Research Article

Collocations in Methods Section of Masters of English Language Theses

Wilhelmina Coker Davis[*]

Ramos Asafo-Adjei

Francis Bukari

Department of Communications and Media Studies,
Takoradi Technical University Takoradi,
Ghana
*Corresponding Author

Abstract

Collocation has long been recognised as a problematic domain for Second language learners of English and acknowledged to lag far behind the acquisition of other Second language acquisition (SLA) aspects. This mixed approach study was, therefore, aimed to identify content words and their collocations in the methods section of the Master of Philosophy in English Language theses. Specifically, the research question was: What are the collocations of the content word in the Methods sections of English language MPhil theses at the University of Cape Coast, Ghana? The data was purposively sampled from 20 MPhil theses submitted to the Department of English and analysed using AntConc 3.5.8 software. The Firthian (1986) Collocation Approach was used for the study. The findings showed that the content word data is the keyword typical of the methods section, and it strongly collocates with both nouns and verbs to both the left and to the right in varied ways. Also, near-synonyms collocate differently with the keyword data. This study makes a vital contribution by revealing how the students in the Master of Philosophy in English programme make use of language to present the methods section of their theses. The study offers implications for engendering students' academic writing, especially at the postgraduate level.

Keywords: academic writing; corpus linguistics; data; methodology; postgraduate

1. Introduction

Vocabulary has a significant role in English learning, whether as a second language or as a foreign language, and is an important part of language learning because it allows people to communicate more effectively (Saud, 2018). During second language learning, students need to comprehend the vocabulary in that language. Students who obtain a large amount of vocabulary as well as knowledge of the different types of vocabulary tend to have better language acquisition levels and communication ability (Zhou, 2016).

Many linguists such as Lewis (2001), Shamsudin *et al.* (2013), Nasrin (2016), Sanguannam (2017) Li-Wei *et al.* (2021) and Bui (2021) all agree that collocation is significant for ESL students to use language accurately and fluently, or that collocation is the key to language fluency. Tanihardjo (2017)

describes collocation as a group of co-occurrences of two or more words that are always used together normally, both in spoken and written languages; such as take a chance, make a mistake or a difficult decision. Ashiba (2017) and Qader (2018) also indicate that collocation is the combination of two or more words to create new meaning.

Collocation is described using a range of terms, such as lexical chunks, lexical bundles, prefabricated patterns, fixed expressions or multi-word units, allowing native-like speech and writing to be produced (Basal, 2019). Collocational knowledge helps second language learners produce natural-sounding speech and writing in the target language (Bui, 2021). It is widely accepted that collocational knowledge is essential for fluency (Shitu, 2015; Boonraksa & Naisena, 2022; Saud, 2018). For this reason, collocations have become an increasingly important aspect of vocabulary knowledge in second language writing including master's thesis (Thongvitit & Thumawongsa, 2017). Sanguannam (2017) points out that, despite the value of collocations, most students' second language learners of English lack competence in using them in writing; many scholars have witnessed similar shortcomings and consider knowledge collocation as a useful guide for students especially masters students whose first language is not English (Bui, 2021; Qader,2018; Zhou, 2016).

In Ghana, there is no known study on the collocation of Master of Philosophy (henceforth, MPhil) theses, especially in English language, focusing on the methods chapters at the University of Cape Coast. Existing studies have largely focused on phraseological patterns (Ngula, 2018) and students' essays (Jones-Mensah *et al.*, 2022). Although these research endeavours have made significant contributions to corpus studies within the Ghanaian context, an empirically-based study that analyses the collocations that typifies the methods section of MPhil theses in English language in the Department of English at the University of Cape Coast (UCC) is lacking, and this study, therefore, aims to address this gap. Therefore, the research question that guided the study is: What are the collocations of the content word in the Methods sections of English language MPhil theses at the University of Cape Coast, Ghana?

The purpose for undertaking this study is hinged on two critical reasons. Firstly, knowledge of collocations helps to improve learners' language skills, enhance learners' communicative competence, and achieve native-like fluency. Hence, knowledge of collocations is essential for L2 learners. That is to say, the mastery of collocations is imperative in the process of acquiring a second language (Bui, 2021). Secondly, evidence revealed that the deficiency in collocations is a factor attributing to ESL writers' errors, which hinder the learners' fluency. In many cases, learners tend to transfer the English collocations to the equivalent ones in their mother tongue if they do not know exactly which words occur together in the context. Therefore, this study will serve as a guide as to how content words in methods section of theses should collocate.

2. Literature Review

2.1 Studies on collocations

Empirical studies, over the years, have evidenced the influence of corpus-based methods of analysis on students' academic works and the benefits of corpus tools in language learning (Ashouri *et al.*, 2014; Grigoriev & Sokolova, 2019; Lai, 2015; Lailiyah & Setiyaningsih, 2021; Taweel, 2020; Ucar & Yukselir, 2015; Zhao, 2023). One aspect of corpus linguistics studies is collocations. Collocations, as one of the units of formulaic language, have received considerable attention and have been defined by scholars using various perspectives (Akhter & Nordin, 2020; Gablasova *et al.*, 2017; Webb *et al.*, 2013; Xu, 2020).

Taweel (2020) compared the discourse markers employed in academic works of Arab ESL students to those of native speakers in a study on collocations. Taweel discovered a difference in the use of discourse markers between Arab pupils and native English speakers. That is, in the domains of conveying attitude, sequencing, cause and effect, addition and comparing and contrasting, ESL learners utilised more discourse markers than native English speakers. Lailiyah and Setiyaningsih

(2021) analysed the most frequently used academic words in an EFL classroom in Indonesia in a related study. Lailiyah and Setiyaningsih found that inappropriate words that did not contribute to making sense of the argumentative essays were used by the learners in their write-ups.

The goal of Phoocharoensil's corpus-based study of the verbs "persist" and "persevere" in the Corpus of Contemporary American English in 2021 was to clear up any confusion between them based on how they are used in different genres, with other words, and in terms of meaning. Phoocharoensil found that the use of synonyms and the selection of their collocants are dependent on the type of genre in which they occur. For example, while "persist" is common in academic texts and frequently co-occurs with words or phrases expressing negative meanings associated with ongoing unpleasant situations, "persevere" is preferred in webpages, denoting strong determination and great effort in completing a difficult task. Similarly, Huang *et al.* (2018) investigated the collocability and semantic prosody of verbs of publicity in China's English Media in comparison to those in American newspapers when creating a Corpus of China's English Media (CCEM). Huang *et al.* claim that differences in collocational associations and semantic prosody can reveal the attitudinal meanings of verbs in Chinese English media advertisements. Huang *et al.* report, for example, that journalists in China use specific collocational patterns to describe certain political, cultural or social concepts, and that collocational patterns with positive or neutral semantic prosody were more dominant in the CCEM than those with negative semantic prosody.

Molavi *et al.* (2014) compared the use of collocations in face-to-face and telephone conversation scripts collected from intermediate-advanced second language learners to collocations in three selected series of English textbooks and the Open American National Corpus (OANC). The findings of Molavi *et al.* suggest that specific types of lexical collocations, such as noun + verb and adjective + noun, were common among learners, despite the fact that no direct correlation was found between the frequency of collocations in their series of English textbooks and the learners' use of those collocations. In a different study, Sadeghi and Panahifar (2013) specifically focused on the use of different types of collocations in the oral productions of 30 intermediate Iranian EFL learners, identifying, categorising and accounting for the inappropriate collocations produced in their speeches. They identified the participants' collocational knowledge and use to be less than satisfactory. Sadeghi and Panahifar also observed an abundance of collocational errors in the oral production of Iranian EFL learners.

Chon and Shin (2013) analysed academic collocations in spoken and written discourse using the British Academic Spoken English and the Academic Corpus in a related study. Chon and Shin discovered that the field of economics has a concentration of common collocations, despite the fact that written collocations were 34% more prevalent than spoken collocations. Chon and Shin (2013) assert that the demand for L2 learners to understand and retrieve lexical items such as academic collocations is high, especially in the tertiary context. Similarly, Ucar and Yukselir (2015) used an experimental and a control group to examine the impact of corpus-based activities on verb-noun collocation learning in EFL classes. There was a statistically significant difference between the experimental and control groups, according to Ucar and Yukselir. That is, rather than the traditional method used to educate the control group, the corpus-based activities employed in teaching the experimental group had a substantial impact on their acquisition of verb-noun collocations.

Some research has focused on the usage of specific lexico-grammatical resources in students' academic writing. Yang (2018), for example, created a Learner Corpus (LC) of academic writing by Chinese learners and compared it to a Professional Corpus (PC) of published research articles. Yang discovered that, as compared to professional writers, Chinese learners utilise modal verbs more frequently, but they also overuse modals like can, will, could and would. However, less use of the modal verb 'may' was seen in the academic writing of Chinese learners. In another study, Grigoriev and Sokolova (2019) examined a corpus of research proposals submitted by Russian students to determine the discourse roles of first-person pronouns and what determines their use of first-person pronouns in academic writing. Grigoriev and Sokolova discovered that socio-cultural writing traditions had a significant influence on how students describe their perspective by using 'polite we'

rather than 'polite I'.

The foregoing discussion shows the fundamental role that collocations play, especially in students' academic writing. This means that collocations play significant roles in the successful production of any spoken or written text. This served as the pivotal factor for the use of corpus linguistics to explore the use of collocations in MPhil theses. Also, it is apparent from the foregoing discourse that studies on corpus linguistics lend credence to the efficiency of corpus linguistics tools in two crucial areas: (a) revealing the relevance of language learning and (b) demonstrating that studies into students' academic writing using corpus linguistic tools as analytical tools have become of great interest to researchers over the years. However, there are virtually few corpus linguistic analyses of the methods section of MPhil theses written by students of English. Therefore, it is the aim of the present study to analyse the MPhil theses – one of the most important types of student academic writings – using corpus tools. Specifically, we explore the key vocabulary (content word) typical of the methods section and words that collocate with that key vocabulary (content word). This is intended to provide useful information for genre-based pedagogy and could inform instruction in academic writing courses.

3. Theoretical Framework (The Firthian Approach)

This paper adopted the Firthian Approach (1986) to collocation as the theory for the study. Firthian defined collocation as a range of collocates parameterized with positional information and association strength. The application of Firthian approach is widely used in empirical studies in second language acquisition such (Xiao & McEnery, 2006, Herbst et al., 2014, Wolter & Gyllstad, 2013, Gablasova et al., 2017; Friginal & Hardy, 2021). These case studies demonstrate that the approach not only enables intuitive and holistic inference of syntactic and semantic functions of node words, but also supports large-scale quantitative investigation of lexical knowledge. The frequency-based studies of Firthian Approach to collocation is classified into Lexical Collocations and Grammatical Collocations. Lewis (2000) defined Lexical Collocations as two or more content words (Noun, Verb, Adjective, and Adverb) that can be divided into 6 sub-categories, while Grammatical Collocation is the combination of content words and function words (mostly preposition) which can be divided into three patterns as shown in Figure 1.
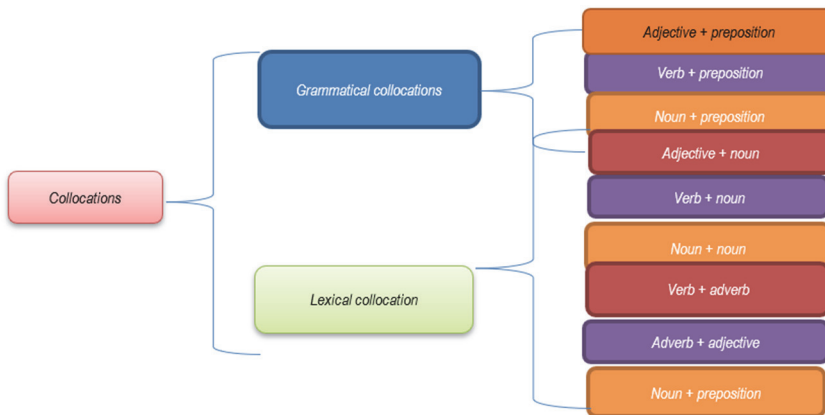

Figure 1. Collocation Classification

The meaning of Grammatical Collocation is the collocation between Noun, Verb or Adjective + Preposition, while Lexical Collocation is the collocation between Noun, Verb, or Adjective. As a

result, it is vital to look at the collocation in methods section of Masters theses of ESL students through the lenses of lexical collocation. The findings can be used to improve existing methods section part in students theses, to help students write effectively, and to develop a natural fluent use of language.

4. Methods and Materials

4.1 Context of the study

The University of Cape Coast (UCC), a comprehensive traditional university, was founded in 1962 in Cape Coast, Central Region, Ghana, to train graduate teachers to serve not only colleges which have the core mandate of training teachers but also technical institutions in Ghana. With research and teacher training as their focus, UCC offers courses ranging from certificate programmes to doctorate degrees. It operates on two campuses: the Southern Campus (Old Site) and the Northern Campus (New Site). It currently has a student population of over 70,000 undergraduate students and over 10,000 postgraduate students (https://yen.com.gh/112416-public-universities-ghana-2020-2021.html). Aside from being the official language of the university, the English language is also offered as a programme at various levels (i.e., proficiency, diploma, bachelor, master's and doctorate) by the Department of English. The Department of English is one of the departments in the Faculty of Arts, College of Humanities and Legal Studies (CHLS). At the Department of English and the University, generally, it is a requirement for every student to complete project work at the undergraduate level, a thesis at the master's level and a dissertation at the PhD level, which is crucial to postgraduate education.

Three rationales informed our choice of using UCC as the context. First, UCC is the only University in Ghana that teaches Academic Writing as a compulsory course for every postgraduate student. This course is intended to equip students, especially postgraduate students, with the requisite research skills and the use of the English language in writing their theses and other academic works. Second, the institution has been rated in the 2023 edition of the Times Higher Education ranking as the number one university in Ghana, the 4th best university in Africa, and the 24th best university in the world. Despite this giant stride, there is virtually no empirical study of this nature targeted at improving their academic writing using corpus linguistic approaches. Hence, this study provides concrete descriptions of how the students make use of language to present the methods section of their theses at UCC.

4.2 Data sampling

All UCC MPhil theses in the English language programme were targeted. Baker (2006) advises that, in dealing with texts on a specific topic, it would be sagacious to only gather texts that conform to these specialised criteria. Therefore, it was essential to build a specialised corpus for this study because none of the already established corpora were found to provide such corpora for the methods section of students' theses. A three-stage sampling technique was used to select 20 MPhil theses submitted to the Department of English.

The first stage involved the purposive selection of the Department of English MPhil theses because of the dearth of knowledge of students' application of corpus in these theses. The second stage involved the convenient selection of 20 MPhil theses, which were easily accessible and were downloaded from the institutional repository site of the UCC (https://ir.ucc.edu.gh/xmlui/handle/123456789/1035). Most of the theses in the repository were repeated; hence, we sampled one of each repeated thesis. The theses selected were submitted from 2019 to 2022 for currency. Ucar and Yukselir (2015) successfully experimented with fifteen (15) participants' academic writing, and Baker and Vessey (2018) successfully analysed nine (9) texts from the ISIS Magazine Dar al Islam, on the acquisition of collocations and discourse meanings,

respectively. We, therefore, expand the focus by using 20 theses Finally, we sampled the methods section of the theses. This is because the RM sections outline the sampling approach and criteria for including certain texts, which directly impact the content and representativeness of the thesis. Also, RM sections often contain technical language and terminology that may be worth specifically examining in this study.

4.3 Procedure and data collection

In building the specialised corpus, the research methods (henceforth, RM) sections of the electronic version of the collected theses saved in pdf files were copied from the theses and converted into Word documents (.doc). After this, we edited the Word documents to remove headings, sub-headings, page numbers, tables, figures and other insignificant repetitions to ensure smooth corpus processing and analysis. We later converted the .doc formats into plain texts (.TXT). In changing the texts from the Word document format to plain text, the Unicode UTF-84, which is compatible with AntConc, was used to ensure that as many as possible of the original characters remained intact after conversion to AntConc-readable form. Table 1 contains the data and their text length. In all, a specialised corpus of 64,973 tokens was built to achieve the purpose of the study.

Moreover, for ethical reasons, the names of the authors of the sampled theses were excluded, and the theses were rather labelled as RM 1, RM 2, to RM 20. By not reporting the names of the students who authored the theses, their rights to confidentiality and anonymity were guaranteed.

Table 1: Text Length of Research Method Section of Each Thesis

RM	Text Length	RM	Text Length	RM	Text Length	RM	Text Length
1	3085	6	4388	11	1250	16	2621
2	2216	7	3678	12	2394	17	5077
3	5126	8	3529	13	1662	18	4116
4	4972	9	2626	14	3081	19	3664
5	3405	10	2931	15	2935	20	2217

Baker (2006) encourages that in carrying out discourse analysis, it is potentially useful to use a reference corpus that reflects some aspect of the smaller corpus that is being studied, even if the reference corpus itself is not the main focus of the analysis. Therefore, the written component of ICE-Ghana, which has 67,965 tokens, was used for this study. The choice of ICE-Ghana as a reference corpus is a good benchmark since it is stored in the English language. Both corpora – RM sections of MPhil theses and ICE-Ghana – are naturally occurring data in the Ghanaian setting. Specifically, we used this reference corpus to identify the keyness of the lexical items that would be selected for the study.

4.4 Data analysis techniques

We used the AntConc 3.5.8 version as a corpus tool for the analysis of the corpus. Corpus tools are tools that inherently do both qualitative and quantitative analysis of the corpus which is uploaded onto it. Thus, the adoption of the corpus tool (AntConc) for the analysis of data in this study positions the present study in the mixed method paradigm. Since the present study is corpus-based, the specialised corpus was approached without any search terms in mind. Therefore, after uploading the corpus into the AntConc tool, we run the wordlist, and, from the wordlist, the first content word that occurred among the first ten (10) most frequently occurring words in the corpus, which is usually expected to be functional words, was selected for the study for further inquiry.

After the identification of the first functional word in the first ten words in the wordlist, there was then the need to find its keyness in the corpus as compared to a reference corpus, the ICE-

Ghana. After the identification of the keyness of the word, we did a collocation analysis through concordance. In the concordance analysis, the selected search term was input into the AntConc software, which brought out the use of the keyword in context displayed on the interface of the software. Finally, we sorted all the collocations from the concordance analysis manually and analysed in detail, with a focus on identifying the use of the keywords and their collocates in the context of the research methods section of the theses.

5. Results

The results of the study are presented in two subsections: the first section focuses on the key vocabulary (content word) typical of the RM while the section subsection deals with the collocations of the key vocabulary (content word) in the RM.

5.1 *Key vocabulary and its collocates in the Methods sections of English language MPhil theses*

5.2 *Wordlist Analysis*

A wordlist is identified based on how often the word occurs in the corpus in conjunction with how much it contributes to the content of the corpus (Baker, 2006). We generated the first ten most frequently occurring words from the RM corpus (see Table 2).

Table 2: First ten words of the Wordlist with their frequencies

Rank	Frequency	Word	Rank	Frequency	Word
1	3395	the	6	495	a
2	1555	of	7	433	is
3	1127	and	8	414	for
4	996	in	9	367	was
5	977	to	10	356	data

As can be seen in Table 2, *the* (F = 3395) was the most frequently occurring word in the RM corpus. This was closely followed by *of* (F = 1555), *and* (F = 1127), *in* (F = 996) and *to* (F = 977). The word *data* (F = 356) was found to be the most frequently occurring content word in the RM corpus. The fact that the word *data,* which is a content word, is ranked the tenth highest frequently occurring word among the first ten (10) most frequent functional words in the RM corpus is noteworthy.

5.3 *Keyness*

We conducted a keyword analysis to confirm or deny the keyness of *data*. The statistical test indicated the extent to which a particular word occurs unusually frequently in the RM corpus. The result of the keyword analysis is presented in Table 3.

Table 3: First ten words on the Keyword list

Rank	Keyword	Keyness	Effect	Rank	Keyword	Keyness	Effect
1	Data	+1481.29	0.0198	6	Research	+666.31	0.0115
2	Study	+1289.96	0.0187	7	Sampling	+522.39	0.0054
3	The	+1164.77	0.0999	8	Qualitative	+427.72	0.0048
4	Analysis	+829.95	0.0103	9	Were	+426.08	0.0152
5	Researcher	+773.09	0.0083	10	Theses	+397.07	0.0045

From Table 2, we identified *data* had the highest keyness of +1481.29 and the highest effect of 0.0198 in the RM corpus. The keyword analysis revealed *data* as essential and paramount when it comes to the research methods section of the MPhil theses.

5.4   Collocates of "Data"

The frequency of occurrence of collocates of a word in a corpus refers to the identification of collocates that are common across a wide variety of contexts and, many a time, are very meaningful (Mackiewicz & Thompson, 2016). This was the starting point for the identification of the collocates of the identified keyword (*data*). To identify the words that collocate with *data*, we set the corpus tool within a 5-word window to the left and right of the keyword to identify the frequency of co-occurrence of the words that collocate with *data* with the highest frequency. Only the first thirty (30) words with a frequency of ten (10) and above occurrences are presented in Table 4. We grouped the collocates of *data* under functional and content word categories. For each group, fifteen collocates were selected. Table 4 presents the results.

Table 4: First Thirty Collocates of Data

Functional Words	Frequency	Rank	Content Words	Frequency	Rank
The	507	1	Collection	74	1
Of	160	2	Study	60	2
In	135	3	Analysis	55	3
And	120	4	Data	30	4
To	144	5	Used	29	5
For	73	6	Procedure	23	6
Was	54	7	Research	21	7
This	48	8	Sampling	19	8
Were	44	9	Researcher	17	9
That	35	10	Collected	17	10
From	35	11	Quantitative	15	11
A	30	12	Gathered	12	12
Is	28	13	Collecting	12	13
Also	19	14	Technique	10	14
Are	17	15	Size	10	15

Table 4 indicates that a great number of functional words collocate with *data*. While *the* ranked as the highest (F = 507), *are* was ranked as the lowest (F = 17) as collocates of *data*. It can be observed that almost fifty per cent (50%) of the first fifteen (15) functional words which collocate with data have a frequency of above fifty (50) occurrences. Specifically, it can be seen that the frequencies of the first five (5) functional word collocates of data run beyond one hundred (100) occurrences to almost one hundred and fifty (150) occurrences in the RM corpus. This is a confirmation of Gabrielatos' (2007) observation that, in every corpus, there are more high frequency functional words than content words.

Regarding the content word category, on the contrary, though the frequencies of occurrence are below one hundred (100), the word *collection* (F = 74) ranked highest in the content word category, while the word *size* (F = 10) ranked the least as content word collocates of *data*. In addition, it was interesting to observe that the word data (F = 30) occurred as the fourth (4th) highest content word collocate of the keyword *data*. This confirms how essential the word *data* is in the RM corpus, as evidenced under Section 4.1.1, supporting the fact that the significance and keyness of the word *data* identified to have a keyness value of +1481.29 in the RM corpus are worth investigating (Mackiewicz & Thompson, 2016).

Furthermore, frequency alone cannot be taken as saliency. Hence, we used the Mutual

Information (MI) test was employed, MI value measures the strength of association between words, and a higher MI score of above 3.0 is a significant indication of a stronger association and coherent relationship between the search term and its collocates (Hong *et al.*, 2017).

MI statistical test to avoid the error of disregarding less frequent yet useful collocates (Walker, 2008). The probability of words co-occurring based on their relative frequencies and the size of the RM corpus was automatically computed. The results of the first fifteen collocates of *data* in descending order of collocation strength are presented in Table 5.

Table 5: Collocates of data using MI Value and Frequency

Rank	Total Frequency	Frequency to the Left	Frequency to the Right	MI Value	Collocates
1	74	13	61	6.82778	Collection
2	23	7	16	6.57929	Procedure
3	12	12	0	6.52521	Collecting
4	8	8	0	6.47077	Analysing
5	12	1	11	6.22565	Gathered
6	17	2	15	6.02771	Collected
7	7	6	1	5.98862	Source
8	8	4	4	5.83334	Encountered
9	9	9	0	5.56269	Analysing
10	10	7	3	5.43906	Size
11	7	6	1	5.36058	Spoken
12	10	7	3	5.05573	Methods
13	10	8	2	5.05573	Techniques
14	7	6	1	5.05573	Categories
15	6	3	3	4.90373	Reliabilities

It is evident from Table 5 that about seven lexical words that initially appeared in the first fifteen collocates of *data* did not appear in the list of collocates after using the MI statistical test. Rather, new lexical words such as *analysing, source, encountered, spoken* etc. were registered as collocates of *data* with MI scores of 5.0 and above, though they appear with frequencies of 10 and below. Since an MI score of 3.0 and above is capable of detecting strong word combinations, for a trade-off between frequency and saliency of collocates, in the present study, words with a frequency of above ten (10) occurrences yet falling within an MI score of 6.0 and above were considered strong collocates of the *data* and for concordance analysis.

Words identified as collocates include *collection, procedure, collecting, gathered* and *collected*. *Collection* occurs with a collocational strength of 6.8 and a high frequency of 74 occurrences, whereas *procedure* occurs with a collocational strength of 6.5 and a frequency of 23 occurrences. Again, though both *collecting* and *gathered* occur with a frequency of 12, the former has an MI value of 6.5 while the latter has an MI value of 6.2. More so, *collected* occurs with MI value of 6.0 and a frequency of 17. There is, consequently, an indication that *data* as a noun, collocates more with both verbs and nouns. Out of the five strongly identified collocates of *data*, though only two are nouns (procedure and collection) and the remaining three (collecting, gathered and collected) are verbs, the nouns appear at higher frequencies than the verbs.

5.5   *Concordances Analysis*

From Table 6, *data* collocates with *collection* and *procedure* more to the right, mostly as a pre-modifier, than distinctively to the left. First, when *collection* collocates with *data* to the right, *data* is used as a pre-modifier to *collection* to describe the type of activity that involves the assemblage of information to be subjected to detailed examination (concordance lines 3 and 4). However, when *collection* collocates with *data* to the left to serve the same purpose supra, the use of the preposition

ISSN 2039-2117 (online)
ISSN 2039-9340 (print)

Mediterranean Journal of Social Sciences
www.richtmann.org

Vol 15 No 6
November 2024

of is rather employed in between them (concordance lines 13, 14, and 15).

Table 6: Concordance of Data Collocating with Nouns

1 Courts thereby slowing down the process of	data collection.	Also, the delay in data collection
2 and qualitative methods which includes viewpoints,	data collection,	analysis and interference techniques use
3 assertion that this combination is necessary in	data collection	and analysis to yield better interpret
4 ability and ethics, challenges encountered in the	data collection	and analysis, and how such challenges
5 sampling technique, source of data, instrument of	data collection	and the data collection procedure. The
6 ive and quantitative technique through sampling,	data collection	and the presentation of statistical me
7 were discussed. The chapter also described the	data collection	approaches and sampling technique used
8 tween researcher and respondents for problem-free	data collection	at both research sites, a few
9 assistants were recruited to help in the	data collection	for the study. These are national
10 cannot be sure of their accuracy. The	data collection	for this study stared when a
11 as is usually the case with other	data collection	instruments such as the interview and
12 number available to represent a group for	data collection	It must be emphasized that there
13 measures to ensure the smooth	collection of data	in the main study. These measures are
14 research were evident in the	collection of data	and issues of ethics. A major challenge
15 in order to make the	collection of data	less stressful and well-organised, the
16 design, the sample size and sampling procedure,	data collection	procedure and methods employed in data
17 and its selection, the research design, the	data collection	procedure, ethics and the data analysis
18 data, though this raises ethical concerns. Its	data collection	procedure is highly reliable since it
19 high reliability. This means that under CDA,	data collection	procedure is highly reliable since it
20 technique and sample size, data source and	data collection	procedure, method of analysis, an
21 the research design, research site, data size,	data collection	procedure, sampling technique, analytic
22 The measures are discussed under data and	data collection	procedure. The data gathered from the
23 As noted under the section 'Data and	data collection	Procedure', the researcher provided th
24 data, instruments of data collection and the	data collection	procedure. The chapter also discusses
25 of materials and tools and then the	data collection	procedure. This research followed the
26 of the study. Also addressed are the	data collection	procedure, sampling method and the to
27 researcher with regard to the	data collection	procedure for one option was to visit the

Again, the noun *procedure* collocates with *data* to the right. There is the use of *collection* again in between *data* and its noun collocate (procedure). That is, *data* and *collection* are used together as pre-modifiers to *procedure* (concordance lines 16 to 27). On the contrary, though *procedure* seems not to be a close collocate of *data* to the left in the RM corpus, it is indicated from the concordance that whenever it collocates with *data* to the left, the preposition '*for*' is adopted in between them. This finding confirms Hong *et al.*'s (2017) observation that, as regards noun-noun collocations, many seem to be specialised combinations dependent on specific contexts. Regarding the concordance analysis verbs, the results are presented in Table 7.

Table 7: Concordance of Data with Verbs

1 in the study concerns efforts at	collecting data	from the two radio panel discussions, namely
2 Accra for assistance. The reason for	collecting data	from the three courts was not for
3 strategy is sometimes appropriate in	collecting data	in descriptive studies. On the strength of
4 I encountered a problem in	collecting the data	because the librarians of the various departments
5 questions (Creswell, 2003). In	collecting the data	for the study, I downloaded the forty-
6 and sampling technique used in	collecting the data	The data were gathered between September, 2010 a
7 procedures that were used in	collecting the data	have also been catered for. In addition
8 of the methodology employed in	collecting the data	as well as the transcription used in
9 Academic year. Convenient sampling is used when	data gathered	from a specific group is desired
10 under data and data collection procedure. The	data gathered	from the pilot study was analyzed
11 of these resources, I first coded the	data gathered	I categorised the theses into their
12 it. The sampling technique and how the	data was gathered	And used for analysis were
13 assigning tags, names or labeling pieces of	data collected	for analysis. According to Punch, the
14 deals with how I interacted with the	data collected	in order to come out with
15 Peace FM respectively (For a detail of	data collected	on the respective stations, please refe

16 (Chapter four) contains the analysis of the	data collected	to show how educated Ghanaians signal
17 research question two. To achieve this, the	data collected	were analysed using simple percentages
18 frequent form of passive voice utilized. The	data collected	were coded for analysis. According to
19 the supra segmental features of prominence. The	data collected	were made up of parliamentary discourse
20 about the issue of standards. After the	data was collected	it was edited, coded and
21 electronic equipment for onward analysis. The	data were collected	between December, 2009 and February
22 second. This chapter has explained how the	data were collected	for the study and how
23 nomic activities in the country (Hasty, 2005).	data were collected	from Sam Jonah library (the
24 the three categories of courts from which	data were collected	The search did not yield

The concordance analysis of how the three verbs *collecting*, *gathered* and *collected* collocate with *data* indicates that, though these lexical items are near-synonymous and belong to the same word class with the same semantic quality, they are found to collocate with *data* in a variety of ways. First, from Table 7, *collecting* collocates with *data* only to the left and not on the right, but *gathered* collocates with *data* to the right (11) more than to the left (1); like *gathered*, *collected* also collocates with *data* to the right (15) more than to the left (2).

Additionally, *collecting* collocates with *data* to the right as *data* becomes an object to *collecting*, but *gathered* collocates with *data* to the right as a post modifier to *data* (concordance lines 9-12). Similarly, *collected* collocates with *data* to the right as a post modifier (concordance lines 13-19), but when the auxiliary verbs *was/were* are employed in between them (concordance lines 20-24), *data* collocates as a subject to *collected*. These findings of near-synonyms collocating differently with *data* corroborate the results of Phoocharoensil (2021), who also found that, although the near-synonyms, *persist* and *persevere* are similar in cognitive meaning; as verbs, they are mostly associated with different noun collocates found in different semantic fields.

6. Discussion

In this study, we sought to explore the keyword and its collocates, and the relationship between them in the methods section of MPhil theses in the English language programme. We built a specialised learner corpus from the RM sections of the theses. The study revealed that the lexical item *data* is the keyword in the RM corpus and with an MI value of 6.0 and a frequency of above ten (10) occurrences, five strong collocates of the keyword *data* were identified for concordancing. The selected collocates include *collection*, *procedure*, *collecting*, *gathered* and *collected*. The results of the study have three outcomes on a noun-noun collocation and noun-verb collocation in academic writing such as the corpus of the RM sections of the theses.

First, the analysis revealed that *data* is a keyword in the research method corpus with the highest keyness value of +1481.29 and a frequency of 356 occurrences, appearing as the 10th highest occurring word in the corpus. This is an indication of the aboutness of the RM section in the MPhil in English theses, which confirms that the RM chapter is mainly about the data that is gathered and processed as proof of facts for a particular study. It confirms that it is the substance around which a study revolves because it forms the basis of that study (Abdulai & Owusu-Ansah, 2014; Bryman, 2012).

Second, *data,* as a noun, collocates with verbs more than with nouns; *collection*, *procedure*, *collecting*, *gathered* and *collected* were identified as strong collocates of *data*. This finding corroborates that of Holtz (2007), who found from her study of verb/noun collocation in interdisciplinary registers that nouns collocate with a greater variety of lexical verbs in their source domain than in the other domains.

Third, in collocating with nouns, *data* occurs at the left as a pre-modifier to nouns such as *collection* and *procedure,* but in collocating with nouns to the right, the prepositions *of* and *for* are adopted in between them. It was also identified in the corpus that there was no concordance hit for *data* collocating closely with *procedure* to the left. Whenever it did, *data* rather combined with *collection* to pre-modify *procedure*. *Data* collocating with *collection* gives the sense that such activity of assembling information is for a specific purpose, which is subjecting it to examination or

investigation for a study. In addition to that, *data* is used to collocate with *collection* to describe the mechanisms adopted in assembling the information (e.g., concordance line 11). *Data* collocating with nouns as premodifiers in the research method corpus is in tandem with the findings of Hong *et al.* (2017), who identified *data* as a noun modifier to *collection* in their collocational analysis of pre-modification types in academic writing like a thesis.

In addition, *data,* as a keyword in the methods section of MPhil in English theses, collocates with verbs such as *collecting*, *gathered* and *collected* in varied ways. In one way, they occur as adjectivals and pre-modifiers to *data* on the left, and, in other instances, *data* also collocates with these verbs either as subject or object to the verbs. *Data,* as used with these collocates in this manner, presents the idea that the assemblage or retrieval of the information for a study is not done haphazardly, but occurs systematically or sequentially, being referred to as procedure (concordance lines 24, 25, 26, 27). *Data* collocating with verbs is found to agree with the discovery of Dunne *et al.* (2005) that the two verbs most often used in collocation with *data* are *gather* and *collect,* suggesting that *data* is like wild fruits or flowers which the researcher finds in the field and thus have an existence independent of the researcher. In addition, the finding of near-synonyms collocating differently with the keyword *data* in the context of the present study is a confirmation of the finding of Phoocharoensil (2021) that near-synonyms, though similar in cognitive meaning, mostly associate with different noun collocates in different semantic fields.

As the findings of this study demonstrate, English language learners will be equipped to choose the right word for each specific context, as this study has exposed the subtle differences in the collocational behaviour between the near-synonyms that collocate with the keyword in the RM sections. This attests to Phoocharoensil's (2021) conclusion that such revelations will increase accuracy and proficiency in the use of English in the L2 setting. Furthermore, the current study demonstrates that all disciplines shape words for their own purposes (Hyland & Tse, 2009) and that taking a discipline-specific approach to language study, particularly collocations in academic writing, is critical and paramount, particularly in the EFL classroom (Taweel, 2020).

7. Conclusions

It was the aim of this study to explore the methods section of MPhil in English Language theses, identifying the keywords typical of that section and the words that strongly collocate with it. The corpus-based analysis of the specialised corpus revealed four key findings. First, the lexical item, *data*, was identified as the keyword in the methods section of MPhil theses, with the highest frequency of 356 and the highest keyness value of +1481.29. Second, five lexical items were identified as strong collocates of *data* in the methods section of the MPhil theses. These collocates were the nouns: *collection* and *procedure*, and the verbs: *collecting*, *gathered* and *collected*, with strong MI values of 6.0 and above. Third, in collocating with these verbs, in one instance, the verbs occur as adjectives to *data* on the left, and, in other instances, *data* also collocates with these verbs either as subjects or objects to the verbs. Finally, in collocating with these nouns to the left, *data* becomes a pre-modifier, but in collocating with nouns to the right, the prepositions *of* and *for* are adopted in between them.

In terms of what this study means, the results is a contribution to the semantic analysis of lexical items and show the semantic connections between words in the methods section of postgraduate theses. This informs faculty and postgraduate students alike on some of the nuances concerning collocation to better inform how they compose the methods sections of their theses. Again, the findings of this study contributes to andragogy, thus the findings of this study brings to the attention of postgraduate students the prominence of the methods chapter of theses and its lexico-grammatical features and inform faculty in guiding postgraduate students in their theses production. This supports students in successfully producing the expected discourse in their discipline. Even so, considering the nature of postgraduate education in Ghana and elsewhere, which is considered a personal educational and training journey for students, globally, the present findings contributes to the development of students' knowledge and awareness of keywords and their collocates not only in

the methods section of the thesis but also in the other sub-genres of the thesis.

For further research, it is suggested that studies of such nature on lexicography could be extended to the other sub-units of theses to achieve the similar purpose of helping novice postgraduate students to write their theses successfully with ease and distinction. The findings of the present study are limited to one educational institution and one department; other studies could be extended to include data from an entire faculty, data to compare faculties, or data from different universities to paint a more holistic picture of the situation, as pertains in Ghana. Also, a comparative analysis could be conducted using different sections of master's theses to extend the findings of the present study.

References

Abdulai, R. T. & Owusu-Ansah, A. (2014). Essential ingredients of a good research proposal for undergraduate and postgraduate students in the Social Sciences. *Sage Open*, 1-15. https://doi.org/10.1177/2158244014548178

Akhter, S. & Nordin, N. R. M. (2020). An analysis of EFL collocations: A corpus-based study. *International Journal of Management (IJM)*, *11*(8), 274-284. https://doi.org/10.34218/IJM.11.8.2020.027

Ashouri, S., Arjmandi, M. & Rahimi, R. (2014). The impact of corpus-based collocation instruction on Iranian EFL learners' collocation learning. *Universal Journal of Educational Research*, *2*(6), 470-479. https://doi.org/10.13189/ujer.2014.020604

Baker, P. & Vessey, R. (2018). A corpus-driven comparison of English and French Islamist extremist texts. *International Journal of Corpus Linguistics, 23*(3), 255–278. https://doi.org/10.1075/ijcl.17108.bak

Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.

Barnbrook, G., Mason, O., & Krishnamurthy, R. (2013). *Collocation: Applications and implications*. New York: Springer.

Bonsu, E. M., Afful, J. B. A., & Hu, G. (2023). A corpus-based genre analysis of letters of regularization: The case of land institutions in Ghana. *Ibérica*, (45), 215-241. https://doi.org/10.17398/2340-2784.45.215

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford: Oxford University Press.

Chon, Y. V. & Shin, D. (2013). A corpus-driven analysis of spoken and written academic collocations. *Multimedia-Assisted Language Learning*, *16*(3), 11-38. http://dx.doi.org/10.15702/mall.2013.16.3.11

Dunne, M., Pryor, J., & Yates, P. (2005). *Becoming a researcher: A companion to the research process*. UK: Open University Press.

Farrokh, P. (2012). Raising awareness of collocation in ESL/EFL classrooms. *Journal of Studies in Education, 2*(3), 55-74. http://dx.doi.org/10.5296/jse.v2i3.1615

Gablasova, D., Vaclav, B., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing and interpreting the evidence. *Journal of Research in Language Learning*, *67*(1), 155–179. https://doi.org/10.1111/lang.12225

Gabrielatos, C. (2007). Selecting query terms to build a specialised corpus from a restricted-access database. *ICAME Journal, 31*, 5-43. https://eprints.lancs.ac.uk/id/eprint/528

Grigoriev, I., & Sokolova, A. (2019). Corpus based analysis of first-person pronouns in research proposals written by Russian students. *The Journal of Teaching English for Specific and Academic Purposes*, *7*(4), 423–430. https://doi.org/10.22190/JTESAP1904423G

Hardie, A. (2012). CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics, 17*(3), 380-409. https://doi.org/10.1075/ijcl.17.3.04har

Holtz, M. (2007). A corpus-based analysis of verb/noun collocations in interdisciplinary registers. Proceedings of the Corpus Linguistics Conference. http://ucrel.lancs.ac.uk/publications/CL2007/paper/14_paper.pdf

Hong, A. L., Hua, T. K. & Mengyu, H. (2017). A corpus-based collocational analysis of noun premodification types in academic writing. *The Southeast Asian Journal of English Language Studies*, *23*(1), 115-131. http://doi.org/10.17576/3L-2017-2301-09

Huang, Q., Xia, L. & Xia, Y. (2018). A corpus-based study on collocation and semantic prosody in China's English media: The case of verbs of *Publicity*. In C. Liu, L. Zou, & J. Li (Eds.), *DASFAA 2018: Database systems for advanced applications* (pp. 203–217). QLD: Springer.

Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes, 24*(2), 123-139. https://doi.org/10.1016/j.esp.2004.02.002

Hyland, K., & Tse, P. (2009). Academic lexis and disciplinary practice: Corpus evidence for specificity. *International Journal of English Studies, 9*(2), 111-129. https://revistas.um.es/ijes/article/view/90781

Hyland, K., Handford, M., & Chau, M. H. (2012). Corpus applications in applied linguistics. London: Bloomsbury Publishing.

Jones-Mensah, I., Owusu Tabiri, M., Arkoh Fenyi, D., Edward Kongo, A. & Amexo, D. (2022). Vocabulary knowledge of collocation in business texts: A case of ESL tertiary students. *International Journal of Education, Technology and Science*, 2(1), 1-23.

Kim, Y. (2009). Korean lexical bundles in conversation and academic texts. *Corpora*, 4(2), 135-165. https://doi.org/1 0.3366/E1749503209000288

Lai, S. L. (2015). EFL students' perceptions of corpus-tools as writing references. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), Critical CALL – *Proceedings of the 2015 EUROCALL Conference*. Padova, Italy (pp. 336-341), Dublin.

Lailiyah, M. & Setiyaningsih, L. A. (2021). Academic vocabulary of EFL students' writing: A corpus-based study. http://ojs.pnb.ac.id/index.php/procedings/

Mackiewicz, J. & Thompson, I. (2016). Adding quantitative corpus-driven analysis to qualitative discourse analysis: Determining the aboutness of writing center talk. *The Writing Center Journal*, 35(3), 187-225. https://www. jstor.org/stable/43965694

McEnery, T., Brezina, V., Gablasova, D., & Banerjee, J. (2019). Corpus Linguistics, Learner Corpora, and SLA: Employing technology to analyze language use. *Annual Review of Applied Linguistics*, 39, 74-92. https://doi.org/10.1017/S0267190519000096

Molavi, A., Koosha, M., & Hosseini, H. (2014). A comparative corpus–based analysis of lexical collocations used in EFL textbooks. *Latin American Journal of Content and Language Integrated Learning*, 7(1), 66-81. https://d oi.org/10.5294/4087

Musa, A. (2014). Hedging in academic writing: A pragmatic analysis of English and Chemistry masters' theses in a Ghanaian university. *English for Specific Purposes World*, 42(15), 1-26.

Ngula, R. S. (2014). Corpus Linguistics and language development in Ghana. *Asian Journal of Humanities and Social Studies*, 2(4), 522-534.

Ngula, R. S. (2017). Epistemic modal verbs in research articles written by Ghanaian and international scholars: A corpus-based study of three disciplines. *Brno Studies in English*, 43(2), 7-27. https://doi.org/10.5817/BSE2017-2-1

Ngula, R. S. (2018). A corpus-based study of the phraseological pattern It+V-link+ADJ+That Clause in L1 and L2 expert academic writing. *International Journal of Applied Linguistics and English Literature*, 7(5), 128-134. http://dx.doi.org/10.7575/aiac.ijalel.v.7n.5p.127

Nkansah, S. K., & Bonsu, E. M. (2022). Foregrounding the verbal process: A corpus stylistic analysis of Adichie's Zikora. *Linguistics Initiative*, 2(2), 155-172. https://doi.org/10.53696/27753719.2250

Owusu, E., Agor, J., & Amuzu, E. K. (2015). Second language learners' family background and their English writing competence: The case of a private tertiary institution in Ghana. *Studies in English Language Teaching*, 3(4), 466-486.

Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms persist and persevere. *Eurasian Journal of Applied Linguistics*, 7(1), 240–258. https://doi.org/10.32601/ejal.911269

Sadeghi, K. & Panahifar, F. (2013). A corpus-based analysis of collocational errors in the Iranian EFL learners. *The Journal of Teaching Language Skills*, 4(4), 53-78.

Taweel, A. Q. (2020). Discourse markers in the academic writing of Arab students of English: A corpus-based approach. *Theory and Practice in Language Studies*, 10(5), 569-575. http://dx.doi.org/10.17507/tpls.1005.10

Uçar, S. & Yükselir, C. (2015). The effect of corpus-based activities on verb-noun collocations in EFL classes. *The Turkish Online Journal of Educational Technology*, 14(2), 195-205.

Walker, C. P. (2008). *A corpus-based study of the linguistic features and processes which influence the way collocations are formed: Some implications for the learning of collocations*. Unpublished doctoral thesis, University of Birmingham.

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91- 120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

Xu, D., Akhter, S., & Qureshi, A. H. (2020). Towards the importance of English collocations for L2 Learners, A corpus-based study. *Revista Argentina de Clínica Psicológica*, XXIX (3), 185- 191. https://doi.org/10.24205/03 276716.2020.712

Yang, B. (2016). A Corpus-based comparative study of Learn and Acquire. *English Language Teaching*, 9(1), 209-220. http://dx.doi.org/10.5539/elt.v9n1p209

Yang, X. (2018). A corpus-based study of modal verbs in Chinese learners' academic writing. *English Language Teaching*, 11(2), 122-130. http://doi.org/10.5539/elt.v11n2p122

Zhao, N. (2023). A corpus-based comparison study of first-person pronoun we in English-language abstracts. *Journal of English for Academic Purposes*, 63, 101244. https://doi.org/10.1016/j.jeap.2023.101244