Development of English Musicology Word List

Soitong Kongnui[1*], Supakorn Phoocharoensil[2]
[1]Language Institute, Thammasat University, Thailand
[2]Professor, Language Institute, Thammasat University, Thailand
*kongnuistark@gmail.com, supakorn.p@litu.tu.ac.th

ABSTRACT

The aim of this study was to collect the most frequent content words found in music and music education research articles, and create a list that was validated by music experts. The music and music education research article corpus was compiled from 300 articles published during January 2015 – February 2021 from nine international journals of SCOPUS and one expert-recommended journal. The music and music education research article corpus consists of 1,938,216 words. The list's word selection criteria were based on Coxhead's frequency criterion for Academic Word List. Every word family occurring 57 and more than 57 times were included in the word list. The AntWordProfiler (1.4.0) software was used to analyze the frequency of words and to generate the word list. The results revealed that 67.73 percent of the MWL's words belonged to 1st 1,000 of GSL, 5.52 percent of words belonged to 2nd 1,000 of GSL, 10.31 percent were AWL, and the other 16.44 percent did not belong to previous word lists. There are 516 high frequency content words after excluded function words, proper names, and meaningless words. Following that, three musicologists were asked to justify the words in the musicology word frequency list. After the justification process, 378 terms (72.87%) were deleted from the list, which was then classified into word families. The word list called the Musicology Word List (MWL) was generated. The MWL aims to provide key terms for reading research papers to music students and people who are interested in music in their preparation for musicology-related disciplines. MWL can also be used in vocabulary pedagogy to help music teachers choose appropriate vocabulary to educate students.

Keywords

corpus-based study; word list; musicology; music and music education research articles

Introduction

Nowadays, education in Thailand has greatly improved. The education at universities and high schools in Thailand is divided into various branches to meet the needs of the students who are interested in studying in different subject areas, which include the music field. Learning Western music can be divided into several sub-subjects which are taught using the same standards around the world (Smith C., 1962). Most content in the various subjects of music is in the form of English.

Reading academic texts in English is one of the most essential requirements for students who are studying English as a foreign language (EFL) and English as a second language (ESL). One of the necessities of music learners and also important for studying music is to read English music research articles. The survey results from previous study show that reading is still a problem for music students (Sun, 2006). Music education in the English language for music students not only requires knowledge of general English vocabulary, but also knowledge of that word in musical terms. So, it is necessary to have a good vocabulary to achieve success in reading comprehension. For learners of English for specific purposes (ESP) , knowledge of specialized vocabulary is very important. Students studying in different fields need to learn different vocabulary, meaning that word lists that meet the specific needs of those learners are in demand. In other words, developing academic word lists for specialized areas is beneficial to the

overall learning outcomes of that area (Ward, 2009). University lecturers and students have had a growing interest in the field of music as well as reading and writing musical articles in English. Although many word lists have been studied in a wide variety of disciplines such as science (Boonyos, 2014; Liu and Han, 2015), medical science (Wang et al., 2008), engineering (Ward, 2009), and political science (Srichai, 2006), there have been very few English-language research studies on Musicology. Therefore, this research study will be conducted to search for high-frequency vocabulary items in music and music education research articles to facilitate music learners or those interested in music in searching for vocabulary and meanings that are specific to the context of music education research articles.

This study focuses on the most frequent content vocabulary items in music and music education research articles in international journals in the field of music and music education. These words are the major obstacles for learners who read English research articles in the music-related fields. The vocabulary usage is also a vital way to analyze certain rhetorical functions in academic texts by utilizing corpus-based methods in order to understand the content of those written texts.

In this study, 300 music and music education research articles retrieved from the online journal websites were used for data collection and analysis. This study aimed to generate frequently used content words occurring in music and music education research articles called the Musicology Word List (MWL) using a concordance program.

As this study aims to establish the Musicology Word List (MWL) as a list of core words used in music discipline, the MWL will contribute to the vocabulary learning and teaching in the field of music in many aspects. First, it provides a useful specific vocabulary list for EFL and ESL learners to focus on relevant and necessary vocabulary and comprehend texts in the music discipline more effectively. Second, ESP teachers, course designers, and material developers of music discipline can use the MWL as a guideline for preparing their teaching or materials.

Literature Review

Corpus Linguistics

A corpus is a collection of written or spoken texts used for linguistic analysis. A corpus usually refers to a principled collection of electronic texts stored on a computer. Computers can simply access the data. Corpus linguistics, as defined by Müller and Waibel (2000), is the study of naturally occurring language using specialized software tools on a computer. The study of specialized word lists is a key component of corpus linguistics, and it makes a substantial contribution to the area of ESP.

Concordance software, usually referred to as a concordancer, is a type of computer program that can help in corpus linguistics. It can build word frequency lists, word lists, and keyword lists automatically. In electronic text forms such as text files, databases, or the internet, a concordancer can quickly count and classify words. A concordancer can be useful in determining the frequency, distribution, or occurrence of specific words or combinations of words, as well as correlations between terms in authentically used language. The concordancers "AntWordProfiler" (Anthony, 2014) and "AntConc" (Anthony, 2020) were used to analyze the data in this investigation.

Word Frequency Lists

A word frequency list is a list usually sorted either in frequency order and alphabetical order of all words in a specified corpus together with the frequency of word occurrence in the corpus (Hunston, 2006). Word lists are useful in various linguistic studies, for example, language teaching, stylistics, content analysis, and information retrieval. It is also very useful for teaching and learning English for specific purposes (ESP). They provide frequency information of the vocabulary in particular texts. Word frequency lists can be generated from a particular corpus by using concordance software. The software will thoroughly search every word in that corpus to find out how many tokens and how many different word types there are. The software then generates the final counts as a frequency list, which can be displayed in rank order frequency and in alphabetical order (Evison, 2010). Word lists are crucial in lexical studies that, aim to gather information on the frequency with which words occur in specific places.

A specialized word list (also called technical word list and field-specific academic vocabulary list) refers to the list of academic words that are closely related to particular disciplines (Liu & Han, 2015). Experts have paid considerable attention to the word lists of specific disciplines because several studies have proved that not all words in the interdisciplinary academic word lists (e.g. West's GSL (1953), Coxhead's AWL (2000), and Gardner&Davies's AVL (2014)) are equally important to learners with highly specific needs. As can be seen in previous corpus linguistic works such as Wang et al's MAWL (2008), Tangpijaikul's Thai-EBEN (2014), and It-ngam and Phoocharoensil's SAWL (2019).

Word Families

One of the methods of word counting is word families. A head word, its inflected forms, and its derived forms make up a word family. To clarify, a head word, words made from the head word with affixes, and a morpheme that can be appended to a base form word to generate a new word or change the word's part of speech are all members of the word family, according to Nation (2001). All new words with the same core word but different affixes will belong to the same word family. To illustrate, the word *music* is in a base form. *Music* can be attached with many affixes to make new words such as *musicology, musicological, musician, and musicianship*. These words are members of the word family *music*.

Methodology

Data Collection Procedure

The sample to be employed in this study is music and music education research articles downloaded from ten international journals websites in the music subject category. One journal is a journal recommended by music experts in this research. And nine journals are in the highest quartile which is Q1, according to SCImago Journal Rankings (SJR).

In order to study the vocabulary frequently used in music and music education research articles, English academic papers in the music and music education category of international journal websites were collected in order to create a Musicology Word List (MWL). This corpus consists of 300 music and music education research articles from ten sub-corpora downloaded from ten journals. Selected articles need to have been published during the years 2015-2021 to avoid outdated words. Some sections of the articles, namely figures, tables, acknowledgment, references and appendices, were removed because these sections did not seem to contain core words specific to the discipline being investigated. All 300 music and music education research articles contain a balanced number of running words, approximately 5,300 to 8,000 running words each. Therefore, the approximate total size of this MWL corpus is 2,000,000 words**.**

Table 1 Corpus Compilation

Journals	Articles	Words
International Journal of Music Education	30	187,420
Journal of Music Teacher Education	30	158,837
Journal of Research in Music Education	30	184,219
Popular Music and Society	30	239,079
Psychology of Music	30	158,580
Journal of New Music Research	30	247,820
JSTOR	30	158,456
Research Studies in Music Education	30	210,534
Bulletin of the Council for Research in Music Education	30	204,021
Music Education Research	30	189,250
Totals	**300**	**1,938,216**

Data Processing

All the selected music education research articles are both in a pdf format and ones that can be copied immediately. The pdf format articles were then converted into text files (.txt.), in order to be processed by a computer software program, AntWordProfiler (Anthony, 2014) and AntConc (Anthony, 2020). As stated earlier, figures, tables, bibliographies, footnotes, acknowledgments, references and appendices in the articles were removed from all the articles.

Research Tools

To analyze the corpus, AntWordProfiler version 1.4.0 (Anthony, 2014) and AntConc 3.5.8 (Anthony, 2020) were the main programs used in this work.
AntWordProfiler and AntConc are freeware, multiplatform tools for carrying out corpus-based research on vocabulary profiling created by Laurence Anthony. It runs on any computer.
AntWordProfiler was used to generate word lists from the music and music education research articles, and to compare the lists against reference word lists: West's (1953) GSL and Coxhead's (2000) AWL.
AntConc is a freeware corpus analysis program for concordancing and text analysis. This program was used to examine the words in the Music Word List. The concordance function was used to identify the MWL words in the music and music education research articles. The

program was used to determine the context of words in a list generated by AntWordProfiler by looking at concordance lines.

Word Selection Criteria

In order to create the Musicology Word List which represents words frequently found in music and music research articles, this work adapted the word selection criteria in the compilation of AWL (Coxhead, 2000). According to the AWL, words are selected based on two criteria: frequency and specialized occurrence.
Specialized occurrence refers to the frequent occurrence of the words that are outside the first 2,000 most frequently occurring English words in GSL and AWL. In this work,

both GSL and AWL were considered essential for music learners, as they should know these words prior to learning specialized academic words. Therefore, for the MWL, the words occurring in the GSL and AWL were removed.
The frequency of a word in a corpus was the second condition. According to the AWL, each word in the list has to occur with a frequency of at least 100 times in the whole corpus of 3.5 million running words. That is equal to approximately 28.6 times in every one million running words of the corpus. In the present study, the corpus contained around 2 million running words. Therefore, the appropriate frequency rate for the MWL was 57 times in the whole corpus (28.6 x 2 = 57.2).

Table 2 Word Selection Criteria

Criteria	AWL	MWL
Corpus size	3.5 million tokens	2 million tokens
Special occurrence	Excluding West's GSL	Excluding Coxhead's AWL and West's GSL
Frequency	100 times on the whole corpus (28.6 times/ 1 million)	57 times on the whole corpus (28.6 x 2 = 57.2)

Data Analysis

Data Analysis

The corpus-based approach consisted of four major steps. First, the music and music education research articles were loaded into the AntWordProfiler program. An overall list of word families occurring was created. Second, the word families in the list were refined and compared with West's (1953) GSL and Coxhead's (2000) AWL. The word families coinciding with the GSL and AWL were removed. Next, the remaining words were further investigated. Words like transparent compounds, proper names, non-words, foreign words, prefixes, and abbreviations were removed from the results. Finally, the words that met all selection criteria were kept. The potential MWL were generated based on the results.

Expert Justifications

The Musicology Word List was validated by three music experts in order to verify the appropriateness and the practicality of the Musicology Word List. Opinions and recommendations were also included in the validation form. The results from the experts were calculated using Item Objective Congruence (IOC) method. A word was removed if its mean score was lower than 0.5.

The irrelevant terms were manually eliminated after the expert justifications procedure. The vocabulary in this word list was organized alphabetically and by frequency from the highest to lowest. As a result, students will be able to notice which words appear more frequently in the music and music education research articles if the most frequent words are shown first in the arranging process. The word list was then categorized to determine which word families the words belong to. Learners can access the word members that appear in the articles by viewing them along with their frequency. The complete version of the Musicology Word List (MWL) was finally obtained.

Results

Lexical Profiles

For word frequency analysis and corpus lexical profiles, the AntWordProfiler tool is essential. This program can classify texts by word frequency into the first and second thousand words of GSL, 570 words of AWL, and off-list words. This data reveals how difficult the texts are. Readers can find texts difficult to read if they contain a large number of academic or off-list words. However, it is an advantage that contributes to this work. This information confirms that there are other terms besides GSL and AWL that the learner or person interested in musical articles should know.

The Musicology Word Frequency List

After the analysis of the AntWordProfiler program, the results were shown. The vocabulary was examined; specific names and meaningless words were removed. The Musicology Word List (MWL) derived from a corpus of music and music education research articles before being validated by music experts contained 516 words.

The Music Experts' Opinions Regarding the Musicology Word List

After the researcher compiled this MWL, the validation form and word list were submitted to three music experts for approval. Using the Item Objective Congruence (IOC) process, the experts were asked to go through the words on the lists and determine whether or not each word should appear on the list. The number of terms that must be deleted was 378 words (72.87 percent). The remaining words in the final MWL are 138 words which is the complete list of the 94 word families of the Musicology Word List (MWL) and their word family members.

Table 3 The final Musicology Word List (MWL)

acoustic	174		chorus	138		
aesthetic	142		chord	1097		
	aesthetics	69			chords	445
articulation	164		choreomusical	80		
	articulated	79	clarinet	85		
audio	337		cluster	65		
auditory	191		composers	256		
aural	282				compositions	150
bass	325				compositional	90
blues	77		concert	505		
cadence	135				concerts	233
chamber	73		conductor	266		
chant	95				conductors	176
choir	417		conservatories	69		
	choirs	112	consonance	88		
choral	467				consonant	72

Term	Count	Variant	Count
diatonic	101		
dissonance	83		
drummer	127		
		drummers	103
ensemble	1201		
		ensembles	471
era	80		
excerpt	121		
		excerpts	153
expressivity	109		
flute	175		
folk	194		
funk	60		
genre	441		
		genres	438
groove	196		
guitar	298		
harmony	296		
		harmonic	531
harp	95		
improvisation	588		
		improvising	102
		improvise	73
instrumentation	72		
		instrumentalists	98
intonation	126		
jazz	686		
keyboard	142		
lyrics	291		
majors	226		
manuscripts	80		
melody	565		
		melodic	403
		melodies	293
metronome	63		
musics	93		
		musicianship	262
		musicality	76
		musicology	63
		musicological	62
notation	433		
octave	140		
opera	103		
orchestra	447		
		orchestral	229
		orchestras	78
pedagogy	436		
		pedagogical	440
		pedagogies	138
percussion	246		
performer	189		
phrase	262		
		phrases	193

Term	Count	Variant	Count
piano	899		
		pianists	193
		pianist	69
pitch	923		
		pitches	218
polyphonic	87		
pop	289		
practicing	214		
		practiced	119
pulse	97		
punk	96		
quantitative	190		
rap	206		
recordings	478		
reference	261		
rehearsal	437		
		rehearsals	147
renaissance	63		
repertoire	398		
rhythm	567		
		rhythmic	451
		rhythms	222
salsa	254		
saxophone	69		
score	585		
		scores	796
		scored	94
semitone	67		
solo	164		
sonata	58		
songwriting	110		
sonorities	86		
studio	267		
tango	298		
tempo	554		
timbre	469		
tonality	84		
		tonal	188
tone	618		
		tones	409
tonic	149		
transcription	88		
		transcribed	91
		transcriptions	91
triad	131		
		triads	67
vibrato	151		
violin	212		
vocal	594		
		vocals	126
woodwind	137		
workshop	141		
		workshops	135

Discussions

The analysis of the most common words in the music and music education article corpus yielded some interesting results. From Nation (2008)'s study, an academic text consisted of GSL at 80%, AWL at 10%, and specialized words at 10%. From the results of this study, AntWordProfiler program was used to (1) analyze the lexical profiles and (2) created an MWL from music and music education research article corpus. For the lexical profiles, overall this particular corpus consisted of GSL at 73.25%, AWL at 10.31%, and specialized words at 16.44%. This adds an interesting vocabulary point to the information composition while also illustrating the corpus's lexical diversity. Furthermore, the percentage of technical words was 16.44%, suggesting that music and music education research papers may be difficult to understand. Meanwhile, it also confirms that this corpus had a high enough percentage of specialized terms to allow for further development into a specialized linguistic corpus according to It-ngam and Phoocharoensil (2019). Readers who are unfamiliar with English musicology terminology or who are at a beginner level can struggle. For better reading comprehension, it is important to concentrate on the specialized words. It can be concluded that building a Musicology Word List (MWL) is beneficial for reading music and music education articles, as evidenced by the data. GSL and AWL should be familiar to readers already. If readers can gain more vocabulary from MWL, it may increase their comprehension of music and music education research articles by 90%.

With respect to building the Musicology Word List, after proper names and meaningless words were removed, there are remained a total of 516 words. Kruawong's (2018) approach was used as the framework to investigate the emergence of content words in music research articles and music studies. The study's results are consistent with previous research indicating that developing technical word lists should take into account more than just a corpus-based approach. Relying only on the corpus-based approach has some drawbacks. For example, important and necessary words cannot be separated from the word list. Hence, a process of justification by experts was included in this study. The three music experts were asked to perform the IOC of each word. Words on the MWL list must win two points or have a minimum average of 0.5 to be included. Thus, the total of 378 words (72.87%) were removed after the validation process. As a result, MWL contains 138 essential words.

However, conducting research on building a word list in different disciplines and gathering data from different sources leads the results to vary. Even though the music and music education research article corpus had a total of 1,938,216 running words, which is similar to the Zoology Research Academic Corpus (Kruawong, 2018) and the marketing corpus (Patanasorn and Kakaew, 2013), the number of words in the word list after the final step was not the same.

Conclusion

The objectives of this work were to (1) create a Musicology Word List and (2) investigate the justifications for the word list from music experts. The research was based on a corpus of music and music education research papers. The corpus of music and music education research papers was collected by the researcher from 300 articles published during 2015 – 2021, from nine journals within the music category on the SJR and one expert-recommended journal. The corpus contains a total of 1,938,216 running words. The Software AntWordProfiler was used to analyze data and determined word frequency. The Musicology Word List (MWL) is specific to the disciplines, besides GSL, AWL, and function words. After the expert removal process was completed, the MWL corpus contained 138 words.

There are several ways to benefit from the Musicology Word List (MWL). Teachers and learners who have access to this word list can use it to achieve their own objectives. According to Nation and Waring (1997), teachers and course designers should have reference word lists to decide which words in a certain register should be taught to students. The MWL was created to provide teachers, course designers, and material developers of music discipline with a ready-to-use word list for English for specific purposes (ESP) in musicology. This study's findings revealed that a teacher can use a corpus program to teach vocabulary to EFL and ESL students, especially in ESP classes. The instructor can instruct students to study this MWL to prepare before reading English music articles. The MWL created by the AntWordProfiler software can provide a list of the most common specialized vocabulary in particular subject areas, which helps students with reading comprehension; this allows teachers to know which vocabulary to concentrate on and also to prepare their teaching or materials. It includes a helpful vocabulary list for EFL and ESL students to help them focus on appropriate and essential vocabulary and better understand texts in music and music education research papers. Overall, this project is another example of how corpus linguistics and empirical evidence can be used to improve students' learning.

Limitations and Future Studies

This work has significant limitations that should be acknowledged. Firstly, the music and music education research article corpus was primarily collected from scholarly articles in the category of musicology published in only nine journals from the Scimago Journal Rank and one expert-recommended journal. Secondly, despite the fact that the corpus of music and music education research articles comprised approximately 1.9 million terms, the higher the word count, the more effective it is. A larger corpus can improve the chances of finding vocabulary that appears frequently in musicology research papers. Finally, ideally, the number of experts involved should exceed the scale of the scoring criteria in order to minimize the irreconcilable and may be able to acquire more essential vocabulary.

It is recommended that, for future research work, it is advisable to compile words from a wider variety of sources

such as music education books so that the vocabulary list can be used in a wider range of subjects and potentially be more practical. It is also important for EFL/ESL students to be aware of terms in musicology papers that have many definitions and meanings. In the future, in addition to building a larger corpus than the MWL's music and music education research article corpus, collocations and polysemy should be added to the research process to make the word list more useful to learners. For the justifications, it is also useful for future terminology research, more than three specialists should be involved in the validation process to reduce the conflicts. In addition to the judging by experts, future studies should include learners' judgments to confirm that the word list is practical, both for instructors and learners.

Acknowledgement

References

[1] Anthony, L. (2004). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. Proceedings of IWLeL, 7-13.

[2] Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

[3] Anthony, L. (2020). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

[4] Biber, D., Conrad, S., & Reppen, R. (1998). Corpus Linguistics: Investigating Language Structure and Use. New York: Cambridge University Press.

[5] Boonyos, K. (2014). Establishment of a microbiology academic word list: [Bangkok, Thailand]: Language Institute, Thammasat University, 2014.

[6] Browne, C. (2014). A new general service list: The better mousetrap we've been looking for. Vocabulary Learning and Instruction, 3(2), 1-10.

[7] Coxhead, A. (2000). A new academic word list. TESOL quarterly, 34(2), 213-238.

[8] Davies, Mark. (2016-) Corpus of News on the Web (NOW). Available online at https://www.english-corpora.org/now/.

[9] Dee Gardner, Mark Davies, A New Academic Vocabulary List, Applied Linguistics, Volume 35, Issue 3, July 2014, Pages 305–327, https://doi.org/10.1093/applin/amt015

[10] Gilner, Leah. (2011). A primer on the General Service List. 23. 65-83.

[11] It-ngam, T. & Phoocharoensil, S. (2019). The development of science academic list. Indonesian Journal of Applied Linguistics, 8, 657-667. doi:org/10.17509/ijal.v8i3.15269

[12] Kakaew, S., &Patanasorn, A. T., (2013). Wordlists of Marketing: A corpus-based Study. 9th International Conference on Humanities and Social Sciences 2013.

[13] Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. English for Specific Purposes, 39, 1-11. doi:10.1016/j.esp.2015.03.001

[14] Müller, F., & Waibel, B. (2000). Corpus linguistics - an introduction. Retrieved December 3, 2018 https://www.anglistik.uni-freiburg.de/seminar/abteilungen/sprachwissenschaft/ls_mair/corpus-linguistics

[15] Nation, P., and Waring, R. (1997). Vocabulary size, text coverage and word lists. Vocabulary: Description, acquisition and pedagogy, 14, 6-19.

[16] Nation, P. (2008). Teaching Vocabulary: Strategies and Techniques. Boston, Heinle.

[17] O'Keeffe, A., & McCarthy, M. (2010). Historical perspective: what are corpora and how have they evolved? Michael McCarthy and Anne O'Keeffe The Routledge handbook of corpus linguistics (pp. 31-41): Routledge.

[18] SCImago, (n.d.). SJR -SCImago Journal & Country Rank [Portal]. Retrieved Date you Retrieve, from http://www.scimagojr.com

[19] Tangpijaikul, M. (2014). Preparing business vocabulary for the ESP classroom. RELC journal, 45(1), 51-65.

[20] Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. Journal of English for Academic Purposes, 12, 248-263.doi:10.1016/j.jeap.2013.07.001

[21] Wang. (2017). A Corpus-based Study of English Vocabulary in Art Research Articles. Journal of Arts and Humanities, Vol 6, Iss 8, Pp 47-53 (2017) (8), 47. doi:10.18533/journal.v6i8.1255

[22] Wang, J., Liang, S.-l., & Ge, G.-c. (2008). Establishment of a Medical Academic Word List. English for Specific Purposes, 27, 442-458. doi:10.1016/j.esp.2008.05.003

[23] Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering

undergraduates. English for Specific Purposes, 28(3), 170-182.

[24] West, 1953, A General Service List of English Words, Longman, London

[25] Yang, M.-N. (2015). A nursing academic word list. English for Specific Purposes, 37, 27-38.