66.62
STTR basis	1,000
mean word length (in characters)	3.54
Word length std.dev.	2.47
sentences	90,564
mean (in words)	14.70
std.dev.	573.84
paragraphs	140
mean (in words)	8,275.74
std.dev.	20,862

TABLE II: TOKENS AND WORD TYPES IN THE RGL CORPUS

Tokens	1,356,029
Word types	40,542
Type/token ratio	1:33.44

In Table II the overall corpus size comprised of 1,356,029 tokens or running words. The total number of word types equals 40,542 words in the corpus since a recurrent word is counted only once. For the whole corpus, the ratio of types/tokens is 1:33.44. The ratio indicated that each word was repeated nearly 34 times on average throughout the corpus.

#### B. Vocabulary Levels

It was found out that the three types of vocabulary used in the rock guitar lesson corpus were 13.7% of general vocabulary, 10.5% of academic vocabulary and 75.8% of off-list words. The off-list words could be divided into 8.32% of technical vocabulary, 44.6% of abbreviations, and 22.87% of symbols as shown in Fig. 1 below:

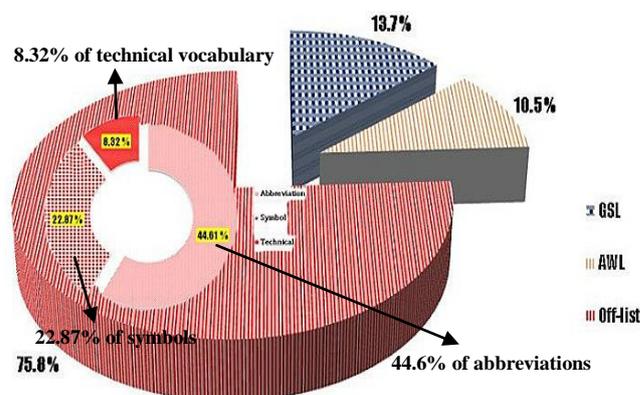


Fig. 1. Proportion of general, academic, and off-list vocabularies.

This finding did not align itself closely with the figures

cited by [5], [9] that on average, the GSL and AWL vocabulary represent approximately 90% of running words in academic text. The technical vocabulary covers around 5% and the other low frequency words such as proper name, abbreviation, and numbers represents around 5% of the total words in the text. We can see that the rock guitar lesson is a text-type that covers a large amount of off-word lists, i.e. abbreviations, symbols, and technical terms.

In terms of word classes, two major types of word classes, namely function words and content words were found in the corpus. The function words were found at the top ten word frequency lists in the RGL Corpus such as ‘THE’, ‘A’, ‘AND’, ‘TO’, ‘OF’, and ‘IT’. Interestingly, abbreviations such as ‘I’, ‘L’, ‘E’ occurred in high frequency as well, as shown in Table III below:

TABLE III: THE TOP 10 WORD FREQUENCY LISTS IN THE RGL CORPUS

Rank	Word	Frequency	% of occurrence
1.	THE	55,214	4.07
2.	I	52,844	3.90
3.	A	32,138	2.37
4.	AND	22,677	1.67
5.	TO	21,073	1.55
6.	OF	18,896	1.39
7.	IN	16,817	1.24
8.	L	15,095	1.11
9.	YOU	13,313	0.98
10.	E	11,860	0.87

### C. Content Words

Content words are those that belong to the large, open classes of the language including: noun, verbs, adjectives, and adverbs.

Table IV shows the ten most frequent content words found in the RGL corpus:

TABLE IV: THE TOP 10 CONTENT WORDS IN THE RGL CORPUS

No.	Rank	Words	Freq.	%
1.	19	GUITAR	7,180	0.53
2.	29	CHORD	5,456	0.40
3.	37	MAJOR	4,628	0.34
4.	38	BLUES	4,627	0.34
5.	40	SCALE	4,361	0.32
6.	41	MINOR	4,324	0.32
7.	44	STRING	3,832	0.28
8.	45	CHORDS	3,809	0.28
9.	47	NOTE	3,767	0.28
10.	55	NOTES	3,600	0.27

For a dictionary of Rock Guitar, we shall be concentrating on high frequency nouns among content words. As shown in Table III, general words of music appeared at the high frequency such as ‘GUITAR’ (7,180 times), ‘CHORD’ (5,456 times), ‘MAJOR’ (4,628 times), ‘BLUES’ (4,627 times), and so on.

### D. Compounds and Multiword Terms

The RGL concordances show that most of the nouns are

compounds and multiword terms. Most of them consist of nouns, with proceeding nouns functioning as adjectives, e.g. blues guitar, chord progression, bar blues, etc. We can see that although there are two or more lexemes, the parts function as a single item, with its own meaning. Table V below shows the top ten compounds in the rock guitar lesson corpus that were chosen from the highest frequency of clusters.

TABLE V: THE TOP 10 COLLOCATIONS IN THE RGL CORPUS

No.	Word (Freq.)	Collocations	Frequency
1.	GUITAR (7,180)	BLUES GUITAR	1,053
2.	CHORD (5,456)	CHORD PROGRESSION	369
3.	MAJOR (4,628)	MAJOR SCALE	1,317
4.	BLUES (4,627)	BAR BLUES	540
5.	SCALE (4,361)	MAJOR SCALE	904
6.	MINOR (4,324)	MINOR SCALE	449
7.	STRING (3,832)	E STRING	119
8.	CHORDS (3,809)	POWER CHORDS	175
9.	NOTE (3,767)	EIGHT NOTE	114
10.	NOTES (3,600)	QUARTER NOTES	55

### E. Technical Vocabularies

In terms of technical vocabularies with high frequency in RGL corpus, the terms such as ‘PENTATONIC’ (1,496 times), ‘ARPEGGIO’ (850 times), and ‘HARMONIC’ (848 times) appeared at the high frequency as shown in Table VI below:

TABLE VI: THE TOP 10 TECHNICAL VOCABULARIES IN THE RGL CORPUS

No.	Rank	Words	Freq.	%
1.	111	PENTATONIC	1,496	0.11
2.	173	ARPEGGIO	850	0.06
3.	174	HARMONIC	848	0.06
4.	210	MELODIC	679	0.05
5.	214	DIMINISHED	664	0.05
6.	271	DOMINANT	511	0.04
7.	318	MODES	437	0.03
8.	332	TAPPING	422	0.03
9.	369	DORIAN	379	0.03
10.	371	CHROMATIC	377	0.03

When observing in-depth about the nature of the specific language of rock guitar lessons, those words do not stand alone. They frequently occurred with other words such as the word ‘PENTATONIC’ always collocates with the word

‘SCALE’ (PENTATONIC SCALE) and used as a compound noun. ARPEGGIO always collocates with the word ‘MAJOR’ or ‘G MAJOR’. It is used as the multi-word terms such as ‘G major arpeggio’ as shown in KWIC concordance as in Fig. 2 below.

For the study of collocation, KWIC concordance is an excellent tool to help finding out the specific collocation of terms or technical vocabularies used in playing rock guitar. Fig. 3 shows the sample of collocation of the word ‘MINOR’ as ‘melodic minor scale’:

HING ROCK St BLUES GUITAR BOOK Q nouns 9-s G-Major arpeggio, long fingering TRRCH 3'1 ;+'?# L.-\$-J \_ 4 : neck. Any combination of G, B, D will make a G-Major arpeggio. noun: 9-a is an unorthodox fingering that s you come up with nouns 12-11. Q noun: 12-10 BMajor arpeggio swept TRRCK S3 tx: 19- F- T J ' ij Ire 1 n - D \$5 3: mun: 15-11 C-Major scale ascending, C-Major arpeggio descending -5:? TRHEK 9? , Ii) 5&5 TT:<;mI\*~F ord 14. Dom7(b9) Chord 15. Dom7(#9) Chord Pure Major Arpeggio Major 7 Arpeggio Major 6 Arpeggio Pure Minor hat you already know. For example, here is a G major arpeggio based on the 6th- string root G barre chord, on the 6th- string root G barre chord, and a C major arpeggio based on the 5th-string root C barre chord.

Fig. 2. KWIC Sample of collocation of the word ‘G MAJOR ARPEGGIO’ in the RGL corpus.

This is one nftthc modes ofthe. melodic minor scale. Youll see more of these modes ARPEGGIO ON 6 This is simply a melodic minor scale without the 4th and 6th. Anyli 1 6th. Anyline you can use the melodic minor scale you can use this arpeggio. Its WITH MAJOR 7 ARPEGGIO Play the melodic minor scale without the 4th or 6th. an W ["1 ie with a I12. Actually, its a melodic minor scale starting on b3, and yes, it is a ?7 This is another mode. of the melodic minor scale For Csus4(,9), play the B8 mulc ar ??h d?grec, For C7, play a G melodic minor scale yes, its another mode of melod .-8 Here is anozher mode of the melodic minor scale. Over altered dominant chords,

Fig. 3. KWIC sample of collocation of the word ‘Melodic Minor Scale’.

Table VII in below shows the high frequency technical noun phrases or multi-word lexical units taken from the top 20 high-frequency content words.

TABLE VII: TECHNICAL MULTI-WORD LEXICAL UNITS IN RGL CORPUS

No.	Technical Noun Phrases or Multi-Word Lexical Units	Frequency
1.	MINOR PENTATONIC SCALE	127
2.	A MINOR PENTATONIC	67
3.	MINOR SCALE PATTERN	66
4.	C MAJOR SCALE	65
5.	MELODIC MINOR SCALE	63
6.	HARMONIC MINOR SCALE	60

Regarding collocations, both lexical and grammatical collocations were found in this study. Most technical compounds and multi-word terms consisted of words from GSL, AWL, and technical vocabularies. From Fig. 3 and Fig. 4, it can be seen that the words were not used in general sense but technical musical meaning such as ‘melodic minor scale’.

F. Abbreviations and Symbols

Abbreviations were found frequently in the rock guitar lessons. In this study, we found many symbols and 5 main types of abbreviations, i.e. 1) clippings, 2) acronyms, 3) initialisms, 4) contractions, and 5) substitutions as shown in Table VIII below:

TABLE VIII: THE 10 HIGH-FREQUENCY ABBREVIATIONS IN RGL CORPUS

No.	Words	Meaning	Freq.	%
1.	E	E CHORD	11,860	0.87
2.	H	HAMMER ON	10,403	0.77
3.	F	F CHORD	7,751	0.57
4.	G	G CHORD	7,528	0.56
5.	M	MUTING	7,441	0.55
6.	T	TAPPING	7,398	0.55
7.	C	C CHORD	7,046	0.52
8.	S	SLIDE	7,019	0.52
9.	D	D CHORD	5,791	0.43
10.	B	BENDING	5,035	0.37

The high frequency abbreviations include ‘E’ (E Chord- 11,860 times), ‘H’ (Hammer on – 10,403 times), and ‘F’ (F Chord – 7,751 times) and so on.

The top five high frequency symbols in RGL corpus are in Table IX below.

G. The Corpus Inputs to the Dictionary

The corpus is a primary source of information about the way words behave. It shows the frequency of words, grammatical information, and collocations. The main factor in deciding whether or not to include a word as an entry in the dictionary is how often it occurs in the corpus. The

sample sentences from the concordance are the main sources of example sentences presented in the dictionary [11].

1) *Headword selection*

The information about the word frequency is very important for choosing or grading vocabularies. Therefore, the single words, compound words, multi-words, abbreviations and symbols with highest frequency were

selected as the samples of rock guitar dictionary.

2) *The design of a sample dictionary*

An Entry Structure:

The entry ‘*Alternate Picking*’ below offers as a headword, pronunciation in Thai, Abbreviation, Thai synonym, Thai definition, example of English usage in a sentence taken from the concordance and illustration as shown in Fig. 4 below:

TABLE IX: THE 5 HIGH-FREQUENCY SYMBOLS IN RGL CORPUS

No.	Symbols	Meaning	Freq.	%
1.		ALTERNATE PICKING	8,579	28.34
2.		HAMMER ON & PULL OFF	4,850	16.02
3.		SLIDE	2,576	8.51
4.		BENDING	2,214	7.31
5.		PULL OFF	1,584	5.23

**Alternate Picking** /อัล-เทอร์-นิกทีฟ พิก-กิ้ง/ (อ่าน Alt.or Altern.)

**การตีคอร์ดสลับมือ**

การสลับด้วยขั้วกลางของนิ้ว ทาบนิ้วหัวแม่มือขวาสลับลง นิ้วหัวแม่มือซ้ายจะถูกลีลาเช่นเดียวกับ 1 ลีลา 2 ลีลา 3 ลีลา 4 ลีลา 1 ลีลา (ใบคอร์ดที่เป็นนิ้วชี้กับนิ้วกลาง) เดี่ยวที่ใส่จะหนักแน่น ทาบจะย้ายไปคอร์ดแล้วคอร์ดแล้วเร็วหนักฝึกฝนจนคล่องแคล่วจะไปถึงขั้นเขียนเพลงด้วยขั้วกลางเช่นนี้

Alternate picking is a big part of my guitar playing style.



Fig. 4. The sample entry of RGL dictionary.

V. CONCLUSION

The purposes of this study were to compile a corpus of the Rock Guitar Lessons and analyze the word frequency

lists, technical vocabularies, abbreviations, symbols, and collocations. The design of a sample rock guitar lesson dictionary was proposed. The results showed that there were 8.32 % of technical terms in the RGL corpus. Some general

and academic words were used as technical vocabularies. Many multi-word terms were frequently found in the texts when using KWIC concordance. Interestingly, symbols and abbreviations were significantly found in the RGL corpus, i.e. clippings, initials, acronyms, contractions, substitutions. The corpus findings on word frequencies, word combination and typical usage were used to create sample entries for the bilingual English-Thai dictionary of rock guitar. The findings thus have applications for a design of a rock guitar dictionary for musical students, musicians, guitar players, and those who are interested in playing rock guitar.

#### ACKNOWLEDGEMENTS

This work was supported in part by Thai Government Fellowship (SP2), Office of Vocational Education Commission. I would like to express my gratitude to Assoc. Prof. Dr. Jirapa Vitayapirak, my advisor, for her valuable suggestions and devotion of time in revising this paper. I would also like to thank Miss Piyanuch Charoen for her advice on software use, other suggestions and encouragement.

#### REFERENCES

- [1] S. Wiwan, "Effect of Yngwie Malmsteen's guitar style on the Thai guitarists," M.A. thesis, Mahidol University, 2004.
- [2] D. Summer, "The role of dictionary in language learning," in *Vocabulary and Language Teaching*, R. Carter and M. McCarthy, Eds. Longman, USA, 1988.
- [3] Z. Harris, *Mathematical Structures of Language*, New York: Wiley (Inter Science), 1968.
- [4] J. Vitayapirak, "A corpus for a dictionary of computer science," presented at the 13th World Congress of Applied Linguistics (AILA 2002): Applied Linguistics in the 21st Century: Opportunities for Innovation and Creativity, Singapore: Suntec City, 2002.

- [5] A. Coxhead, "A new academic word list," *TESOL Quarterly*, vol. 34, no. 2, pp. 213-238, 2000.
- [6] D. Crystal, *An Encyclopaedia Dictionary of Language and Linguistics*, Cambridge: Cambridge University Press, 1992.
- [7] P. Mallikamas, *Application of Corpora in Language Teaching*, Thai TESOL BULLETIN, pp. 1-17, 1999.
- [8] M. Scott, *Oxford Word Smith Tools*, Oxford: Oxford University Press, 2012.
- [9] P. Nation. (2005). Range program with GSL and AWL. [Online]. Available: <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- [10] L. Anthony. (2014). AntConc (Windows, Macintosh OS X, and Linux) build 3.4.3. [Online]. Available: <http://www.laurenceanthony.net/software.html>
- [11] J. Vitayapirak, "A Corpus-based approach to ESP lexicography: The case study of English for Thai computer science students," PhD. thesis, Macquarie University, 2001.

**Nantakarn Imphong** was born on November 1, 1976 in Chaiyaphum, Thailand. In 2000, he graduated from Ramkhamhaeng University. He had received his bachelor degree in the field of arts (English). Then in 2010 he got a master degree in the field of education (guidance) from Sukhothai Thammathirat Open University, Thailand. At present, he is a Thai government English teacher at Minburi Technical College, Bangkok Thailand. His research interests include corpus, lexicography, and English teaching.



**Jirapa Vitayapirak** works at the Department of Applied Arts (English), Faculty of Industrial Education, King Mongkut's Institute of Technology Ladkrabang, Thailand. In 1986, she got a M.A. degree in applied linguistics, King Mongkut's University of Technology Thonburi, Thailand. In 2001, she got a Ph.D. degree in linguistics, from Macquarie University, Australia.

In 1996 she was the associate dean for academic affairs, Faculty of Faculty of Industrial Education, King Mongkut's Institute of Technology Ladkrabang. Her research interests are corpus lexicography, ESP, ELT and terminology. In 2008 to 2009 she had been the president of the Asian Association for Lexicography (ASIALEX).