

**A CORPUS-BASED STUDY ON FREQUENCY AND PURPOSE
SENTENCES IN LABORATORY ANIMAL REVIEW ARTICLES
USING WORDS FROM WORD LISTS**

SIRAWICH TAMPANICH

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF ARTS
(APPLIED LINGUISTICS)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2016**

COPYRIGHT OF MAHIDOL UNIVERSITY

Thesis
entitled

**A CORPUS-BASED STUDY ON FREQUENCY AND PURPOSE
SENTENCES IN LABORATORY ANIMAL REVIEW ARTICLES
USING WORDS FROM WORD LISTS**

.....
Mr. Sirawich Tampanich
Candidate

.....
Assoc. Prof. Songsri Soranastaporn,
Ph.D. (Educational Administration and
Foundation)
Major advisor

.....
Lect. Natthapong Chanyoo,
Ph.D. (Instruction and Learning)
Co-advisor

.....
Lect. Yuwadee Tirataradol,
Ph.D. (Curriculum and Instruction)
Co-advisor

.....
Prof. Patcharee Lertrit,
M.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University

.....
Asst. Prof. Wiwat Puntai,
Ph.D. (Information Studies)
Program Director
Master of Arts Program in
Applied Linguistics
Faculty of Liberal Arts
Mahidol University

Thesis
entitled
**A CORPUS-BASED STUDY ON FREQUENCY AND PURPOSE
SENTENCES IN LABORATORY ANIMAL REVIEW ARTICLES
USING WORDS FROM WORD LISTS**

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Arts (Applied Linguistics)

on
May 30, 2016



.....
Mr. Sirawich Tampanich
Candidate



.....
Lect. Sompratana Ratanakul,
Ph.D. (English Language and Applied
Linguistics)
Chair



.....
Assoc. Prof. Songsri Soranastaporn,
Ph.D. (Educational Administration and
Foundation)
Member



.....
Lect. Yuwadee Tirataradol,
Ph.D. (Curriculum and Instruction)
Member



.....
Lect. Natthapong Chanyoo,
Ph.D. (Instruction and Learning)
Member



.....
Prof. Patcharee Lertrit,
M.D. Ph.D. (Biochemistry)
Dean
Faculty of Graduate Studies
Mahidol University



.....
Asst. Prof. Aphilak Kasempholkoon,
Ph.D. Program in Thai (Literature)
Dean
Faculty of Liberal Arts
Mahidol University

ACKNOWLEDGEMENTS

I would like to acknowledge my sincerest gratitude and greatest gratitude to many people for the indispensable assistance, support, and encouragement without which I would not have been able to complete this thesis.

To start with, I would like to offer the most sincere thanks to my supervisor, Associate Professor Dr. Songsri Soranastaporn, for her patient guidance to me throughout my academic program as well as my thesis and for her beneficial comments and invaluable suggestions during the all my work on this thesis.

I wish to thank Dr. Natthapong Chanyoo and Dr. Yuwadee Tirataradol, co-advisors, for their generous assistance in terms of comments, guidance, and suggestions. Special thanks are also extended to Mr. William Martin for his professional advice, kind suggestions for my thesis, and editing my English grammar.

I would also like to give my greatest appreciation to Assistant Professor Kanchana Rojanapanich, my former advisor at Kasetsart University and Lect. Paweenai Boonpok, Kasetsart University. They provided me with a lot of friendly encouragement and supported me in my studies in this program and in doing the current research.

I wish to thank my friends in Applied Linguistics, Mahidol University and my friends from secondary school, Buranakrunluk School and Kasetsart University, who have listened, given superior advice, read many drafts, and supported me throughout my attempts.

More importantly, I wish to extend extra special thanks to my family for their support and understanding. Also I wish to thank them for being there for me whenever I felt discouraged in my studies. Without their love, support, patience, and encouragement, this thesis could not have been completed.

Sirawich Tampanich

A CORPUS-BASED STUDY ON FREQUENCY AND PURPOSE SENTENCES IN LABORATORY ANIMAL REVIEW ARTICLES USING WORDS FROM WORD LISTS

SIRAWICH TAMPANICH 5736188 LAAL/M

M.A. (APPLIED LINGUISTICS)

THESIS ADVISORY COMMITTEE: SONGSRI SORANASTAPORN, Ph.D.,
NATTHAPONG CHANYOO, Ph.D., YUWADEE TIRATARADOL, Ph.D.

ABSTRACT

The current study aimed 1) to identify top 100-high frequency of content words in a corpus of laboratory animal journals, 2) to investigate an academic word list (AWL) occurring in the same corpus, and 3) to investigate the frequency of use of words and forms in the statement of research purposes which laboratory animal researchers use in review articles. In the analytical framework, three criteria of Coxhead (2000) including *specialized occurrence*, *range*, and *frequency* as well as the limitation of a word family were used for choosing the word lists. Moreover, the purposive patterns applied the framework of Swales and Feak (2012) and Soranastaporn (2013).

The size of corpus consisted of 160 review articles during the year 2010 to 2014 from the Journal of the Institute for Laboratory Animal Research (ILAR). All articles included 840,773 running words excluding their bibliography, references, appendix, footnotes, captions, and acknowledgements. The research instruments used for analyzing the data in this study were AntConc Version 3.4.3 and The RANGE Program. Moreover, the purpose sentences were collected from all 160 review articles. Words and patterns presenting research purposes of each sentence were investigated. Moreover, the verb tenses of purpose sentences were examined. The statistical device used in the data analysis was descriptive statistics: the frequency data and percentages.

The main findings exposed that 1) there were 703,296 high-frequency words (83.65%), which were composed of 365,934 function words (43.52%) and 337,362 content words (40.13%). 2) the number of academic words in the laboratory animal review article (LARA) corpus was 36,658 word tokens (4.36%). 3) This corpus contained 133 purpose sentences, which were formed into eight purpose patterns. The pattern with a personal subject was found the most. All purpose sentences were also composed of 68 content words occurring in three word lists. Moreover, each pattern was composed of four different verb tenses; present simple, future simple, past simple and present perfect tenses. The present simple tense occurred in the pattern III the most, and the present perfect tense also occurred in the pattern III.

KEY WORDS: COUPUS-BASED STUDY / HIGH FREQUENCY WORDS /
ACADEMIC WORD LIST / PURPOSE FUNCTION / REVIEW
ARTICLE / LABORATORY ANIMAL

205 pages

การศึกษาความถี่และประโยคแสดงจุดประสงค์ในคลังข้อมูลภาษาบทความปริทัศน์สาขาสัตวทดลองโดยใช้คำศัพท์จากรายการคำศัพท์

A CORPUS-BASED STUDY ON FREQUENCY AND PURPOSE SENTENCES IN LABORATORY ANIMAL REVIEW ARTICLES USING WORDS FROM WORD LISTS

สิริวิษณุ ธรรมพานิช 5736188LAAL / M

ศศ.ม. (ภาษาศาสตร์ประยุกต์)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์ : ทรงศรี สรณสถาพร, Ph.D., ณัฐรุพงษ์ จันทร์อยู่, Ph.D.,
ยุวดี ธีรชราดล, Ph.D.

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อ 1) จำแนกคำเนื้อหาที่มีความถี่สูงสุด 100 คำแรกในคลังข้อมูลภาษาสาขาสัตวทดลอง 2) ศึกษารายการคำศัพท์ทางวิชาการ (academic word list: AWL) และ 3) ศึกษาความถี่ของคำและรูปแบบประโยคแสดงวัตถุประสงค์ในคลังข้อมูลภาษาสาขาสัตวทดลอง นอกจากนี้การวิเคราะห์รายการคำศัพท์ในคลังข้อมูลภาษาสาขาสัตวทดลองใช้กรอบทฤษฎีของ Coxhead (2000) ซึ่งประกอบด้วย *specialized occurrence*, *range*, และ *frequency* รวมถึงข้อจำกัดของตระกูลคำศัพท์ และการศึกษารูปแบบประโยคแสดงวัตถุประสงค์ใช้กรอบทฤษฎีของ Swales and Feak (2012) และ Soranastaporn (2013)

คลังข้อมูลภาษาสาขาสัตวทดลองประกอบด้วยบทความปริทัศน์จากวารสาร The Institute for Laboratory Animal Research (ILAR) จำนวน 160 บทความซึ่งตีพิมพ์ตั้งแต่ปี ค.ศ. 2010 ถึง 2014 คำในบทความปริทัศน์มีจำนวนทั้งหมด 840,773 คำ ซึ่งไม่รวมคำในบรรณานุกรม เอกสารอ้างอิง ภาคผนวก เชิงอรรถ คำบรรยายใต้ภาพ และกิตติกรรมประกาศ เครื่องมือที่ใช้ในการวิเคราะห์ประกอบด้วย โปรแกรมวิเคราะห์ภาษา AntConc Version 3.4.4 และโปรแกรม RANGE นอกจากนี้ นักวิจัยได้รวบรวมประโยคแสดงวัตถุประสงค์จากบทความปริทัศน์ทั้งหมด 160 บทความ พร้อมศึกษาคำและรูปแบบที่ใช้แสดงวัตถุประสงค์ รวมทั้งกิริยาแสดงกาลในแต่ละประโยคและในการวิเคราะห์ข้อมูลครั้งนี้ได้ใช้ค่าความถี่และร้อยละ

ผลการศึกษาพบว่า 1) คำที่มีความถี่สูงมีจำนวน 703,296 คำ (ร้อยละ 83.65) ซึ่งประกอบด้วยคำหน้า 365,934 คำ (ร้อยละ 43.52) และคำเนื้อหา 337,362 คำ (ร้อยละ 40.13) 2) คำศัพท์ทางวิชาการในคลังข้อมูลภาษาสัตวทดลองมีจำนวนทั้งหมด 36,658 คำ (ร้อยละ 4.36) และ 3) ประโยคแสดงวัตถุประสงค์มีทั้งหมด 133 ประโยคและรูปแบบประโยคแสดงวัตถุประสงค์มีทั้งหมด 8 รูปแบบ โดยรูปแบบที่ 3 คือ 'the personal subject' พบมากที่สุด นอกจากนี้คำแสดงวัตถุประสงค์มีจำนวนทั้งหมด 68 คำซึ่งปรากฏใน 3 รายการคำศัพท์ของคลังข้อมูลภาษาสัตวทดลองหนึ่ง กิริยาแสดงกาลที่ปรากฏในประโยคแสดงวัตถุประสงค์ประกอบด้วย 4 กาล ได้แก่ present simple, future simple, past simple และ present perfect โดย present simple เกิดขึ้นมากที่สุดในรูปแบบที่ 3 และ present perfect เกิดขึ้นในรูปแบบที่ 3 เช่นกัน

CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
ABSTRACT (ENGLISH)	iv
ABSTRACT (THAI)	v
LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER I INTRODUCTION	1
1.1 Background of the Study	1
1.2 Rationale of the Study	3
1.3 Statement of Problems	6
1.4 Purpose of the Study	10
1.5 Research Questions	10
1.6 Significance of the Study	10
1.7 Conceptual Framework	11
1.8 Scope and Limitation of the Study	12
1.9 Definition of Terms	12
CHAPTER II LITERATURE REVIEW	14
2.1 Corpus-Based Analysis	16
2.1.1 The Definition of Corpus	17
2.1.2 Type of Corpus	18
2.1.3 Concordancing Software	19
2.1.4 Corpus Design	26
2.1.5 Use of the Corpus	28
2.2 Laboratory Animal Science	32
2.3 Review Articles	33
2.4 Language Features of Scientific English	36

CONTENTS (cont.)

	Page
2.5 Word Lists	44
2.5.1 High-Frequency Words (HFW)	45
2.5.2 Academic Words (AW)	47
2.5.3 Technical Words (TW)	50
2.5.4 Low-Frequency Words (LFW)	51
2.5.5 Previous Studies of Word Lists	53
2.6 Forms and Functions of Language	55
2.6.1 Purpose of the Research	55
2.6.2 Research Methodology	58
2.6.3 Data Collection	65
2.6.4 Research Results	68
Chapter Summary	69
CHAPTER III RESEARCH METHODOLOGY	71
3.1 Research Procedure	73
3.2 Selection of the Journal	75
3.3 Designing the Laboratory Animal Review Article Corpus	77
3.4 Research Instruments	79
3.5 Procedures for Analyzing the Corpus	80
3.6 Data Analysis and Statistical Devices	87
3.6.1 Framework of the AWL	87
3.6.2 Analyzing the Forms of the Purpose Function	89
3.6.3 Statistical Device	90
Chapter Summary	91

CONTENTS (cont.)

	Page
CHAPTER IV RESULTS	92
4.1 The High-Frequency Words	92
4.2 The Content Words of AWL	98
4.3 The Words and Patterns Describing the Research Purposes	101
4.3.1 Content Words	101
4.3.2 Verb Tenses	105
4.3.3 Patterns	111
Chapter Summary	126
CHAPTER V DISCUSSION	127
5.1 High-Frequency Words	127
5.2 Content Words from the Academic Word List	131
5.3 Words and Patterns Presenting the Research Purposes	134
5.4 Implications of the Study	139
Chapter Summary	140
CHAPTER VI CONCLUSION	142
Research question 1	143
Research question 2	143
Research question 3	144
Recommendations for Further Studies	143
REFERENCES	146
APPENDICES	158
Appendix A	159
Appendix B	172
Appendix C	185
Appendix D	195
BIOGRAPHY	205

LIST OF TABLES

Table	Page
2.1 Outline of Literature Review	16
2.2 Summary of Tools and Features in AntConc 3.4.4	21
2.3 Examples of Frequent Word Formation in Bank of English (BOE)	29
2.4 Examples of Compound Nouns	36
2.5 Frequencies of Selected Features in Research Article Sections	38
2.6 Use of Active and Passive Voice in Scientific Research Articles in Methodology Section	39
2.7 Overall Frequency of Active and Passive Verbs in the Stoeger and Lightman Papers	40
2.8 Examples of Noun Suffixes	41
2.9 Summary of the Previous Corpus-Based Studies on the Frequency of Occurrence of Words on the HFW and AWL	53
3.1 Data for the <i>ILAR Journal</i> as Reported by <i>JCR Science Edition</i> <i>2014</i>	75
3.2 Composition of the Laboratory Animal Review Article Corpus	81
3.3 Composition of Academic Corpus	88
3.4 Criteria for Word-Family Distribution in the LARA Corpus	89
4.1 Statistical Details of LARA Corpus	93
4.2 Statistical Details of High-Frequency Words in LARA Corpus	93
4.3 Top-100 Most Occurred Content Words Used in LARA corpus	96
4.4 Statistical Details of AWL in LARA corpus	98
4.5 Statistical Details of Headwords of AWL in LARA corpus	99
4.6 Top-100 Most Occurred Headwords of AWL in LARA corpus	99
4.7 Content Words Describing the Research Purposes in High- Frequency Word List	103

LIST OF TABLES (cont.)

Table	Page	
4.8	Eight Patterns Describing the Research Purposes in the LARA Corpus	111
4.9	Pattern I: Signal Words for Research Purposes and To-Infinitive	112
4.10	Pattern II: Signal Words for Research Purposes as the Subject and Phrases Explaining the Details of Research Purposes	114
4.11	Pattern III: Personal Subjects	116
4.12	Pattern IV: Impersonal Subjects	118
4.13	Pattern V: Main Verbs and To-Infinitive	120
4.14	Pattern VI: Main Verbs in the Passive and To-Infinitive	122
4.15	Numbers and Percentage of Verb tenses in the Patterns Presenting the Research Purposes	125
5.1	Comparison between the AWL in LARA Corpus and Previous Studies	132
5.2	Comparison of the Purpose Patterns Used in Textbooks of Each Author	136

LIST OF FIGURES

Figure		Page
1.1	Results of searching for the key words using ScienceDirect	8
1.2	Results of searching for the key words using Scopus	8
1.3	Conceptual frameworks of the study	12
2.1	Road map of Chapter 2	15
2.2	Seven text analysis tools in AntConc 3.4.4	22
2.3	Concordance software used for analyzing corpora (cited in Anthony, 2013)	23
2.4	Frequency of occurrence and range of use of vocabulary in RANGE	25
2.5	Types of review articles	35
2.6	Four categories of words in the academic texts and their percentage of text coverage	45
2.7	Types of low-frequency words	52
3.1	Road map of Chapter 3	72
3.2	Procedures in the current study of the frequency and functions of words from various word lists in LARA corpus	74
3.3	The quartile of the <i>ILAR</i> journal from 1999 to 2014	76
3.4	Example of adding files for processing	82
3.5	Example for clicking on the “Treat all data as lower case” option	83
3.6	Example of word list sorted by frequency	83
3.7	Steps for uploading files to analyze using RANGE	84
3.8	Example of the <i>range</i> as determined by RANGE	85
3.9	Chart detailing steps in the investigation of the forms and functions used to signal research purpose	86
3.10	Framework for AWL according to Coxhead’s criteria	87
3.11	Studying the forms and functions used to signal research purposes	90

LIST OF FIGURES (cont.)

Figure		Page
4.1	Proportions of function words and content words in the high-frequency words	94
4.2	The most occurred function words in the high-frequency words	95
4.3	The first content words found in the LARA corpus	95
4.4	Percentage of word lists describing the research purposes	102
4.5	Content words describing the research purposes in AWL	104
4.6	Tenses in purpose sentences	105
5.1	Summary of the findings for high frequency words in the LARA corpus	128
5.2	Comparison of high-frequency words in this study with previous studies	129

CHAPTER I

INTRODUCTION

1.1 Background of the Study

Nowadays English is very significant and has an influence on various academic fields; especially the scientific fields because members of the discourse community of this area; namely, science researchers or students, generally have read and/or write articles or texts in English. In other words, they have to understand and be able to control the use of vocabulary, sentence patterns, and discourse patterns in order that they master the knowledge in the fields. As a result, both readers and writers have to understand more of the background of English in the scientific field, the use of scientific English, and the problems of using English of readers and writers in the scientific field.

Firstly, the members in the scientific discourse community have to understand the background of scientific English. There are several factors behind the domination of English, ranging from the geopolitical to the local. In the old Soviet bloc, English has replaced Russian because “the countries of the old Soviet bloc seek to remove themselves from the Russian influence and its language and position themselves participants in the increasingly global economy and academic community” (Flowerdew & Peacock, 2001). Since the mid-1990s, the trend for using English has increased in scientific articles; that is to say, approximately 96% of the world’s scientific articles are written in English (Engber, 2013). More than 75% of scientific papers today are published in English (Montgomery & Crystal, 2013). Currently English is used as the scientific language throughout the world (Hamel, 2007). In fact, the universality of scientific English has affected scientific communication extraordinarily (Drubin & Kellogg, 2012). In other words, worldwide scientists access the scientific literature and communicate with others all around the world through English (Drubin & Kellogg, 2012; Paltridge & Starfiels, 2013). English has also

become the language used for educated communication (Parkinson, 2000; Armer, 2011). It is accepted that English plays a gradually increasing role in scientific publication all over the world (Wood, 2001; Cook, 2003; Armer, 2011).

Moreover, scientific English is used in several ways. The words with their morphemes are used in various scientific disciplines (Flowerdew, 2002). Words from word lists such as the high-frequency word list, the academic word list, or the low-frequency word list are also used in the corpus of scientific English (Paltridge & Starfields, 2013). In addition to the word lists, various grammatical forms such as tenses and voices are used in each section of articles in scientific texts or journals (Swales & Feak, 2012; Glasman-Deal, 2010). Because of these features of scientific English, it is important for scientists to read, write, and cite academic work as often as possible in order for their accomplishments to be recognized (Flowerdew & Peacock, 2001). In other words, the academic scientific works are available to as many as possible and written in English (Wood, 2001). In addition, it is clear that the most cited journals in a variety of scientific fields are published in English. Scientists may also cite their own work published in English (Wood, 2001). The use of English as the language of science will probably continue. More and more scientists are deciding to use scientific English. Generally, there are a lot of scientific materials for scientists to read today. More scientific papers are available in English, so the value of knowing this variety is also increasing (Nation & Coxhead, 2001; Wood, 2001). Furthermore, the perceived advantages of knowing scientific English make new users connect with the international network of scientific English.

However, both native and non-native writers face problems in using scientific English. It is more obviously problematic for non-English speakers who cannot easily publish in peer-reviewed journals or simply read the journals. Those potential scientists possibly are not fluent in the language, and their inability to work or publish in English may limit their effectiveness (Brotherton, 2008). Likewise, Thai students in the field of science face problems with the English journals: writing in or reading, international journals. Pawapatcharandom (2007) found that the main English problem for Thai students in Mahidol University was writing. Major problems in writing are the inability to write academic texts in English, imperfect use of grammatical rules in writing academic papers, and inability to develop an appropriate

structure for the content. In addition to writing, the reading skills of decoding meaning through knowledge of sentence structure, guessing the meaning of vocabulary in context, and following the organizational structure of paragraphs organization was at an average level among science students in Thailand (Chawwang, 2008). Similarly, Ponmanee and Sinsuwan (2001) found that Thai students have problems with vocabulary words, sentence structure and language functions when they read texts, so they did not understand what they have read. Their study indicated that Thai students' reading ability is at a low level. Consequently, one main problem in the use of scientific English is words used in the scientific context and another problem is forms and functions of language which appear in the sentence patterns of scientific texts.

All in all, when members of the discourse community of science (science learners or researchers) understand more the background and use of English in the scientific field, it is useful for them to use scientific English in terms of reading, writing, or citing the academic works in the field of science. Nevertheless, members of the scientific discourse community such as English Thai teachers, students, or researchers still face problems in using English in the scientific field. These problems consist of lack of knowledge of vocabulary words and their forms and functions. In the next section, the reasons for analyzing vocabulary words and their forms and functions in this study are explained.

1.2 Rationale of the Study

Words and the forms used in communicative functions are the most crucial issue in the area of English for Specific Purposes (ESP) (Paltridge & Starfield, 2013). This study focuses on gathering data related to word lists and forms used in various functions; it involves a corpus-based study in the field of science. The specific area of science in this study is laboratory animal science. In the study, animals are used as subjects of the experiment or investigation. To have a better understanding of how to conduct experiments involving laboratory animals, researchers have to keep themselves up to date with current studies in the field. Reading articles in the field is mandatory for them. Knowledge of scientific English helps them understand the papers in the field. Moreover, as members of the discourse community of laboratory

animal science, the researchers must make their work available for other members by publishing papers, based on their experiments. Crucially, knowledge of scientific English helps them to produce effective reports in the field. As mentioned above, knowledge of words and functions is fundamental for developing language proficiency. Therefore, providing a study of English words and functions used in laboratory animal research is important for at least three reasons: understanding the word lists, understanding the language function of the research purposes, and understanding the guidelines in treating animals as laboratory subjects.

Firstly, there are several limitations on the word lists in each field of science. Moreover, Sprenger (2013) stated that using vocabulary words helps users display their knowledge. Vocabulary words are necessary for these people; they use words to acquire knowledge; learning the words of ESP is useful for critical reading and writing, especially academic writing. However, members of the scientific discourse community will have problems reading and writing academic texts and articles if they are not familiar with vocabulary words (Nation, 2001). In other words, vocabulary words are a difficulty for members of the discourse community of laboratory animal scientists. It is also possible for people in the field of laboratory animal science to face problems of the use of vocabulary words. If they do not understand or use words incorrectly, the meaning of academic articles can be misunderstood, which affects the humane care and use of laboratory animals. Therefore, this present study will investigate high-frequency words and academic words in the laboratory animal corpus for the following reasons. Firstly, words on the high-frequency word list (HFW) are considered as core or basic vocabulary; these words are the most frequent in both spoken and written texts and occur in all kinds of uses of the language (Nation, 2001; Paquot, 2010). Thus, members of the discourse community of laboratory animal science, who have to read and write academic scientific articles in the laboratory animal field, should understand high-frequency words; knowledge of the top-100 high frequency content words allows a better degree of comprehension of a text (Nation, 2001). In addition to HFW, academic words (AW) from the word lists are also needed by students in higher education. Academic words are used extensively nowadays in language teaching, testing, and the development of pedagogical materials (Paquot, 2010). Also, academic words are closely related to

academic writing. Hence, this study focuses on academic words to support the goal setting of learning laboratory animal science and the assessment of vocabulary in this field. When members of the discourse community of laboratory animal science know words from the various word lists in the field of laboratory animal research, they can read sentences the previous researches more effectively and also and write more effectively.

Learners have to acquire forms and functions of language to construct English sentences and reach a higher level of proficiency. Both forms and functions of language are used in both oral language and academic language. Language functions represent the active use of language for a specific purpose. Learners use language functions to engage with content and interact with others; they use language functions in order to express ideas, communicate with others, and show understanding of content in academic setting. In academic writing, learners use a range of specific functions such as describing processes, comparing or contrasting things or ideas, and classifying ideas or purposes in order to communicate ideas clearly (Swales & Feak, 2012). Recent research emphasizes the purpose function in articles because the research purpose is a common feature occurring in every article. The research purpose provides the readers the general idea; it contains the keywords of the study, which briefly generates some ideas on the part of the readers (Swales & Feak, 2012; McMillian, 2012). The statement of the purpose of the research consists of words and forms in special patterns. As a result, the current study on words and forms used for presenting research purposes is of benefit to members of the discourse community of laboratory animal science and may help them to read and write papers in the future.

Finally, thoroughly understanding the English used in the field helps researchers and students to follow guidelines in treating animals as laboratory subjects. According to the book, *Guide for the care and use of laboratory animals* (The National Academic of Sciences, 2011: 1), “all who care for, use, or produce animals for researching, testing, or teaching must assume responsibility for their well-being.” If researchers or students in the field do not understand the language used to give precautions to be used for treating animals, they might make the mistake of using more subject animals than necessary, meaning that more animals will be tortured when

the experiment is repeated. Therefore, simply understanding the language used in previous studies or could save numbers of lives.

In conclusion, the study on words and forms of language in laboratory animal science are investigated to serve the linguistic needs of members in the discourse community of laboratory animals. That is to say, if there is a model of English for laboratory animal science, it will be helpful in the care and use of laboratory animals in research, testing, and teaching. Consequently, this study will help members of the discourse community of laboratory animals, including students, teachers, and researchers, to benefit from word lists and lists of language functions, especially for research purposes and for the study of academic works.

1.3 Statement of the Problems

Because of the care for the laboratory animal use and lack of the role models and the sources of English in terms of laboratory animal science, this section outlines three problems of the study, such as the problems indicated in the interview with the director of the National Laboratory Animal Center, Mahidol University, the search for relevant research articles, and the lack of an academic English course in this field in the curriculum of the university.

The director of the National Laboratory Animal Center, Mahidol University, was interviewed on 20 March 2015; she said that laboratory animal (LA) researchers have problems with language usage. Specifically, they lack the skills of reading and writing English scientific articles. When they cannot understand English academic texts about laboratory animal science, they may make the mistake of using too many subjects in their experiments. Moreover, they cannot write satisfactory academic papers in English though they can conduct their research well. As a result, there are few articles written by Thai researchers in the field of laboratory animals. It is clear that LA researchers need to use English to read or write academic papers or texts in English. The lack of reading and writing ability on the part of LA researchers due to a lack of knowledge of word lists and language functions derived from a corpus approach in the field of laboratory animals, as shown in Figures 1.1 and 1.2. Because of the lack of knowledge about the language used in the laboratory animal field,

serious problems could arise in reading and writing such academic articles among members of the discourse community of laboratory animals.

Then, on June 1st 2015, the researcher attempted a search of two e-databases at Mahidol University for research articles involving the frequency analysis of words lists in various fields of study by using the keywords, “word lists”, “communicative functions” and “laboratory animal”. The first e-database was *ScienceDirect*, which is “a leading full-text scientific database offering journal articles and book chapters from nearly 2,500 journals and more than 30,000 books”. The other was *Scopus*, which is “the largest abstract and citation database of peer-reviewed literature: scientific journals, books, and conference proceedings”. These two databases were used to search for the data because they are considered as the most comprehensive ones; they provide a wide range of scientific studies. It was found that corpus-based frequency analyses of word lists and language functions have been conducted in a variety of fields such as engineering (Para, 2004), medical science (Chen & Ge, 2007; Wang, Liang & Ge, 2008), applied linguistics (Vongpumivitch, Huang & Chang, 2009), agriculture (Martínez, Beck & Pannza, 2009), chemistry (Valipouri & Nassaji, 2013), nursing (Yang, 2015), and environmental science (Liu & Han, 2015). However, no AWL corpus-based studies have been published in the field of laboratory animals, as shown in Figure 1.1 for *ScienceDirect* and 1.2 for *Scopus*. The lack information in the laboratory animal field this could pose serious problems for researchers and students in reading and writing academic texts and articles because L2 input is the key to building competence. Also, this information would be beneficial for teachers in planning ESP courses, especially, English for Academic Purposes course.

The screenshot shows the ScienceDirect search interface. The search bar at the top contains the keywords: "word lists in laboratory animal articles" and "language functions in laboratory animal articles". The search results page displays a total of 6,426 results. The left sidebar shows various filters such as Year (2017-2013), Publication title (The Lancet, International Journal of Psychophysiology, etc.), Topic (patient, brain, etc.), and Content type (Journal, Book, etc.). The main content area lists several articles, including "A corpus-based study of academic vocabulary in chemistry research articles" and "A new medical academic word list: A corpus-based study with enhanced methodology".

Figure 1.1 Results of searching for the key words using ScienceDirect

As can be seen in Figure 1.1, there were 6,426 results, but no studies concerned word lists or language functions in the laboratory animal science field. The search results showed studies such as *A New Medical Academic Word List* and *A Corpus-Based Study of Academic Vocabulary in Chemistry Research Articles*.

The screenshot shows the Scopus search interface. The search bar at the top contains the keywords: "word lists", "language functions", and "laboratory animal articles". The search results page displays a message: "No documents were found." The left sidebar shows various filters such as Search, Alerts, Lists, and My Scopus. The main content area is empty, indicating that no results were found for the search query.

Figure 1.2 Results of searching for the key words using Scopus

As can be seen in Figure 1.2, when the key words “word lists”, “language functions” and “laboratory animal” were entered in *Scopus*, no documents were found.

Finally, a search of the curriculum of the Faculty of Liberal Arts of Mahidol University, which is well-known for its ESP courses, revealed that the faculty provides ESP courses for students in specific fields of study, including English for Information Technology, English for Tourism, English for Hotel Business, English for Science and Technology, English for Nursing, English for Services, Business Talk, Business Writing, English for Media and Journalism, English for Logistics, and English for Career Preparation. The only ESP courses for science learners are English for Science and Technology and English for Nursing. Moreover, when a study of the curriculum for the Doctor of Veterinary Medicine Program, revealed that the program provides veterinary science learners with just English Levels 1 to 4. ESP courses such as English for Veterinary Science or English for Laboratory Animal Science are not provided for veterinary science students. Therefore, they have not had sufficient preparation to read or write the special works of their field. In other words, ESP courses are to be designed for the curricula of science learners, it would be useful for them to study specific pieces of work in English from the laboratory animal fielding order to improve their use of the four skills of academic communication, namely listening, speaking, reading, and writing. Specifically, the skills of reading and writing are essential for members of the discourse community of laboratory animals if they are to be able to read and write academic texts or papers published in English.

In short, in an interview with the director of the National Laboratory Animal Center and a search using the e-databases of Mahidol University revealed that no studies involving word lists or frequently used functions had been conducted in the field of English for laboratory animal science. Moreover, no academic English course ESP courses (English for Laboratory Animal Science or English for Veterinarians) have been offered to learners. As a result, to fill the gap and gain new knowledge regarding word lists and language functions in the field of laboratory animals, this corpus-based study on the academic word list and language functions was conducted.

1.4 Purposes of the Study

Laboratory animal science is a branch of science studied mainly by veterinarians. Reading of laboratory animal texts or articles and writing academic papers in English are problems for Thai learners or researchers. Importantly, vocabulary knowledge as well as knowledge of forms and functions of language is core factors in the reading and writing of academic articles. Based on the use of an extensive corpus of laboratory animal research articles, this study addressed the following objectives:

1. To identify the hundred most frequently used content words in a corpus of laboratory animal journal articles.
2. To investigate words from the academic word list (AWL) occurring in the same corpus.
3. To investigate the frequency of use of words and forms in the statement of research purposes which laboratory animal researchers use in review articles.

1.5 Research Questions

In order to achieve the abovementioned objectives for the research, this study explored the following questions:

1. What hundred content words appear most frequently in laboratory animal review articles?
2. What content words from the AWL occurred with high frequency in the same corpus of laboratory animal review articles?
3. What frequency of use of words and forms signal the function of indicating research purpose in laboratory animal studies?

1.6 Significance of the Study

The study of words frequently appearing on word lists can play a vital role in the teaching or learning of English for academic course or ESP program of English for laboratory animal science or English for veterinarians. Therefore, knowledge of

word lists and forms used to signal purpose are essential to language teaching and learning for the following reasons.

Findings of the study can serve as essential input for teachers and course designers of materials for English for Academic Purposes or English for Laboratory Animal Science or English for Veterinarians courses. Moreover, knowledge about which words from which word lists and which forms of language functions appear most frequently in the corpus will help in selecting the materials for the development of vocabulary and use of language functions.

Moreover, the results of this study will provide English for Academic Purposes instructors, especially English for Laboratory Animal Science, teachers who are requested to teach Academic English courses for learners of veterinary or laboratory animal science with background knowledge of words appearing on the word lists as well as forms and language functions.

All in all, it is anticipated that information from this study in terms of word frequency and language forms and functions will assist education specialists such as developers of English reading and writing materials or teachers of writing courses for science learners.

1.7 Conceptual Framework

The criteria of Coxhead (2000) determine which word lists were used in laboratory animal review articles. Then the forms and functions of academic words were analyzed in terms of the research purpose (Swales & Feak, 2012; Soranastaporn, 2013) as can be seen from the following framework.

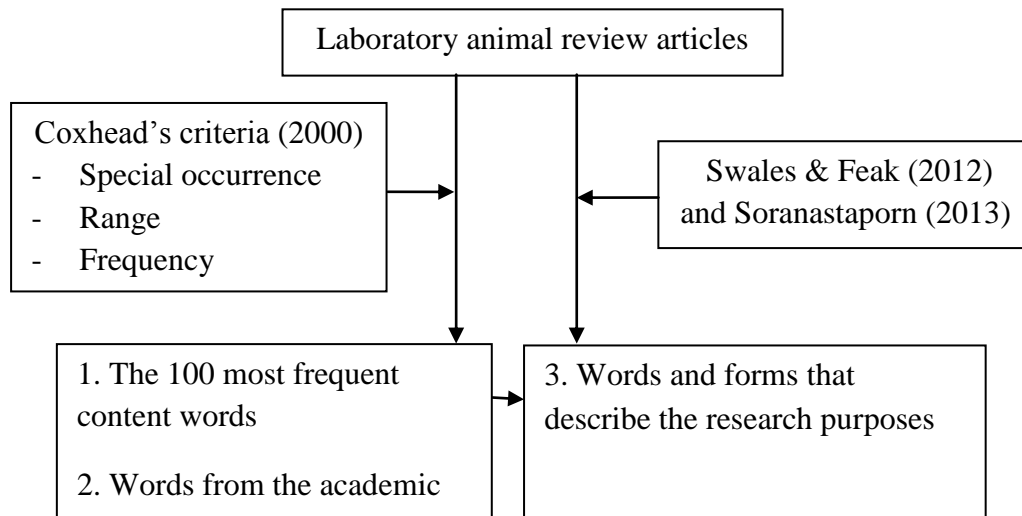


Figure 1.3 Conceptual frameworks for the study

1.8 Scope and Limitations of the Study

1. Coxhead's criteria, including *special occurrence*, *range*, and *frequency* was used in this study to determine which words from which lists appeared most frequently in laboratory animal journals.

2. Only content words and word families were counted as high frequency words in the laboratory animal corpus.

3. Verbs that can function into two categories (can be either a main verb or an auxiliary verb) such as the verb *to be*, verb *to have*, and verb *to do* were excluded from the study to avoid ambiguity in classification.

4. In this study, only words appearing in the body of the text were investigated. Titles, references, appendices, footnotes, captions, graphs, tables, formulae, acknowledgements, and bibliographies were excluded.

1.9 Definitions of Terms

Academic words refer to words contained in laboratory animal corpus as determined by using the criteria of Coxhead (2000).

Academic word list (AWL) refers to a list of word families that are commonly found in the laboratory animal corpus by using the criteria of Coxhead (2000).

Corpus refers to a simply large collection of naturally occurring instances of language or databases of language, consisting of anything from sentences to a set of written texts or tape recordings, which has been compiled for linguistic study (Cheng, 2012; Huntson, 2002; Schmitt, 2000).

Concordance refers to a list of the examples for a search item, an index of the words, with co-text from either side of the centered search item (Cheng, 2012; McEnery & Wilson, 2001; Sinclair, 1991).

Content words refer to one category of open-class words, which are meaningful words referring to things or situations or properties in the world (Murphy, 2010).

English for Specific Purposes (ESP) is an area of second or foreign language teaching and learning related to the specific needs of a particular group of learners (Hutchinson & Waters, 1987; Paltridge & Starfield, 2013).

Forms and functions refer to a link between forms and functions. In other words, the function the writers used to express the objectives of the articles is linked to the forms used.

High-frequency words (HFW) refer to words making up approximately 80% of the text, including function words and content words (Nation, 2001).

Laboratory animals or animals are any vertebrate animal (e.g. traditional laboratory animals, agricultural animals, wildlife, and aquatic species) produced for or used in research, testing, or teaching. (The National Academic of Sciences, 2011).

Word outside word list (OWL) refers to the list of words which are outside the high-frequency word list and academic word list. It covers technical words and low frequency words.

Word families refer to words consisting of a headword, its inflected forms, and closely related derived forms (Nation, 2001).

CHAPTER II

LITERATURE REVIEW

This chapter aims to summarize relevant research so as to provide a background to the present study. In other words, the section reviews the studies and presents the findings related to the problems being investigated. It is organized into six parts: corpus linguistics, laboratory animal science, review articles, language features of English in scientific fields, word lists, and forms used to signal language functions, as shown in Figure 2.1.

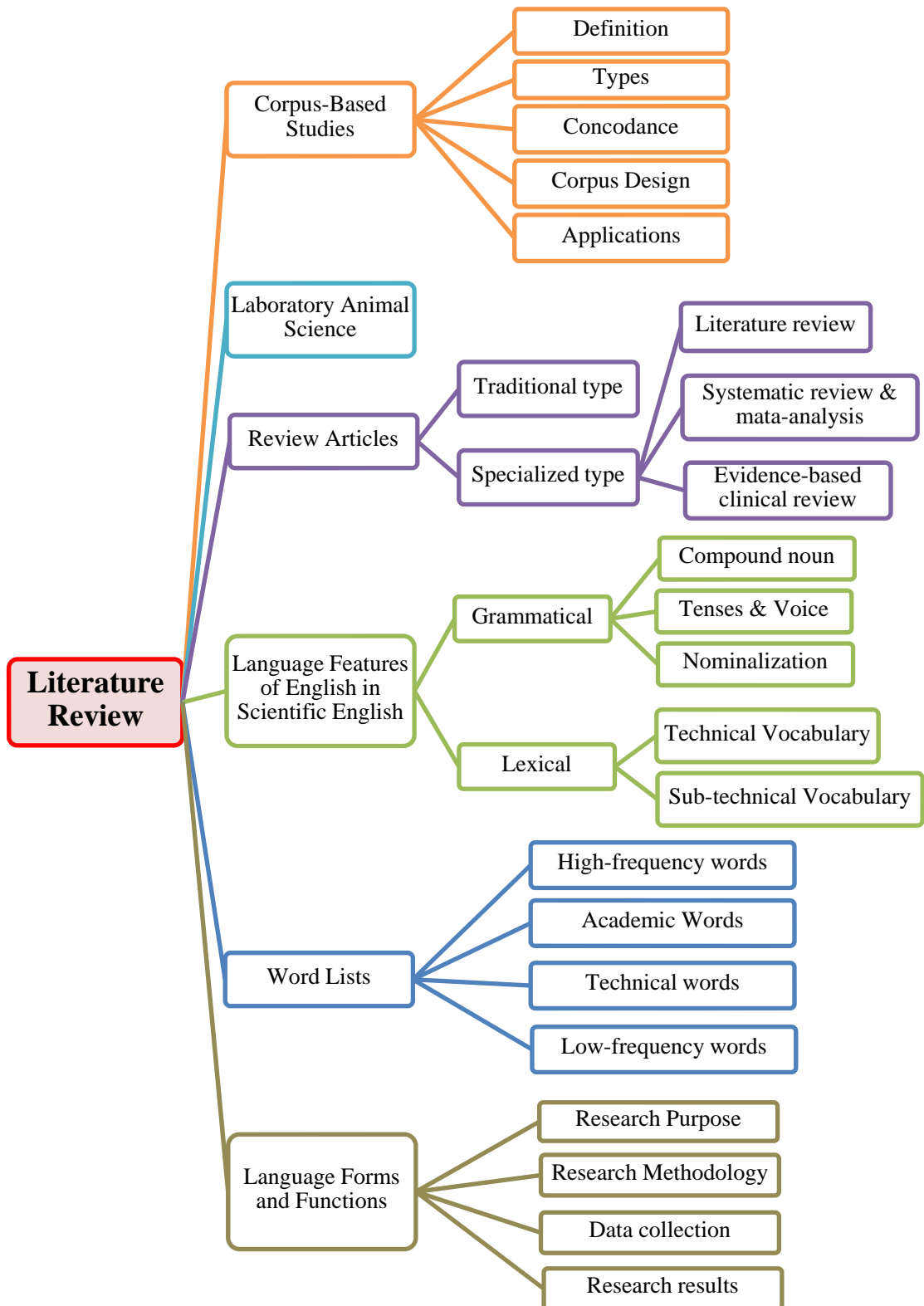


Figure 2.1 Road map for Chapter 2

As can be seen in Table 2.1, the theories and previous research related to the current study are reviewed in order to establish the conceptual orientation, refine the research problems, develop the significance, and identify methodological limitations. The theory and the previous research for this present study comprise corpus linguistics, laboratory animal science, review articles, language features of scientific English, word lists, and language forms and functions.

Table 2.1 Outline of Literature Review

Chapter Outline	Objectives of Reviewing
<ul style="list-style-type: none"> • Corpus linguistics <ul style="list-style-type: none"> - Definition - Types - Concordance - Corpus design - Applications 	<ul style="list-style-type: none"> - To understand the definition of corpus and identify corpus types before doing the recent study. - To understand the development and uses of concordance programs (AntConc and RANGE) to identify methodological limitations of this research. - To understand corpus design to design the current corpus. - To understand the use of corpus for use in this research.
<ul style="list-style-type: none"> • Laboratory animal science 	<ul style="list-style-type: none"> - To refine the research problems in terms of laboratory animal science.
<ul style="list-style-type: none"> • Review articles 	<ul style="list-style-type: none"> - To understand the components of review articles to analyze the language used in the review articles.
<ul style="list-style-type: none"> • Language features of English in scientific fields • Word lists • Forms of language functions 	<ul style="list-style-type: none"> - To establish the conceptual orientation in terms of the language features of English used in scientific fields, including word lists and forms used to express language functions for analysis of the current study.

2.1 Corpus-Based Analysis

Empirical research may be conducted using written or spoken texts. Indeed, such individual texts form the basis of many kinds of literary and linguistic analysis (McEnery & Wilson, 2001; Hunston, 2002) such as the stylistic analysis of a

poem or novel or the conversation analysis of a television talk show. Such data analysis is carried out using the corpus-based approach. Consequently, the definitions of corpus, types of corpus, concordancing software, corpus design, and applications for corpora are explained as follows:

2.1.1 The Definition of Corpus

The word *corpus* can be seen in the literature referring sometimes to a couple of short stories stored in electronic form and sometimes to the whole World Wide Web because there is no unanimous agreement on the necessary and sufficient conditions for a collection of texts to be considered a corpus. In order to discuss the fundamental principles of corpus linguistics, it is important to first establish certain limits around what can and cannot be considered definitions of corpus by emphasizing different aspects.

Crystal (1991: 95) states that a corpus (corpora) is “a collection of linguistic data, written texts or recorded speech, which can be used as a starting-point of linguistic description or as a means of verifying hypotheses about a language.” Sinclair (1991) defines a corpus as a collection of language naturally occurring in the form of texts, which are selected to typify a state or variety of a language. Schmitt (2000) say that a corpus is simply great collection or database of language, incorporating discourse ranging from a few words to the whole books. McEnery and Wilson (2001) say that the notion of a corpus as the basis for a form of empirical linguistics differs in several fundamental ways from the examination of particular texts. In principle, any collection of more than one text can be called a corpus. Similarly, Hunston (2002) defines a corpus in terms of both its form and its purpose; in other words, she describes a collection of naturally occurring examples of language, consisting of anything from a few sentences to a set of written texts or tape recordings, which have been collected for linguistic study. Sinclair (2005) also defines the corpus as a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research. Furthermore, Groom and Littlemore (2011) add that a corpus involves a very large collection of text. Likewise, Cheng (2012) defines a corpus as a collection of texts that has been compiled for a particular reason. In other

words, it is a collection of texts based on a set of design criteria, one of which is that the corpus aims to be representative.

To sum up, from the various definitions for corpus, the conclusion can be drawn that the **corpus** is collection of language in a set of written texts or a transcription of recorded speech based on a set of design criteria to conduct a linguistic study varying from words to discourse.

2.1.2 Types of Corpus

In corpus linguistics, a corpus is always designed for a specific objective. The type of corpus will also depend on the objective. Thus, corpora can be divided into eight types as follows (Hunston, 2002).

The first type is a *specialized corpus*. It is a corpus of texts of a particular type such as newspaper editorials, geography textbooks, and academic articles in a particular subject, lectures, casual, conversations, and essays written by students. The aim of this corpus type is to be representative of a given type of text. It is used to investigate a particular type of language. There is no limit to the degree of specialization involved.

Secondly, a *general corpus* is a corpus of texts of many types. It may consist of written or spoken language, or both. Also, it may include texts produced in one country or many. It is probably not representative of any particular “whole”, but will include as wide a spread of texts as possible. It is used to produce reference materials for language learning or translation.

Moreover, two or more corpora in different languages or in different varieties of a language form a *comparable corpus*. They are designed along the same lines. This corpus type in terms of varieties of the same language can be used to compare those varieties. Different languages in the corpus can be used by translators and by learners to identify differences and equivalences in each language.

The *parallel corpus* is two or more corpora in different languages. Each contains texts having been translated from one language into the other. Translators and learners can use this type to find potential equivalent expressions in each language and to investigate differences between languages.

Furthermore, a *learner corpus* is a collection of texts; especially essays, produced by learners of a language. It aims to identify in what respects learners differ from each other and from the language of native speakers, for which a comparable corpus of native-speaker texts is required.

A subsequent type is a *pedagogic corpus*. It includes all the language a learner has been exposed to. It is used to collect for the learners all instances of words or phrases they have come across in different contexts.

Additionally, a *historical or diachronic corpus* is a corpus of texts from different periods of time. It is used to trace the development of aspects of a language over time.

The last corpus is a *monitor corpus*, which is designed to track current changes in a language. This corpus is added to annually, monthly, or daily. Thus it rapidly increases in size.

In brief, there are eight corpora, each of which is designed based on a particular purpose. They are specialized corpora, general corpora, comparable corpora, parallel corpora, learner corpora, pedagogic corpora, historical or diachronic corpora, and monitor corpora. Each corpus is classified by purpose. The current corpus-based study on laboratory animal science is considered a specialized corpus for the study of linguistic features in terms of words and forms used to signal purpose in review articles in this field.

2.1.3 Concordancing Software

Concordance software, one of the oldest programs used to study a corpus, is extensively used in terms of the field of the humanities (Oakes & Netliberary Inc., 1998; Watt, 2015). The concordance software is widely used to help analyze the linguistic features of text (Fukshima, Watanabe, Kinjo, Yoshihara & Suzuki, 2012; Watt, 2015). This simple concordance software generates a word list from an authentic text corpus, which is exhibited in the center of the page and shown in context (Lamy & Klarskov Mortensen, 2012; Watt, 2015). This is also known as Key Words in Context concordance (KWIC). In other words, the essential function of the concordancer is to create word frequency lists in KWIC concordance lines (Paltridge & Starfield, 2013). But some concordance software is also able to produce a full

concordance including various linguistic elements of the corpus with all words (Lamy & Klarskov Mortensen, 2012). It is important to review the concordance programs such as AntConc and the RANGE Program used for the analysis of the words in the current corpus.

2.1.3.1 AntConc

AntConc is freeware; it is a multi-platform tool for carrying out corpus linguistics research and data-driven learning, developed by Anthony (2013). This program can operate in conjunction with many computer programs such as Microsoft Windows (tested on Win 98/Me/2000/NT, XP, Vista, Win 7), Macintosh OS X (tested on 10.4.x, 10.5.x, 10.6.x), and Linux (tested on Ubuntu 10, Linux Mint). Moreover, many corpus linguists have a great deal of confidence in AntConc for their work (Anthony, 2013). In the following section, a description of AntConc will be given: its background, its development, its tools, and its reliability and validity.

In 2002, AntConc was first released for use in a new technical writing course with over 700 graduate school engineering students at Osaka University in the west of Japan (Anthony, 2006). The main objective of the course was to produce mini-corpora of target texts. It was necessary to teach corpus linguistics as part of the course because students needed to be able to recognize and hypothesize about significant and unique language features in their specialized corpora (Anthony, 2006). However, there were two problems. One problem was the language laboratory was a Linux environment, which prevents students from using Windows-based corpus software applications. The other was the problem of funding to develop the new software, which was in-house. This software was designed for Linux and adjusted for students who might be unfamiliar with computers (Anthony, 2006). Aware of these two problems, Anthony began to develop a program which was intuitive to use and looked like an updated Windows-based application for simple corpus analysis.

In 2002, AntConc was first released; which was an easy KWIC concordancer. AntConc was developed in a Windows environment using the PERL 5.8 programming language. The graphical user interface (GUI) was also developed using the PERL/TK 8.0 toolkit. The program can be easily ported to a Linux/Unix environment, which was essential as the scientific and technical writing course was

initially taught in a Linux-based CALL (Computer Assisted Language Learning) laboratory before being changed into a Windows-based CALL laboratory the next year. Next AntConc 1.0 was released, which the program was uploaded to the author's website. Researchers, teachers, and learners around the world could easily download it and use it free of charge for non-profit use. This generated wide interest in the program and many users reported successes, problems and features they would like to see added, resulting in new, improved versions of the software. Interest in the program increased further after it was chosen to be included in Morphix NLP, a CD Linux distribution containing a wide range of natural language processing (NLP) tools. The latest version is AntConc 3.4.4. It was released in September 2014, and includes numerous tools and features, as summarized in Table 2.2.

Table 2.2 Summary of Tools and Features in AntConc 3.4.4

Tool and Features
Freeware License
Small memory requirement (2 MB of disk space)
Multiplatform
- Windows (3.4.4)
- Macintosh OS X 10.7-10.10 (3.4.3)
- Macintosh OS X 10.6 (3.4.1)
- Linux (3.4.3)
Extensive set of text analysis tools
- Concordance Tool
- Concordance Plot Tool
- File View Tool
- Clusters/N-Grams
- Collocates
- Word List
- Keyword List
Powerful Search Features
- Regular Expressions (REGEX)
- Extensive Wildcards
Multiple-Level Sorting
HTML/XML Tag Handling
Unicode Support
Easy-to-use, intuitive GUI

(Anthony, 2014)

AntConc 3.4.4 consists of seven text analysis tools accessed either by clicking on their tabs in the tool window or using the function keys F1 to F7 (Anthony, 2014); they include *Concordance Tool*, *Concordance Plot Tool*, *File View Tool*, *Clusters/N-Grams*, *Collocates*, *Word List*, *Keyword List* (See Figure 2.2).

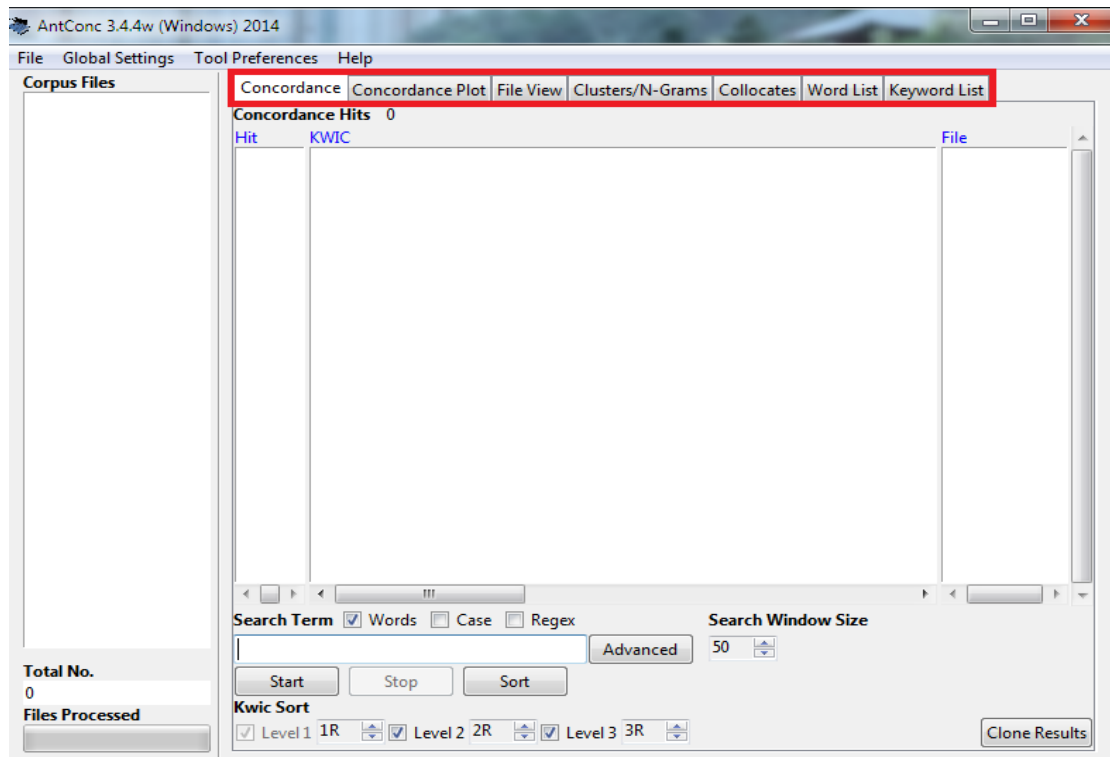


Figure 2.2 Seven text analysis tools in AntConc 3.4.4

Anthony (2014) explained each tool as follows. The first tool is *Concordance Tool*. This tool illustrates search results in a KWIC format. That is to say, how words and phrases are commonly used in a corpus of texts. Secondly, *Concordance Plot Tool* is the search results for plotting in a barcode format, which is the position of the search results in the target texts. Next, *File View Tool* is a tool which confirms the text of individual files. This investigates the results generated in other tools of AntConc in more detail. *Clusters/N-Grams* is a tool which explains clusters hinged on the search condition. It sums up the results made in the *Concordance Tool* or *Concordance Plot Tool*. *N-Grams Tool* scans the entire corpus for ‘N’ (e.g. 1 word, 2 words, ...) length clusters. This expresses the common appearances in a corpus. This next tool, *Collocates* finds the collocations for a search

term; it investigates non-sequential patterns in language. *Word List* is the tool counting for all the words in the corpus and showing them in an ordered list. This tool quickly finds which words are the most frequent in a corpus. This tool is used to analyze the frequency of words in the corpus being investigated. This last tool is *Keyword List* which determines which words are infrequent in the corpus in comparison with a reference corpus. Such words are identified in the corpus; for example, as part of a genre or ESP study.

AntConc has quickly grown in popularity among researchers, teachers, and language learners because of its rich set of features, multiplatform support, easy-to-use interface and the fact that it is freeware (Anthony, 2012). Tribble (2012) reviewed concordance software for use in analyzing corpora, as seen in Figure 2.3 (as cited in Anthony, 2013). AntConc is as frequently used to analyze the corpora as is WordSmith Tools. Both programs offer access to raw data to create corpora. However, WordSmith Tools comes with a limited commercial license that the users have to pay for. When users need to use the program on another computer, they encounter difficulties because of commercial licensing issues. AntConc, on the other hand, is accessible online; it is fast and easy to use (Anthony, 2013). It has many tools such as determining the frequency of all words, finding collocates for a search term, and finding concordance lines for the target words, as mentioned above. As a result, AntConc is considered a suitable and convenient tool for analysis in the current study.

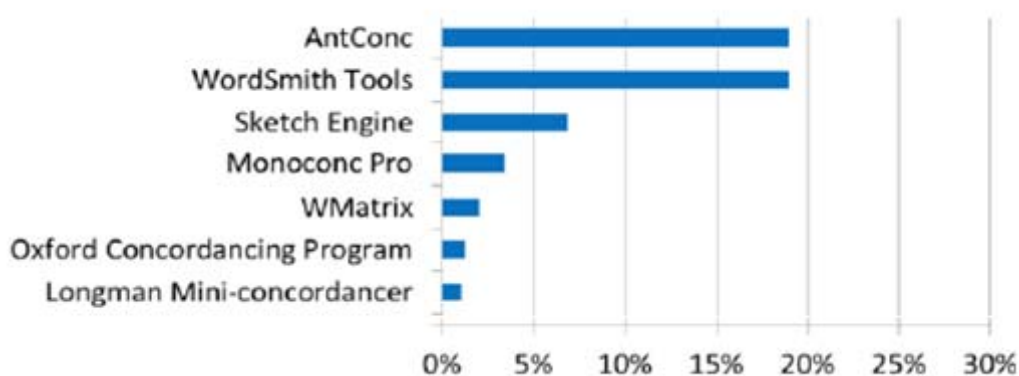


Figure 2.3 Concordance software used for analyzing corpora
(cited in Anthony, 2013)

In short, AntConc is a concordance program which is easy-to-use, fast and free. This tool can process raw corpus data of a variety of kinds. It was easy to use for data analysis of the current corpus. It is a KWIC (Key Word in Context) concordancer designed to be used by more than 700 students in a course in scientific and technical writing at the Osaka University Graduate School of Engineering in 2002. The latest version is AntConc 3.4.4, released in September 2014, comprising seven tools; that is, *Concordance Tool*, *Concordance Plot Tool*, *File View Tool*, *Clusters/N-Grams*, *Collocates*, and *Word List*, *Keyword List*. *Word List* is used for analysis of the frequency of word in the corpus. Thus, the current study used this program to determine the frequency of words in the corpus.

2.1.3.2 The RANGE Program

RANGE provides information about the allocation of words or other lexical components across a set of two or more texts. It can be used to compare corpora of texts or subdivisions of a corpus, or a set of texts supplied by a user. Thus, this program can be useful in many ways as follows.

RANGE investigates the vocabulary words in a text. RANGE is best known for the four following purposes, especially, the first three. Users are able to determine the number of vocabulary words, which is essential for gaining an understanding of the vocabulary in the texts. Moreover, users can generate word lists derived according to the frequency of occurrence and range of use in diverse discourse categories, as shown in Figure 2.4. Users also determine the number of words in the text. Finally, users can assess the vocabulary load of a text for teaching purposes. Most particularly, the development of the Academic Word List (Coxhead, 2000) and the 14 British National Corpus (BNC) 1000 word lists (Kilgarriff, 1997) as well as a series of studies by Nation and his colleagues used this program to determine vocabulary size, essential for understand different types of discourse.

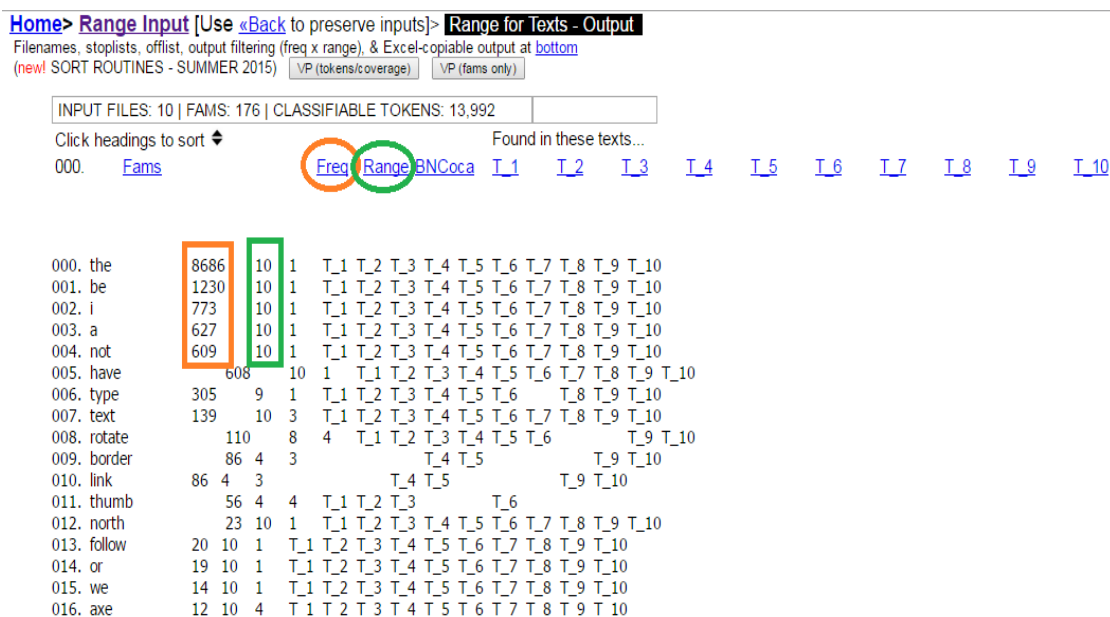


Figure 2.4 Frequency of occurrence and range of use of vocabulary in RANGE

RANGE has been used with West’s (1953) General Service List (GSL) and the Academic Word List (AWL) to determine how often the 2,000 highest frequency words and academic words were represented in a text. Furthermore, the development of the 14 BNC 1000 word lists with RANGE assesses the text more precisely because RANGE users can determine the vocabulary size, which is necessary for comprehension of authentic texts; these words usually include a small percentage of words found among the most frequent 3,000-14,000 word families. (Webb & Nation, 2008). RANGE and the word lists are free for downloading from Paul Nation’s website: <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>.

The useful features of RANGE are: 1) The software is used for analysis of the vocabulary in an individual text or as many as 32 texts at the same time. 2) It can also show the frequency with which each word is encountered in one text as well as in multiple texts because the number of encounters with unknown words may provide some indication of their potential for incidental learning and teaching (Webb, 2007). In particular, the results from analyzing the raw data via this program are helpful in learning and teaching language in terms of determining which vocabulary words are to be included in the syllabus.

To conclude, RANGE investigates the vocabulary words in text. This program can be used in conjunction with the Academic Word List (Coxhead 2000) and the 14 British National Corpus (BNC) 1000 word lists (Kilgarriff, 1997); Nation and his colleagues have used this program to examine vocabulary size. The outcomes of the raw data analyzed via this program can support language learning in terms of helping determine vocabulary words used in each lesson. The current study used this program to find the range of words in the corpus because Coxhead's study (2000) used this program to analyze an academic corpus, and it was easy to use for analyzing the range of the current corpus.

In brief, concordance software such as AntConc and RANGE can be used to create word frequency lists and to determine key words in context. AntConc is an easy-to-use, fast, and free tool. It was designed for use by more than 700 students in a course in scientific and technical writing at the Osaka University Graduate School of Engineering in 2002 to process raw corpus data of a variety of kinds. The latest version is AntConc 3.4.4, released in September, 2014, comprising seven tools; that is, *Concordance Tool*, *Concordance Plot Tool*, *File View Tool*, *Clusters/N-Grams*, *Collocates*, *Word List*, and *Keyword List*. RANGE is used to study the vocabulary words in text. The Academic Word List (Coxhead 2000) and the 14 British National Corpus (BNC) 1000 word list (Kilgarriff, 1997) as well as the vocabulary size of Nation and his colleagues' studies were developed using this program. The word lists from analyzing the raw data via this program are useful in teaching vocabulary words used in language courses. In the current study, the two programs were used for analysis of the words in the laboratory animal corpus; AntConc 3.4.4 for the frequency of words from the *Word List* tool and RANGE for the range of the corpus.

2.1.4 Corpus Design

Designing a corpus is extremely important. In order to benefit from the advantages of corpus, it has to be well designed so as to include all the information required for a study. Consequently, the principles which should be taken into consideration to design a corpus include the objectives of each study, size, representativeness and balance, variety, and chronology (Tribble & Jones, 1990; Sinclair, 1991; Kennedy & Bolitho, 1984; McEnery & Wilson, 2001, Hunston, 2002).

Firstly, *the purposes of each study* need to be considered as a major step in designing a corpus. Corpus designers have to know what exactly the aims of each study are, and who will benefit from each corpus-based study. It is likely to be either a general corpus or a specialized corpus with specified purposes

The next principle is *the corpus size*. The possible size of corpus is not limited so much by the capacity of a computer to store it, as by the speed and efficiency of the software (Hunston, 2002). Likewise, Coxhead (2000) claims that a reasonable number of words should be gathered so as to provide a better chance for the occurrence of particular items and to examine more lexical items in depth. However, there is no justification for establishing a limit on corpus size since it depends on research objectives (Hunston, 2002).

Furthermore, *representativeness and balance* are other significant matters in corpus design. A corpus is intended to be representative of a particular sort of language. Representativeness is defined as the extent to which a sample includes the full range of variability in a population (Pearson, 1998; McEnery & Wilson, 2001; Hunston, 2002). A corpus is representative of the whole when the object of the study is academic prose, casual conversation, the language of newspapers, or American English. In addition, a balanced corpus consists of equal numbers of words in each category. Nevertheless, the problem is that “representativeness” inevitably involves knowing what the character of the “whole” is. The difficulties with representativeness and balance become more complicated when the corpus is expected to represent a regional variety of English with all its complexity of internal variation.

An additional criterion for designing a corpus is *the variety of the texts*. In linguistics, a whole variety of a language is more interesting than an individual text or author (McEnery & Wilson, 2001). Thus, the corpus ought to be taken from a variety of sources (Clany, 2010). Furthermore, different registers of the language have to be included to offer a better chance of finding language from different sources in the corpus analysis (Clany, 2010). It is important to include language samples from a variety of authors, topics, registers, and sources (Sinclair, 1991).

Finally, *chronology* is another important aspect in corpus design. Texts or journals have been published in ten years, preceding the date of collection. The corpus

compiler should follow innovations in the field and consider these current changes collecting texts.

In summary, in order to obtain a larger and more diverse corpus and in order to make definitive statements about language, the design of the corpus is a significant issue. As a result, the purposes of each study, size, representativeness and balance, variety, and chronology have to be considered in corpus design.

2.1.5 Use of the Corpus

After a corpus is compiled, it needs to be analyzed to be of any value. The computer revolution has also changed this aspect, with powerful new programs that can explore the corpus and isolate more aspects of language behavior than ever before. Most readily available software packages process data from a corpus, showing frequency (Cheng, 2012). Determining the frequency of occurrence is a basic procedure, so this was the major type of information coming out the earlier corpora. The words in a corpus can be arranged in order of their frequency in that corpus. This is most interesting when corpora are compared in terms of word frequency. As a result, frequency lists from corpora can be helpful in recognizing possible variations between the corpora studied in more detail; that is, the main use of corpora is to study different aspects of linguistics including lexis, grammar, phraseology, discourse, pragmatics, as well as register and genres.

The first aspect is *lexis*. The basic information about lexis found in a corpus consists of word frequency and word formation. First, all corpora reveal which words are used in their constituent texts and how frequent they are used (Moon, 2010: 197). In addition to word frequencies, the general lexicon or the main words of language can be found. Moreover, word formation can be found in the corpus. Word formation in the words used in the corpus includes derivation and compound words (Moon, 2010: 199). For example, Table 2.3 shows frequency derivations from *colour / color*, frequent words formed with the prefix *hyper-*, and frequent compounds. Furthermore, at the basic level of word frequency, “the base and *-ed* form occur more frequently than the *-ing* and *-s* forms” (Hunston, 2010: 146). In brief, searching a corpus can provide information concerning word frequency and the formation of words.

Table 2.3 Examples of Frequent Word Formation in the Bank of English (BOE)

Derivations from color	Words formed with prefix		Frequent compounds of color		
	<i>hyper-</i>				
colorful,	colorless,	hypertension,	hyperactive,	watercolor,	color-washed,
discolor,	colorant,	hyperinflation,	hypermarket,	color-fast,	hair color (dye),
colorist,	coloration,	hypertext,	hyperventilate,	color coded,	two/four color
uncolored,	colorable,	hypersensitive		full-color,	(of printing), etc.
colorism,	coloristic,			colorway,	
recolor,	etc.				

Cheng (2012: 41)

Next, the frequency of use of words provides grammatical information. *Grammar* is regarded as “how sentences and utterances are formed” and “two basic principles of grammar are ‘the arrangement of items (syntax) and the structure of items (morphology)’” (Carter & McCarthy, 2006: 2). Corpus-based studies of English grammar have proven to be particularly practical for descriptions of language use. In other words, they facilitate to understand what speakers and writers essentially do with the linguistic resources available in English. Three types of results can be used to teach grammar; namely, frequency information, register comparisons, and associations between grammatical structures and words (lexico-grammar) (Carter & McCarthy, 2006; Conrad, 2010: 228). According to Carter and McCarthy (2006), the grammar of academic English is closer to the grammar of general written English than to the grammar of general spoken English (Paltridge & Starfield, 2013). For example, adjectives are frequently pre-modified by adverbs.

Moreover, *phraseology* is another linguistic aspect in a corpus. The frequency of use of a word or phrase in a corpus shows that not only do individual words create the meaning, but phraseology also creates meaning. Corpus linguistics has shown how words were used in texts, particularly the importance of the surrounding language (co-text) and the strong recurrent patterns of language (Cheng, 2012: 45). To different degrees, corpus studies have found that words occur as parts of phraseologies, whether collocation, structural, or both (Moon, 2010). Phraseologies range from idioms, phrasal verbs, jargon, fixed phrases, flexible phrases, fixed sentences, to collocations (Cheng, 2012). Another kind of phraseology involves non-adjacent word co-occurrences.

In terms of *discourse*, Baker (2006: 1) stated “a corpus-based approach helps to provide quantitative evidence of the existence of discourses by enabling researchers to identify repetitive linguistic patterns of language use and to uncover hidden meanings in lexical items e.g. by examining collocations.” Corpus linguistics also allows researchers to find the linguistic indications for prevailing/majority and resistant/minority discourses as a large corpus is likely to show a range of ideological positions—something which an analysis of a single text may be less likely to reveal (Flowerdew, 2002). A number of existing studies in linguistics such as corpus-based discourse analyses, recent sociolinguistic research, studies in critical discourse analysis, and the two-pronged approach use both corpus and discourse analysis.

The next application of corpus is *pragmatics*. Pragmatics is “meaning in context” (McEnery & Wilson, 2001: 98). In other words, language in use does not have a fixed meaning, or a set of meanings. Corpora strip a lot of context from utterances. They tend to use smaller samples of texts rather than the entire texts. Moreover, these samples are removed from their social and textual contexts although relevant social information is sometimes encoded with the corpus. Nevertheless, it is not always impossible to infer context from corpus texts and corpora have formed the basis for some important work in these areas. That is to say, the relation between pragmatics and corpus linguistics is problematic, as “corpora record text, not meaning” (Rühlemann, 2010:289). A corpus needs to be annotated pragmatically in order for corpus linguistics methods to be used to search for any pragmatic patterns.

Finally, corpus-based studies are used for *register and genres*. Most modern corpora are organized in terms of the text categories (Biber, 2010: 241). Many corpus studies have focused on the overall description of written registers such as advertisements, reports, email massagers, business letters, and academic prose rather than spoken registers such as conversation, service encounters, call center interactions and public speeches to examine “a suite of linguistic features that are characteristic of the register” (Biber, 2010: 246). Other corpus studies of register variation focus on a particular linguistic feature. For instance, Hyland (1998) investigated the use of hedge within the register of scientific research articles. Swales (2010) also studied the discourse functions in university speech in MICASE.

In brief, the frequency of word appearance in a corpus can be helpful in investigating different aspects of linguistics such as lexis, grammar, phraseology, discourse, pragmatics, as well as register and genres.

In conclusion, the definition of corpus, types of corpus, concordancing software, corpus design, and uses for a corpus have been presented. A corpus is a collection of the language in a set of written texts or a transcription of recorded speech based on a set of criteria. There are eight types of corpus; namely, specialized corpora, general corpora, comparable corpora, parallel corpora, learner corpora, pedagogic corpora, historical or diachronic corpora, and monitor corpora. Each corpus type is designed based on a particular purpose. Thus, the recent study on laboratory animal science used specialized corpus to study linguistic features such as words used in the field and forms used to signal purpose; the corpus was designed to be representative of review articles in the field. Moreover, two concordance programs—AntConc and RANGE—were helpful in generating word frequency lists and key words in context for this present study. AntConc is an easy-to-use, fast, and free tool to process raw corpus data of a variety of kinds. The latest version is AntConc 3.4.4 released in September 2014, comprising seven tools; that is, *Concordance Tool*, *Concordance Plot Tool*, *File View Tool*, *Clusters/N-Grams*, *Collocates*, *Word List*, *Keyword List*. The other is RANGE, used to study the vocabulary words in text. This program was used to develop the Academic Word List (Coxhead 2000) and the 14 British National Corpus (BNC) 1000 word lists (Kilgarriff, 1997) as well as to determine the vocabulary size of Nation and his colleagues' studies. The current study used these two programs to analyze the frequency and range of words in the laboratory animal corpus. In designing a corpus, certain principles should be considered. The principles are the objectives of each study, size, representativeness and balance, variety, and chronology. These principles were also considered in this study. Finally, the word lists obtained from analyzing the raw data via this program can help in the teaching vocabulary words used in language courses. The frequency word list for a corpus can be useful in different aspects of linguistics such as lexis, grammar, phraseology, discourse, pragmatics, as well as register and genres.

2.2 Laboratory Animal Science

Laboratory animals (animals) are generally defined as any vertebrate animals; that is, traditional laboratory animals, agricultural animals, wildlife, and aquatic species, which are produced for or used in research, testing, or teaching (the National Academic of Sciences, 2011). Universities, schools and companies of veterinary medicine that provide contract animal-testing services for test purposes use laboratory animals. Every area of animal use should be judged individually, and replacing the use of animals with humane alternatives must be the principal goal.

Animal use is defined as the proper care, use, and humane treatment of laboratory animals produced for or used in research, testing, or teaching (the National Academic of Sciences, 2011). Experiments on animals are carried out for many different purposes: 1) developing and testing medicines and vaccines for humans or animals, 2) studying how animals' and humans' bodies function, and 3) assessing the safety of chemicals, such as pesticides, for their possible effects on human health or the environment.

Animals can experience pain and distress in experiments; the pain and distress can be severe. The way that animals are bred, transported, housed and handled may also cause suffering. A practical approach, which coordinates with people involved in animal use in government, industry, and science ensures that the necessity and justification for using animals is always critically reviewed. Everything possible is also done to develop quickly humane alternatives. Moreover, every possible step is taken to reduce the numbers of animals used, and to significantly reduce their suffering and improve their welfare.

Some laboratory animal science groups develop and promote innovative approaches in animal-based research, aiming to improve both the scientific soundness and translational value of animal experiments, and the welfare of laboratory animals. In other words, their research addresses laboratory animal behavior, health and welfare, laboratory animal anesthesia, and ethics of animal-based research. Russell and Burch (1959) described the Three Rs (3Rs), which are guiding principles for more ethical use of animals in testing. The 3Rs are: 1) *Replacement*, which refers to the preferred use of non-animal methods over animal methods whenever it is possible to achieve the same scientific aims. These methods include computer modeling. 2)

Reduction, which refers to methods which enable researchers to obtain comparable levels of information from fewer animals, or to obtain more information from the same number of animals. 3) *Refinement*, which refers to methods that alleviate or minimize potential pain, suffering or distress, and enhance animal welfare for the animals used. These methods include non-invasive techniques.

In short, to test or use animals ethically, researchers have to read the guidelines in treating animals as laboratory subjects, or papers in the laboratory animal field. Moreover, researchers have to write reports on their laboratory results or produce research papers. They have some difficulties reading or writing in English. Knowledge of scientific English helps their comprehension of the guidelines or papers in the field. Therefore, this study investigated the features of the English language in the field of laboratory animal science so that researchers in laboratory animal science can carry out their research ethically.

2.3 Review Articles

In the world of academia, researchers are expected to publish academic articles early in their career. Publications are not done until the research is completed, but original research occasionally takes many years to complete. A variety of academic studies, original research (categorized as primary literature) and other published work (secondary literature) make up the diverse types of articles which can be published in journals. These will help in understanding the how the work can be published and help identify what kind of article would be appropriate for the study. Review articles are one type of published articles. Therefore, this following section contains a detailed description of review articles in terms of the definition and types of review articles.

Review articles are one type of article published in the scientific field. They are dictionary-like and encyclopedia-like, and range from the general to the universal (Salager-Meyer, 1992; Rethlefsen, 2014). Review articles also have a pretension to universality; they are universal. Moreover, review articles are like an essay; they consist of a topic, an introduction, a body, and a conclusion (Taylor, 2011).

The review article is both a summary and an evaluation of another writer's article (Rethlefsen, 2014). It is a non-original publication, which is only a re-organization of an acknowledged idea. However, some theory holds that the review article is original because it provides a new idea, a new approach, and practical insights to the readers (Taylor, 2011). These factors cause review articles to differ from research articles. For example, Salager-Meyer's study on communicative functions (1992) reveals that verb tenses in review articles also differ from research articles. That is, the present tense is used most frequently in review articles, whereas research articles overwhelmingly use the past.

Review articles consist of two types—the traditional type and the specialized type, as shown in Figure 2.5. The former type organizes known details in a meaningful, ideally evidence-based way. This type is sometimes called a clinical updates. The clinical review article discusses disease-related topics such as diagnosis, treatment, and prevention. The later type is the specialized type, which includes three sub-types (Rethlefsen, 2014). Firstly, the *literature review* is a long list of references, generally more than 100. This is useful to summarize research on the topic, to synthesize knowledge for further research, to encourage academic rigor, and to function as a database for health policy decisions (Taylor, 2011). The second sub-type is *systematic review and meta-analysis*. Sub-type one, the (qualitative) systematic review, refers to a review and summary without statistical analysis. Swales and Feak (2012) say that the systematic review follows a very strict method for choosing the research to review in response to carefully chosen research questions. Such review type generally follows the IMRD pattern (Introduction-Methodology-Result-Discussion). Another sub-type, the meta-analysis, is a method of summarizing the results of several research studies (Taylor, 2011). That is, it requires a knowledge of statistical methods. Finally, *the evidence-based clinical review* concentrates on a clinically related question (Siwek, Gourlay, Slawson, & Shaughnessy, 2002). It also draws special attention to the quality of the studies included for analysis (Harris, Helfand, Wolff, et al, 2001).

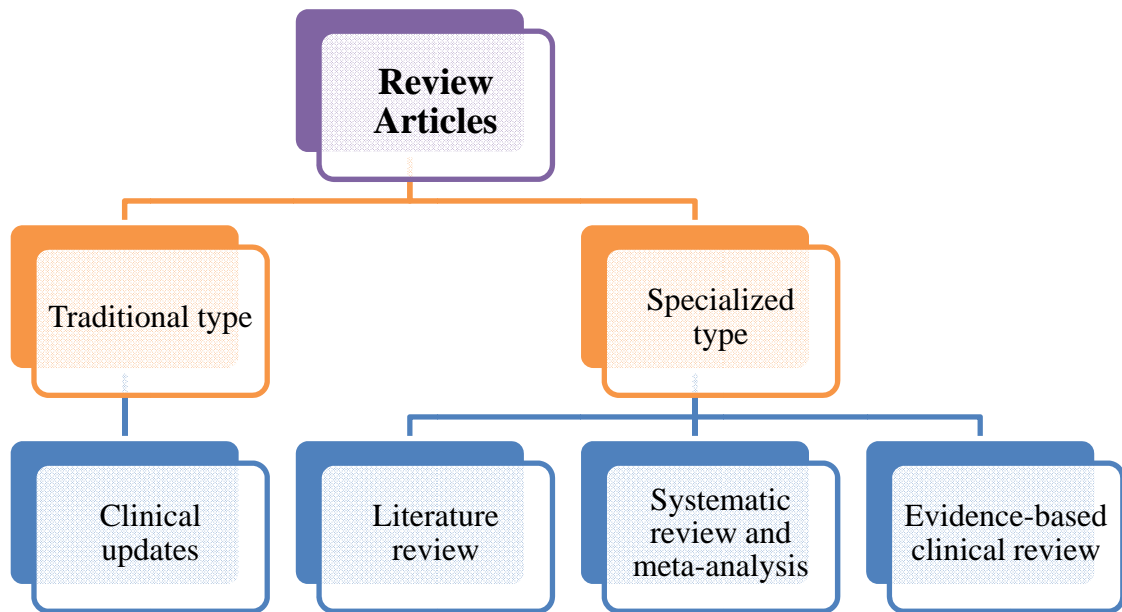


Figure 2.5 Types of review articles

The articles used in the present study are taken from the Institute for Laboratory Animal Research (ILAR) journal. These articles are classified as being of the specialized type. The journal is a peer-reviewed, theme-oriented publication of the ILAR, which provides suitable data for all who use, care for, and supervise the animals used in research. The journal publishes original articles reviewing high-quality research for the promotion of humane care in the use of animals as well as the proper consideration of the use of alternatives. According to its policy, any previously unpublished animal research reported in the ILAR Journal has to be conducted adhering to the scientific, technical, and humanely appropriate guidelines of the *Guide for the Care and Use of Laboratory Animals*.

To conclude, review article refers to both a conclusion and an assessment of the writer's article; the writing style is the same as that of the essay including Introduction, Body, and Conclusion. In the present study, the review articles, which are of the specialized type, were chosen because their elements are useful for the study of language patterns and functions. Moreover, this recent study is helpful in teaching the reading and writing of laboratory animal science academic articles.

2.4 Language Features of Scientific English

English for Specific Purpose (ESP) covers various academic fields such as English for science, medicine, technology, economics, secretarial work, psychology, or teaching (Hutchinson & Waters, 1987; Jordan, 1997; Hyland, 2006; Paltridge & Starfield, 2013). English for science is considered as a major issue for learners of science or researchers in order for them to carry out their research. Scientific English is slightly different from general English. That is, scientific articles entail a precise choice of words and avoid ambiguity or polysemy (Hyland, 2006). Moreover, ESP has tended to emphasize communication in the specialized target context (Hyland, 2007). Halliday (as cited in Parkinson, 2000), pointed out that a scientific text is composed of special features throughout the text. These special features such as compound nouns, passive voice, and nominalization as well as lexical features of scientific and technical language such as technical vocabulary and sub-technical vocabulary can be considered as the grammatical features of scientific and technical language.

2.4.1 The Grammatical Features of Scientific English

The first feature is the grammar of scientific and technical language (Drubin & Kellogg, 2012). Many grammatical areas must clearly be emphasized and reinforced in ESP courses as they are more relevant to learners' needs than other areas are. Some of the areas that need special attention in EST courses such as compound nouns, passive voice, and nominalization are reviewed.

2.4.1.1 Compound Nouns

A compound noun can be defined as two or more nouns or a noun plus adjective joined to indicate a single concept; that is to say, the total expresses a "single noun" idea (Trimble, 1985) (See Table 2.4.).

Table 2.4 Examples of Compound Nouns

Modifying noun	Head noun
a water purification	system
an air quality	program

Master (2003:5)

Master (2003) indicates that compound nouns are common in scientific writing because they function as a way to characterize information in a concise and dense form. Compound nouns are suitable in scientific writing and carry out many different functions. Some functions performed by compound nouns have been summarized as follows (Trimble, 1985; Master, 2003; Drubin & Kellogg, 2012):

1. Compound nouns are sometimes used as technical terms and can be utilized in this sense for the first mention/introduction of a technical concept.

2. Compound nouns may be used for the second and third mentions of a cluster of concepts. The compound nouns are very often the final stage in a perfect summary of a set of concepts. Hence, when the idea/concept is introduced for the first time, it may be mentioned in its full form. When the concept is mentioned again, it may be abridged and the writer may use it in a dense form; i.e. as a compound noun. For instance:

[1] The study of how quickly cracks in the glass grow has made significant progress recently. [2] Glass crack growth rate was found to be related to ...

Drubin & Kellogg (2012:1)

3. It is easier to change a compound noun within a clause than the fuller form. Thus, compound nouns are useful for presenting given information at the beginning of the sentence in subject or theme position. It is also normal to present given information in as short and compressed a form as possible.

In fact, the use of compound nouns involves some difficulties. Although readers can decode the individual words, it is not easy to understand the compound noun (Master, 2003). That is, no rules can tell when compound nouns are written as a single word. If one is unsure, checking a dictionary or a speller is the only solution.

2.4.1.2 Passive Voice

In general, various language features are used in the articles. Swales and Feak (2012) studied the use of language including the topics of tenses, passive voice, citations/references, qualification, commentary and hedging. According to their findings, the present tense is the highest used tense in the Introduction and

Discussion sections, while the past tense is the highest used tense in the Methods and Results sections. The present perfect is frequently used in the Introduction and Discussion sections. Furthermore, the passive voice is most frequently found in the Methods section. Citations are frequently used in the Methods and Discussion sections. Qualifications are highly used only in the Discussion section. Commentary is used in the Introduction and Discussion sections. Finally, hedging is most frequently used in the Discussion section. Details are shown in Table 2.5.

Table 2.5 Frequencies of Selected Features in Research Article Sections

	Introduction	Methods	Results	Discussion
Present Tense	high	low	low	high
Past Tense	mid	high	high	mid
Present Perfect	mid	low	low	mid
Passive Voice	low	high	variable	variable
Citations	high	low	variable	high
Qualifications	mid	low	mid	high
Commentary	high	low	variable	high
Hedging	mid	low	mid	high

(Swales & Feak, 2012: 287)

In scientific/technical writing, the use of the passive is one of the most frequent grammatical features. Master (1991) said that the passive form in science is frequently used in sentences because the significant idea is not who did something, but what was done. Similarly, the passive always describes an activity whether or not the agent performing that activity is stated (Trimble, 1985). Moreover, Master (1991:16) stated:

The passive has been found to be as much as ten times more frequent in one text than in another. The major stylistic factor determining its frequency seems to be related to the distinction between informative and imaginative prose rather than to a difference of subject matter or of spoken and written English. The passive is generally more commonly used in formative than in imaginative writing, notably in the objective, impersonal style of scientific articles and news items.

Likewise, Admad (2012) found that the research methodology and the description of the research tools used for experiments in scientific research articles are written in the passive voice. Table 2.6 shows the statistical results for the use of the active and passive in scientific research articles. It was found that 395 sentences (70%) used the passive voice, which is more than the use of the active voice.

Table 2.6 Use of Active and Passive Voice in Scientific Research Articles in the Methodology Section

	Number of uses	Percentage (%)
Active Voice	170	30
Passive voice	395	70
Total	565	100

Admad (2012: 54)

However, one study examining the actual use of the passive is a study by Tarone et al. (1981:201). This study used a corpus to analyze two astrophysics journal papers. The result in Table 2.7

...shows that the active voice is used much more frequently than the passive and, more importantly, that the active first person plural 'we' verb form seems to be regularly used at strategic [points].

In other words, the use of the passive voice in the study of Tarone et al. (1981) was lower than that in the studies of Master (1991) and Admad (2012).

Table 2.7 Overall Frequency of Active and Passive Verbs in the Stoeger and Lightman Papers

	Stoeger		Lightman	
	N	%	N	%
Total number of verbs	244		370	
Active verbs	217	88.5	301	81.4
Active 'we' verbs	58	23	40	10.8
<i>Passive verbs</i>	27	11.5	69	18.6
Total verbs, existential omitted.	137		248	
Active verbs	110	80	179	72.2
Active 'we' verbs	52	37	40	16.1
<i>Passive verbs</i>	27	20	69	27.8

Tarnoe et al. (1981:201)

In short, the use of the active or passive in scientific/technical writing depends on the function. Likewise, Dudley-Evans & St. John (1998:76) said that:

“The idea that scientific...writing uses the passive voice more frequently than the active is a myth; what is true is that such writing uses the passive voice more frequently than some other types of writing....The choice of active or passive is constrained by functional considerations...”

2.4.1.3 Nominalization

Halliday (1993) analyzed written scientific texts in terms of expressing the meaning nominally rather than clausally. In other words, nominalization is a grammatical feature in scientific texts. Nominalization means making something into a noun; particularly, the process of turning a verb into a noun, e.g. *approve* / *approval*; *announce* / *announcement* (Tallerman, 2005). Nominalized words usually end with suffixes such as *-ation*, *-ity*, or *-ment*. Soranastaporn (2013) provides examples of noun suffixes and their meanings, as shown in Table 2.8.

Table 2.8 Examples of Noun Suffixes

Noun suffixes	Origin	Examples	Area of meaning
-acy, -cy	(noun) state or quality	priv + acy in + fan + cy	privacy: the state of being alone infancy: the state of being a baby or young child
-ant, -ent	(noun) an agent, something that performs the action	dis + in + fect + ant de + pend + ent	disinfectant: an agent that destroy germs, something that cleans dependent: a thing supported by another; a thing determined by another
-er, -or	(noun) person or thing that does something	port + er col + lect + or	porter: a person who carries things collector: a person who collects or gathers things

Soranastaporn (2013: 8)

According to Dudley-Evans & St. John (1998), nominalization is a kind of abstract or economical language. Thus, nominalization is employed to make the phrases simpler. Dudley-Evans & St. John (1998:) indicate that nominalization enables complex information to be packaged into a phrase that is simple from a grammatical point of view and that can be picked up in the theme of the following sentence:

A high primary **productivity** is almost invariably related to a high crop yield. High **productivity** can be achieved by ensuring that all the light which falls on the field is intercepted by the leaves.... Greater **efficiency** in photosynthesis could perhaps be achieved by....”

Dudey-Evans & St. John (1998:78)

Moreover, the functions of nominalizations has been studied in science and engineering discourse (Halliday, 1993). Nominalization allows packing complex information into a dense unit and the high level of abstraction such as *curvature of material surfaces* (Pueyo & Val, 1996).

2.4.2 The Lexical Features of Scientific English

According to Trimble (1985: 30), lexis is defined as “all the words in a language” or “all the words in a language that a person knows”. However, the meaning of lexis in the field of EST is slightly more restricted, referring only to the following two areas: technical vocabulary and sub-technical vocabulary.

2.4.2.1 Technical Vocabulary

A technical word is one that is obviously specific to a specialized topic, field, or discipline. Nation (2001) has divided technical vocabulary into four categories, ordered from the most technical in category 1 to the least in category 4. See below.

Category 1: The word form appears rarely if at all outside this particular field.

Law:	<i>jactitation, per curiam, closture</i>
Applied Linguistics:	<i>morpheme, lemma</i>
Electronics:	<i>anode, impedance, galvanometer,</i>
Computing:	<i>wysiwyg, rom, pixel</i>

Category 2: The word form is used both inside and outside this particular field but not with the same meaning.

Law:	<i>cite (to appear), caution (v)</i>
Applied Linguistics:	<i>sense, reference, type, token</i>
Electronics:	<i>induced, flux, terminal, earth</i>
Computing:	<i>execute, scroll, past</i>

Category 3: The word form is used both inside and outside this particular field, but the majority of its uses with a particular meaning, though not all, are in this field.

Law:	<i>accused (n), offer</i>
Applied Linguistics:	<i>range, frequency</i>
Electronics:	<i>coil, energy, positive, gate, resistance</i>
Computing:	<i>memory, drag, window</i>

Category 4: the word form is more common in this field than elsewhere. There is little or no specialization of meaning. However, someone knowledgeable in the field would know the meaning better.

Law: *judge, mortgage, trespass*

Applied Linguistics: *word, meaning*

Electronics: *drain, filament, load, plate*

Computing: *print, program, icon*

It is evident that there is a degree of “technicalness” in which, as Nation (2001) states, this degree is required by a word’s restriction to a specific field. Thus, words under category 1 are clearly technical words, since they are only found in a specific area in both form and meaning. On the contrary, words in category 4 are common in more than one field and thus can be considered as less technical.

2.4.2.2 Sub-technical Vocabulary

The term sub-technical covers

a whole range of items which are neither highly technical and specific to a certain field of knowledge nor obviously general in the sense of being everyday words which are not used in a distinctive way in specialized texts.

(Mudraya, 2006: 105)

In general, sub-technical vocabulary is not considered a category of vocabulary, but has been called academic vocabulary (Nation, 2001; Donley & Reppen, 2001; Coxhead, 2000; Paquot, 2010). The term academic vocabulary has been used more frequently in recent years.

In summary, there are significant differences between scientific/technical language and general language in vocabulary and the higher frequency appearance of some grammatical forms. Thus, the characteristics of scientific English are different from those of general English. The grammatical features consisting of compound nouns, use of the passive, and nominalization as well as the lexical features consisting of the use of technical and sub-technical vocabulary. More details for the lexical features are explained in the next section.

2.5 Word Lists

Although less attention has been paid to the lexicon has been than to other features of language, the lexis is very important for second language learners of English for Specific Purposes (Nation, 2001; Gass & Selinker, 2008). However, research in terms of vocabulary involves a wide range of teaching and research, from English for Academic Purposes (EAP) to English for Specific Purposes (ESP) such as medicine, environmental science, and engineering (Paltridge & Starfield, 2013). The research covers everyday words with specific meanings in particular contexts and words in a narrow range of usage (Nation & Coxhead, 2001; Paltridge & Starfield, 2013). Moreover, several studies have investigated which type of vocabulary words are needed for the language comprehension of L2 learners (Paquot, 2010). Nation and Coxhead's study (2001) suggests a large number of words are needed by L2 learners in the long term, but it is not necessary for their short-term goals. Studies on the frequency of words show that some words are much more useful than others (Nation 2001; Paquot, 2010). Hence, vocabulary words are divided into four categories in academic texts; namely, high-frequency words, academic words, technical words, and low-frequency words, as shown in Figure 2.6.

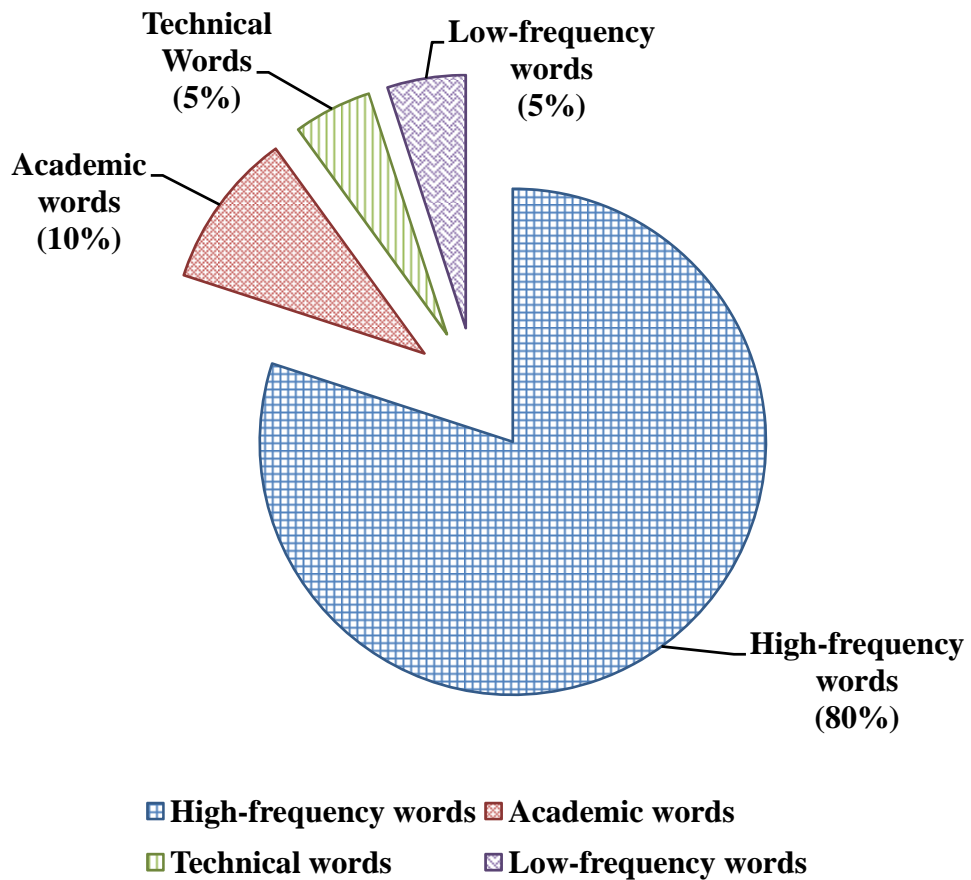


Figure 2.6 Four categories of words in academic texts and their percentage of text coverage

2.5.1 High-Frequency Words (HFW)

Counting frequency of occurrence of words is the systematic criterion to select the useful words for language learning (Schmitt, 2000). Thus, word frequency is the important principle for choosing the words on the list. Significantly, the first word list published was the list of high-frequency words because these words are found the most often (Nation, 2001). For that reason, this section consists of a definition of HFW, the number of HFW, and types of words in HFW.

2.5.1.1 Definition of High-Frequency Words

High-frequency words have been given various names. Firstly, the most famous high-frequency word list is West's (1953) *General Service List of English Words (GSL)*. Nation (2001) called these words high-frequency words. Stubbs

(1986) also called high-frequency words as nuclear words, consisting of the essential common core of pragmatically neutral words. A core vocabulary is neutral by not indicating degrees of intensity or formality (Jordan, 1997). It is also called the subject-core vocabulary, which is only expressive of a particular field (Carter, as cited in Jordan, 1997). Finally, Paquot (2010) uses the name of a core (or basic or nuclear) vocabulary, which consists of words that are of high frequency in most uses of language. In the recent study, the words with the highest frequency are called *high-frequency words*.

2.5.1.2 Number of High-Frequency Words

Deciding on the number high-frequency words is carried by examining the text coverage of groups of words sequentially ranked (Nation, 2001). The most appropriate limit for high-frequency words is regularly set at the level of 2,000 words (Nation, 2001). That is to say, a wide range high-frequency words is a crucial basic for all language use (Flowerdew & Peacock, 2001). The best known high-frequency words are composed of approximately 2,000 word families. In other words, it covers almost 80% of most texts (West, 1953; Nation, 2001; Paquot, 2010). Additionally, a variety of high-frequency words studies have been conducted as shown in Table 2.9. It shows the percentage of text coverage is different because of changes in English language and culture (Paquot, 2010). Importantly, the 2,000 highest frequency words in English serve as the best basis for academic study (Nation, 2001).

2.5.1.3 Word Types among High-Frequency Words

West's GSL (1953) consists of the 2,000 most frequently occurring word families. Words appearing in the list of high-frequency words are divided into function words and content words. Function words are words which have a grammatical function (Murphy, 2010). Function words include adverbial particles, auxiliary verbs, prepositions, conjunctions, determiners, pronouns, and numbers (Nation, 2001). Content words are words that are meaningful and denote things or situations or properties in the world (Murphy, 2010). In the high-frequency word list, there are around 165 word families consisting of function words such as *a*, *some*, *two*, *because*, and *to* (West, 1953; Nation, 2001; Paquot, 2010). The other word type is

content words, including nouns, verbs, adjectives, and adverbs (Nation, 2001; Paquot, 2010).

In short, high-frequency words have been called the General Service List of English Words (GSL), nuclear words, the core vocabulary, the subject-core vocabulary, or the core (or basic or nuclear) vocabulary. In the current study, they are called high-frequency words, which denote the 2,000 words with the highest frequency in the corpus. High-frequency words also consist of both function words and content words.

2.5.2 Academic Words (AW)

Academic words are one type of word in academic texts. Moreover, academic words are a lexical characteristic of scientific English; they are called sub-technical vocabulary as mentioned above (in subtopic 2.3.2.2). This characteristic distinguishes scientific/technical language from general language. The next section gives a definition of academic words, the creation of an academic word list, and the importance of academic words.

2.5.2.1 A Definition of Academic Words

There are various terms of academic words such as “generally useful scientific vocabulary (Barber, 1962), “semi-technical vocabulary” (Farrell, 1990), “sub-technical vocabulary” (Cowan, 1974; Baker, 1988; Mudraya, 2006), “academic vocabulary” (Nation 2001; Paquot, 2010) or “academic words” (Martin, 1976; Coxhead, 2000). The term *academic word* is used in the current study. Categorizations of academic words have been done by many researchers in linguistics. For example, Coxhead (2000) states that academic words consist of 570 word families commonly found in academic texts in the academic word list (AWL). The academic word list covers 10% of academic texts, on average. Similar to Coxhead (2000), Nation (2001) defines academic vocabulary as words other than those among the top 2,000 words in English but occurring reasonably frequently in a wide range of academic texts. Likewise, Hyland (2006) states that academic words are reasonably frequent in academic writing across disciplines and genres and comprise about 8%-10% of coverage in academic texts. Similarly, Paquot (2010) says that academic

vocabulary comprises a set of options to refer to those activities that characterize academic work, organize scientific discourse, and build the rhetoric of academic texts. To conclude, academic words are words that frequently occur in a wide range of academic texts but are excluded from the 2,000 most frequent English words. These academic words also typify academic work, organize scientific discourse, and construct the rhetoric of academic work. Furthermore, an academic word list for learners with academic goals is like a specialized extension of the high frequency words (Coxhead, 2000).

2.5.2.2 Making an Academic Word List

An academic word list can be developed in three ways.

The first method is to take a specialized area and classify the kinds of vocabulary found in the field. For example, Farrell (1990) studied vocabulary words in ESP. He analyzed English lexical entries for electronics and described the semi-technical vocabulary. The study created a list of 467 semi-technical words from the Lancaster-Oslo/Bergen Corpus (LOB Corpus), which consists of around 160,000 running words. The second technique is to collect words above which learners write first language translations in their academic texts (Lynn, 1973; Ghadessy, 1979). The last method was suggested by Coxhead (2000). This method is to take a different academic corpus and see that wide range and reasonable frequency of words are not the part of the highest frequency vocabulary. For instance, a corpus of 3.5 million running words of academic articles and books was collected to develop the academic word list. His academic corpus covers four main disciplines: arts, law, commerce, and sciences. Each discipline is divided into several subject areas. Moreover, word selection for AWL was considered by three criteria: *specialized occurrence*, *range*, and *frequency*. *Specialized occurrence* is words outside the 2,000 highest words on West's GSL (1953). *Range* is words occurring at least 10 times in each academic discipline. *Frequency* is words occurring over 100 times in the overall corpus.

The last technique, that of Coxhead (2000), provides criteria which are the most appropriate for this analysis and were therefore used in this study.

2.5.2.3 The Importance of Academic Words

Academic words are considered a source of information for learners who are preparing to read and write academic texts or articles, and teachers who are preparing to teach academic English courses. One of the challenges is making reasonable decisions about which words are worth concentrating on during valuable class and autonomous study time (Coxhead, 2000; Nation, 2001). Academic words are important for learning and teaching (Nation, 2001). As a result, there exist several reasons for the importance of academic words.

Academic words are important for learning and in the teaching of academic English courses. Nation (2001) has said it is essential for academic words to be taught in the classroom for four important reasons. First, academic words commonly occur in a wide range of academic texts, and not in non-academic texts. Academic words can also explain a considerable number of words in academic texts. Moreover, academic words are normally not known as widely as technical vocabulary words. Lastly, academic words are the sort of specialized vocabulary which English teacher can teach to help learners. In addition, academic words are helpful in achieving vocabulary goals in language courses, guiding learners in their independent study, and selecting vocabulary to develop or design academic courses and materials suitable for texts and learning activities (Nation, 2001; Paquot, 2010).

All in all, academic words are words with suitable frequency in a wide range of disciplines; they exclude the 2,000 most frequently occurring English words. Academic words are called generally useful scientific vocabulary, semi-technical vocabulary, sub-technical vocabulary, or academic vocabulary. Moreover, the attention should be given to academic words because of their high occurrence (approximately 10%) in academic texts. Academic word lists are produced by using three methods. The current corpus used the three criteria of *specialized occurrence*, *range*, and *frequency* to create an academic word list. With the importance of academic words, the academic words play important role in teaching and learning. In other words, a list of academic vocabulary can be a significant help in planning and assessing learning in teaching ESP course.

2.5.3 Technical Words (TW)

Technical words such as *indigenous* (demography), *regeneration* (science and technology), *beech* (botany), and *timber* (botany) occurring in a text are very closely associated with the topic and subject area of the text (Nation, 2001). In other words, the frequency of use of these words is related to the topic and subject area, and they are not very common elsewhere. Technical words cover approximately 5% of running words. Each subject area uses different technical words (Nation, 2001). The section discussing technical words describes categories of technical words and the making of lists of technical words.

2.5.3.1 Categories of technical words

The categorization of technical words depends on the specific topic, field, or discipline. These technical words can be classified into four categories as mentioned in subtopic 2.2.2.1. All four categories are further explained in this section. Words in category 1 are undoubtedly technical words, which are typical of a special field (Nation, 2001; Flowerdew & Peacock, 2001). These words can be learned and actually understood by studying the field. Words in category 2 are also obviously technical words because the more general meaning of the words when used outside the field does not provide ready access to its technical use. (Nation, 2001; Flowerdew & Peacock, 2001). However, words in the categories 3 and 4 are less obviously technical because both have distinctive forms or meanings depending on the text. Words in category 4 are commonly used outside the field.

2.5.3.2 Making lists of technical words

Two ways to make technical word lists are the use of a dictionary and a corpus-based frequency count. The first way for making a technical word list is the use of a dictionary. This way entails all the methodological problems of sampling, defining what is actually being counted, and classification (Nation, 2001; Flowerdew & Peacock, 2001). Another problem with this way is the choice of dictionary because the compilation of the base data and the principles which underlie the collection of the data are not done by the researcher (Nation, 2001; Flowerdew & Peacock, 2001). The other way is to use a corpus-based frequency count. The count

shows the difference between frequency and range. Technical words occur with very high or at least moderate frequency within a very limited range of texts (Nation & Hwang, 1995).

In conclusion, technical words are very closely associated with the topic and subject area of the text. They provide coverage of up to 5% of running words. Technical words are divided into four categories; categories 1 and 2 are clearly technical words, but words in categories 3 and 4 are less clearly technical. Moreover, technical word lists are made in two ways, using a dictionary and using a corpus-based frequency count. Finally, learners can learn technical words by recognizing and interpreting them.

2.5.4 Low-Frequency Words (LFW)

The final type of word list is low-frequency words. The words in this list have both a very narrow range and low frequency (Nation, 2001). In other words, low frequency words consist of words with very infrequent occurrence and small coverage (Flowerdew & Peacock, 2001). The discussion of low-frequency words in the texts is divided into two sections: the number of low-frequency words and the kinds of words found on this list.

2.5.4.1 Number of low-frequency words

Low-frequency words consist of all words except high-frequency words, academic words, and technical words for particular subjects (Nation, 2001; Flowerdew & Peacock, 2001). The unit used to count the number of words is word families. The words in this list occur once or twice in the corpus (Flowerdew & Peacock, 2001). They cover about 5% of the academic texts (Nation, 2001).

2.5.4.2 Kinds of words on the low-frequency word list

Low-frequency words are those occurring only once or twice; low-frequency words are those words outside the high-frequency boundary, proper nouns, one person's technical vocabulary, and words rarely used (See Figure 2.7). Firstly, words outside the high-frequency boundary are words occurring with moderate frequency and which do not appear on the high-frequency word list (Nation, 2001).

For instance, in the Brown Corpus, words outside the high-frequency list are those such as *curious*, *wing*, *arm (v)*, *gate*, *approximately*. Secondly, proper nouns are low-frequency words occurring in the list the most. The meaning of proper nouns is closely linked with the context (Nation, 2001). For example, the words *Carl*, *Johnson*, and *Ohio* are low-frequency words in the Brown Corpus. The next kind is one person's technical vocabulary. This kind refers to the vocabulary people use to discuss their jobs, interests, and specializations (Nation, 2001). Lastly, daily words rarely used alludes to rarely expressed ideas, old-fashioned expressions, words exhibiting very formal markedness; namely, those from a particular dialect, or vulgar or foreign words (Nation, 2001). For instance, the Brown Corpus included the low-frequency words *eponymous*, *gibbous*, *bifurcate*, and *plummet*.

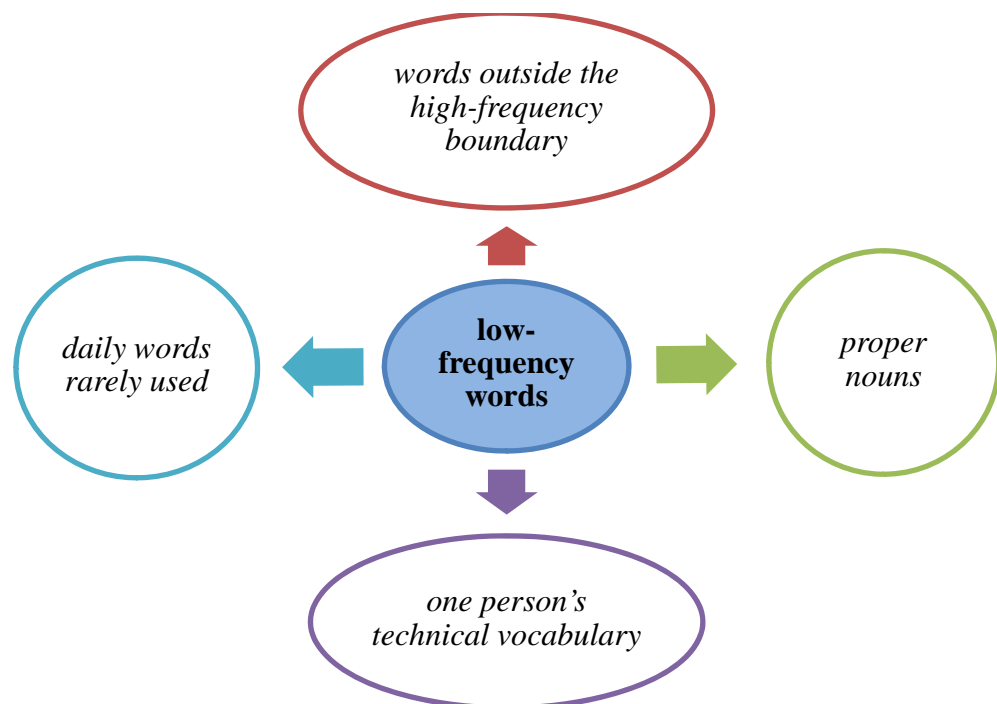


Figure 2.7 Types of low-frequency words

In summary, low-frequency words are words with a narrow range and low frequency. They cover about 5% of academic texts. Moreover, LFW consists of four kinds of words; that is to say, words outside the high-frequency boundary, proper nouns, one person's technical vocabulary, and words rarely used.

2.5.5 Previous Studies of Word Lists

There has been a lot of discussion and some research on word lists. Several studies have investigated high-frequency words (HFW) and/or the academic word list (AWL), which is needed in the academic study of ESP, as shown in Table 2.9.

Table 2.9 Summary of the Previous Corpus-Based Studies on the Frequency of Occurrence of Words on the HFW and AWL

Authors (Year)	Corpus	Number of Words*	Percentage of Text Coverage	
			HFW	AWL
Coxhead (2000)	an academic text corpus	3,513,330	76%	8.5%
Kaewphanngam (2002)	a psychology corpus	236,086	51.48%	7.78%
Para (2004)	a civil engineering research articles corpus	577,623	72.54 %	12.46%
Hyland & Tse (2007)	an academic corpus	3,213,477	74%	10.6%
Chen and Ge (2007)	a medical research articles corpus	190,425	-	10.07%
Wang, Liang, and Ge (2008)	a medical RAs corpus (online resources)	1,093,011	-	12.24%
Vongpumivitch, Huang and Chang (2009)	an applied linguistics corpus	1,500,000	-	11.17%
Martínez, Beck, and Panza (2009)	an agricultural sciences corpus	826,416	67.53%	9.06%
Li and Qian (2010)	Hong Kong financial services corpus	6,300,000	72.63%	10.46%
Valipouri and Nassaji (2013)	a chemistry corpus	4,000,000	-	9.60%
Cherngchawano and Jaturapitakkul (2014)	a Thailand University Admission test corpus	55,161	85.05%	4.58%
Liu and Han (2015)	an environmental science corpus	862,242	70.61%	12.82%
Yang (2015)	a nursing corpus	1,006,934	-	13.64%

* *running words*

The statistical device used for data analysis in the previous studies was descriptive statistics including frequency and percentage. High-frequency words are responsible for almost 80% of text coverage, whereas academic words cover approximately 10%. The highest AWL coverage was 13.54% of the nursing text, as shown in Table 2.9, while the lowest was 4.58% for the Thailand University Admission test corpus (Cherngchawano & Jaturapitakkul, 2014). The results shows that words inform the AWL do not occur as frequently at the secondary school level as at the university level. Furthermore, the HFW coverage for the secondary school level was the highest (85.05%) because these high-frequency words are met in direct teaching, direct learning, incidental learning, and planned encounters (Nation, 2001). Meanwhile, the figure for AWL coverage was around 10 % on average, which means academic writers in each special field use these academic words to work in an academic way (Nation, 2001).

In brief, various previous studies analyzing the use of words (HFW and AWL) have been conducted in different fields as shown in Table 2.9. Frequency and percentages is the statistical device used for data analysis in the previous studies. The results for the frequency of use of words in each field are different because the English used in each field is different. Moreover, there are no corpus studies in the laboratory animal field. Therefore, these previous studies give some useful guidelines to the current study and offer some ideas for data analysis of the study on the word lists in the laboratory animal field.

In summary, the word lists include the high-frequency word list, the academic word list, the technical word list, and the low-frequency word list. The high-frequency word list, or the first 2,000 word families, which contains function words and content words, accounts for approximately 80% of text coverage. Words from the academic word list, responsible for about 10% of text coverage, occur in a wide range of academic fields but are not among the high frequency words. Technical words, which account for up to 5% of text coverage, are associated with the topic and subject area of the text. Low-frequency words with a narrow range and low frequency account for about 5% of text coverage. Additionally, analyses of the words used in various academic fields have been conducted, as shown in Table 2.9. In the current study,

high-frequency words and academic words in the laboratory animal science field were selected for study

2.6 Forms and Functions of Language

To write review articles in scientific fields, scientific researchers need to know the forms used to express various functions in order to write sentences about the purpose of the research, research methodology, data collection, discussion, and research results. Writing such sentences depends on several factors such as the proficiency in language use, writing style, or love of writing (Soranastaporn, 2013). It is necessary for research writers to be able to correctly express language functions needed to express the purpose of the research, research methodology, data collection, and research results.

2.6.1 Purpose of the Research

The research purpose is outlined or stated in the Introduction (Pyrzczak & Bruce, 1998; Swales & Feak, 2012). The research purpose can usually be found at the end of the Introduction section. (Pyrzczak & Bruce, 1998; McMillian, 2012). This element is an obligatory orientation for readers to the proposed investigation (McMillian, 2012). It has two variants: Purposive (P) and Descriptive (D) (Swales & Feak, 2012).

Purposive (P): The author or authors indicate their main purpose or purposes.

e.g. The aim of the present paper is to give ...

Descriptive (D): The author or authors describe the main feature of their research.

e.g. This paper reports on the results obtained ...

(Swales & Feak, 2012: 355)

The purpose in the abstracts of research articles can be written using three forms including the use of words signaling purpose, the pattern describing the design or conduct of the study, and verbs meaning study or research (Soranastaporn, 2013) as followings:

2.6.1.1 Words signaling purpose

The words signaling purpose consist of various words such as *investigate, explore, discover, find, determine, and develop*. These words are modified by a reference to the present text, which serves as the subject of the sentences. The main verb of the sentences is *to be*, followed by *to-infinitive*.

The purpose(s) of	this	paper book research study	is (was) are (were)	V. inf.*		Noun details of purpose (s).
The aim(s) of				(to + VI)		
The goal(s) of				investigate	find	
The objective(s) of				explore	develop	
The reason(s) of				determine	discover	

* *V. inf* = *Verb infinitive*

Soranastaporn (2013: 122)

e.g.

(1) This aim of this study was to investigate whether cloze exercises could be used in teaching English in order to improve reading and writing abilities.

(2) The purpose of this study was to find the correlation between standardized and teacher-produced tests.

(3) **Objective** To examine the association of education with coronary artery calcium (CAC), an indicator of subclinical atherosclerosis, and cardiovascular risk factors, and their changes as potential mediators.

(4) The purposes of this research and development were: (1) to analyze situations and participation in the Patient Safety Management of a Medication Administration System (PSM-MAS), (2) to develop the PSM-MAS, and (3) to evaluate the PSM-MAS at Ubonrak Thonburi Hospital.

2.6.1.2 Statements explaining the design or conduct of the study

A reference to the present text is the subject of the sentences. The main verbs are in the passive forms, followed by *to + VI* or *in order to + VI* to present the purpose of the study.

This	paper book research study	(was)	designed conducted	to in order to	V. inf.	noun or noun phrase.
					study compare analyze ascertain examine evaluate	

Soranastaporn (2013: 125)

e.g.

- (1) This native study was designed to enlarge understanding about healing and spirituality in the nurse-client interaction.
- (2) This study conducted to explore the competencies for caring HIV and AIDS patient by professional nurses.

2.6.1.3 Verbs meaning study or research

A reference of the present text is also the subject of these sentences. The main verbs are in the past, followed by nouns or noun phrases.

This study	The past form of verbs (V2)	noun or noun phrase.
	investigated explored discovered analyzed developed	

Soranastaporn (2013: 125)

e.g.

- (1) This study investigated the relationship between the power of department chairs and the empowerment, compliance, and conflict perceived by faculty members at one Thai public university.
- (2) This study analyzed the discourse of visually impaired older adult friends while they talk about sexuality.
- (3) This research used quantitative and qualitative methodologies (triangulation) to determine the behavioral characteristics of nursing leadership associated with effective coalition building and work among the

Colleagues in Caring (CIC) coalitions and the influence of mentorship on transformational nursing leadership behaviors.

In each type of purpose statement, there are various words expressing purpose. Swales and Feak (2012) stated that some reference to the present text such as the uses of *this*, *the present*, *reported*, and *here* is a signal for the purpose statement. They come before the purposive sentences. Furthermore, it is common to change from the impersonal to the personal by using “*we*” or, more *rarely*, “*I*”. In examples (1) and (2), statements similar to example (1) are found more frequently than sentences like example (2).

(1) We present the results of three experiments in this paper.

(2) In this paper we present the results of three experiments.

(Swales & Feak, 2012)

Swales and Feak (2009) found that both the present tense and the past tense are used in purpose statements. In fact, how verb tenses are used relies on the reference to the work. There are two options: 1) referring to the type of *text* or *genre* and 2) referring to the type of *investigation*. The former refers to the paper, article, thesis, report, or research note. The present tense is used with this option, but the past tense is also used when referring to an original aim which has now changed. The later refers to the experiment, investigation, study, or survey. This option uses either tense. However, the present increasingly tends to be used to indicate the freshness and current relevance of the research.

2.6.2 Research Methodology

The research methodology typically comprises subjects, populations, samples, sampling, and research instruments. When writing sentences to explain these, the researchers have to write short sentences with specific details (Soranasaporn, 2013). The language forms used to describe research methodology can be divided into four patterns:

2.6.2.1 Statements about population or sample and sampling,

2.6.2.2 Statements about research instruments,

2.6.2.3 Statements about the validity of research instruments, and

2.6.2.4 Statements about the reliability of research instruments.

2.6.2.1 Populations and Sample

A statement referring to the population or sample can be written in three ways.

Statement about the subjects

Subjects indicating the subjects of the study are followed by the verb *to be*.

Parts (noun or noun phrase)	Verb to be	Noun or noun phrase
Subject (s)	was / were	details of subjects

Soranastaporn (2013: 129)

e.g.

(1) The population was professional nurses and the head of the Medicine Department.

(2) The population in this study was 208 administrators who were at the first level of nursing management.

(3) The subjects of this study were 177 second-year Medical Students of the 2004 academic year attending the Faculty of Medicine at Mahidol University.

Statements referring to the population and sample

Subjects indicating the population and the sample and are followed by the main verb in the past tense.

Whole (noun or noun phrase)	The past tense of verbs (V2)	Parts (noun or noun phrase)
The population The sample	included comprised consisted of	subjects (Noun = PEOPLE)
	was made up of was composed of was formed of was constituted of	procedures operations the method (Noun = ACTION)
	was formed by was constituted by	procedures operations

Soranastaporn (2013: 130)

e.g.

(1) The study included 680 older people diagnosed with respiratory tract infections in 22 Canadian nursing homes.

(2) Faculty sample consisted of 551 faculty members at Mahidol University in Thailand.

(3) The target population of this study comprised a representative sample of people age 30-64 years residing in Finland during 2000-2001. Of the 5,871 subjects, 4,783 (81.5%) were included in this study.

Statements referring to sampling

Subjects describing the sample are followed by main verbs in the passive form.

Sample	Verb to be	Past participle	Sampling method
This group	was	selected by	Sampling method

Soranastaporn (2013: 131)

e.g.

(1) The subjects were 230 nurses. They were selected based on hospital size by stratified random sampling.

(2) The sample comprised 351 registered nurses who worked at community hospitals Region 8. They were selected by two-stage cluster sampling.

(3) *The sample group comprised 230 professional nurses and was selected by proportional stratified random sampling and simple random sampling techniques according to the size of hospitals.*

2.6.2.2 Research Instruments

It is necessary to include details about all research instruments used in a study when describing the methodology. The sentences in this section should be tightly written, and include types of instruments, the name of the instruments, and the reliability and validity of the instruments.

Research tools

The subject of the pattern is the type or name of a tool. The main verb is in the form of passive *be + used as*, followed by the phrases *a research tool/tools* or *an instrument/instruments*.

Type	of	a tool / tools or an instrument / instruments	was/ were	used as	a research tool / tools or an instrument / instruments
Name					

Soranastaporn (2013: 133)

e.g.

(1) Questionnaires were used as research tools.

(2) An individual interview, a focus group interview, observation, field notes, and in-depth interviews were used as tools to collect data.

Research tools for collecting data

The subject of the pattern is the type or name of a tool. The main verb is in the form of passive *be + used for*, followed by a gerund.

Type	of a tool/tools or an instrument/instruments	was were	used for	V-ing	data
Name					

Soranastaporn (2013: 133)

e.g.

(1) Observations, interviews, and documents were used for collecting data.

The outcome measure

The subject of the pattern is the outcome measure (s). The main verb is in the past form of verb to be, followed by the name of a tool / tools or an instrument/ instruments.

The outcome measure(s)	was	the name of a tool/tools or an instrument/
	were	instruments

Soranastaporn (2013: 134)

e.g.

(1) Main outcome measures were the Brief Psychiatric Rating Scale (BPRS), the Specific level of Functioning Scale (SLOF), the Patient Satisfaction Instrument (PSI), the Readmission rate, and Cost-Benefit analysis.

Research tools and researchers or reasons

The subject of the pattern is the type or name of a tool. The main verb is in the form of passive *be + used*, followed by *to-infinitive* or *in order to + V inf.* to present the reason for using the tool.

Type	of a tool/tools or an	was	used	to	V inf.	from subjects
Name	instrument/instruments	were		in order to		reasons

Soranastaporn (2013: 135)

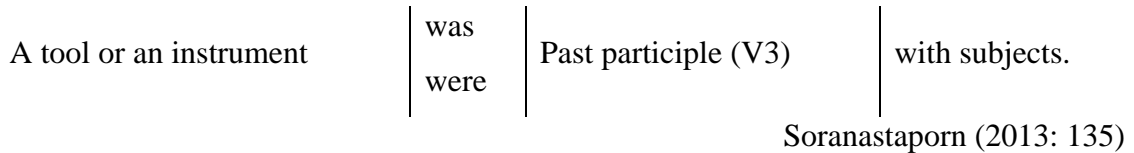
e.g.

(1) Descriptive survey methods were used to collect data from twelve school administrators and 153 teachers.

(2) Think aloud and individual interviews were used in order to find more in-depth data about the use of language learning strategies.

Research tools with subjects

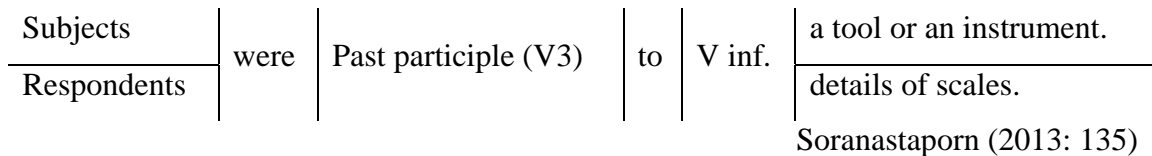
The subject of the pattern is the type or name of a tool. The main verb is in the form of passive *be + V3*, followed by the subject.



- e.g.
 (1) Interviews were conducted with ten nurses.

Subjects or respondents as the subject of a sentence

The subject of the pattern is the experiment all subjects or respondents. The main verb is in the form of passive *be + V3*, followed by *to-infinitive* and a tool or an instrument and details about the rating scales.



- e.g.
 (1) Respondents were asked to answer mailed questionnaires.
 (2) Respondents were asked to rate the appropriateness of 19 specified teacher behaviors on mailed questionnaires.

The composition of the research tools

The subject of the pattern is a tool or an instrument. The main verb is verb phrase introducing a discussion of its components, followed by the name of each component.

Whole (noun or noun phrase)	Verb phrase	All parts must be named
Instruments Tools	comprised consisted of be composed of be made up of	the name of each section Soranastaporn (2013: 136)

e.g.

(1)The instruments consisted of two sections: the head nurses’ exemplary leadership and the staff nurses’ organizational commitment.

(2) The instruments consisted of six sections: (1) demographic data, (2) knowledge, (3) skills, (4) attitudes, (5) nursing administration process, and (6) the application of the nursing process.

2.6.2.3 The Validity of the Research Instruments

The subject of the pattern is a word introducing the concept of the validity of the research instruments. The main verb is in the passive form *be + V3*.

The validity of	the tool(s) the instrument(s)	was	past participle (V3)
-----------------	----------------------------------	-----	----------------------

Soranastaporn (2013: 138)

e.g.

(1)The validity of the tools was verified (by five experts).

(2) The index of item objective congruence (IOC) was 1.

(3) The content validity indexes (CVI) of those two sections were .83 and .80 respectively, while the internal consistency coefficient alphas were .94 and .97 showing their high reliability.

2.6.2.4 The Reliability of the Research Instruments

The subject of the pattern is a word the concept of the reliability of the research instrument. The main verb is the past form of the *verb to be*, followed by a figure representing the reliability value.

The reliability	of	the tool(s)	was	figure of the reliability value.
The reliabilities		the instrument (s)		

Soranastaporn (2013: 139)

e.g.

- (1) The reliability of the questionnaires was .96.
- (2) The Cronbach's alpha reliability coefficient of the second tool was .93.
- (3) The Cronbach's alpha reliability coefficient of part 2 to 6 were .86, .97, .90, .94, and .98 respectively.
- (4) The internal consistency reliabilities were .89 to .93 across all four samples.

2.6.3 Data Collection

Researchers collect two types of data: data for quantitative research and data for qualitative research.

2.6.3.1 Data Collection for Quantitative Research

The way to provide subjects with the research tool

The subject of the pattern is a tool or an instrument. The main verb is in the passive, followed by a description of the way to present subjects with the research tool.

A tool/Tools	was	Past participle (V.3)	to	subject.
An instrument/Instruments				

Soranastaporn (2013: 142)

e.g.

- (1) Questionnaires and letters of co-operation were mailed to faculty members.

The number of questionnaires returned

The subject of the pattern is the number of questionnaires returned. The main verb is in the passive form *be + V3*.

A return rate of | percentage | (proportion of respondents) | was | past participle (V.3)
Soranastaporn (2013: 142)

e.g.

(1) A return rate of 84.94% (468 out of 551 potential respondents) was obtained.

2.6.3.2 Data Collection for Qualitative Research

The researchers provide details about data collection for qualitative research such as the place, the time period, and the questions used for interviews.

Instruments, Place, and Duration

The subject of the pattern is a tool or an instrument. The main verb is in the passive form *be + V3*, followed by the place or duration.

A tool / Tools | was | Past participle | in | place | duration
An instrument/Instruments | were | (V.3) |
Soranastaporn (2013: 143)

e.g.

(1) Observations and interviews were conducted in the Medical Department over an eight-month period.

Data Collection and Duration

Observations, interviews, and documents are used for collecting data. The subject of the pattern is a word introducing the idea of data collection. The main verb is in the passive form *be + V3*, followed by the duration.

Data	were	Past participle (V.3)	Adv. of time	on	D1, D2, D3, D4, and D5.
------	------	--------------------------	--------------	----	-------------------------

Soranastaporn (2013: 143)

e.g.

(1)[Observation, interviews, and documents were used for collecting data.]
Data were collected before, during, and after implementation on nursing costs, empowerment, staff, physician, and patient satisfaction, continuity of care, and patient outcomes.

How and How Long the Data Were Collected

The active and the passive forms show how and how long data were collected and what data was collected.

Active voice

The subject of the pattern is the data collector. The main verb is in the past tense, followed by what data was collected.

Subject	The past form of verbs	object
Personal Noun	V2	Noun.

Soranastaporn (2013: 144)

e.g.

- (1)The investigator asked one broad question, “What has been your experience with healing and spirituality?”
- (2) The interviews lasted from 35 minutes to one hour.

Passive voice

The subject of the pattern is what data was collected. The main verb is in the past passive.

Subject	Verb to be	Past participle
Noun	was/were	V3.

Soranastaporn (2013: 144)

e.g.

- (1) The nurses' responses to this question were taped-recorded.
- (2) Tapes were transcribed and then analyzed to elicit key ideas and themes.

2.6.4 Research Results

Research results are indicated by words that signal to the readers that the results will be presented. The subject of the pattern is a signal words meaning *result*. The main verb is in the past tense, followed a *that-clause* indicating the results. The pattern for showing research results is presented below.

The results of	data analysis this study this research	The past form of verbs	that	clause indicating results.
		showed revealed illustrated demonstrated explained yielded		

Soranastaporn (2013: 150)

e.g.

- (1) Results of this study showed that the case management model seemed to have a positive impact on clients' mental status, functional level, and satisfaction.
- (2) The results of data analysis showed that expert power was the most effective power base in this study, followed by legitimate power and referent power; conversely, reward power and coercive power were ineffective in terms influence.

(3) Multivariate analysis revealed that the following client characteristics predicted nurse case management contact: age, income, living alone, and stage of cancer.

In summary, in scientific articles in English, the relationship between form and function is important for writers and readers. The functions in scientific articles consist of research purposes, research methodology, data collection, and research results. The research purpose is the first part that readers read to know what the article describes. Importantly, each function has different forms. Therefore, it is necessary to study these forms and functions to be able to teach ESP learners in English academic courses. In the present study, the forms used to indicate the research purpose have been studied because the research purpose is the first part that the readers read to understand the general idea of the research. Also, the purpose of the research is given in every article.

Chapter Summary

The literature review section has summarize and reviewed the corpus-based study, laboratory animal science, review articles, language features of scientific English, word lists, and language forms and functions in order to provide a framework for the study and guide a corpus-based study of word lists and purpose functions. The term of review article refers to both articles giving conclusions and those providing an assessment of the article being reviewed. The format is the same as that of for an essay including Introduction, Body, and Conclusion. Moreover, there are two types of review article. The first is the traditional type or clinical update, which mentions disease-related topics such as diagnosis, treatment, and prevention. The other is the special type consisting of a literature review, a systematic review and meta-analysis, or an evidence-based clinical review. In addition, such articles feature a mixture of scientific English and general English and include numerous grammatical variations. Grammatical features include compound nouns, use of the passive, and nominalization. Lexical features include a mixture of technical vocabulary and sub-technical vocabulary. Furthermore, word lists used to analyze academic texts include the high-frequency word list, the academic word list, the technical word list, and the

low-frequency word list. Words on the HFW comprise approximately 80% of text coverage. Secondly, academic words account for about 10% of text coverage; these words are used in a wide range of academic fields, but do not figure on the list of high frequency words. The next word list is technical words covering up to 5% of running words; these words are associated with the topic and subject area of the text. The last word list covers about 5% of text; it consists of low-frequency words with a narrow range and low frequency in academic texts. Finally, previous corpus-based studies of word lists are the guideline for the recent study. Both previous and current studies on word list are also helpful for language pedagogy. Knowledge about the relationship between form and function in scientific articles in English is essential for writers and readers. The functions in scientific articles include research purposes, research methodology, data collection, research results, and discussion. The research purpose is the first part that readers read and learn what the article is about. Importantly, each function uses different forms. Therefore, it is very useful to study these forms and functions so as to teach ESP learners in English academic courses. Lastly, corpus analysis consists of the study of lexis, grammar, phraseology, discourse, pragmatics, as well as register and genres.

CHAPTER III

RESEARCH METHODOLOGY

The objectives of the recent study are 1) to identify the hundred most frequently appearing content words in a corpus of laboratory animal journal articles, 2) to determine which words from the academic word list (AWL) occur in the same corpus, 3) to investigate frequencies of the words and forms used to signal purposes which are used by laboratory animal researchers present their objectives. The research procedures for the study are given in this section to reach the research objects. To this end, the following items are presented: research procedure, selection of the journal, designing the laboratory animal corpus, research instruments, corpus development, and data analysis and statistical devices. (See Figure 3.1)

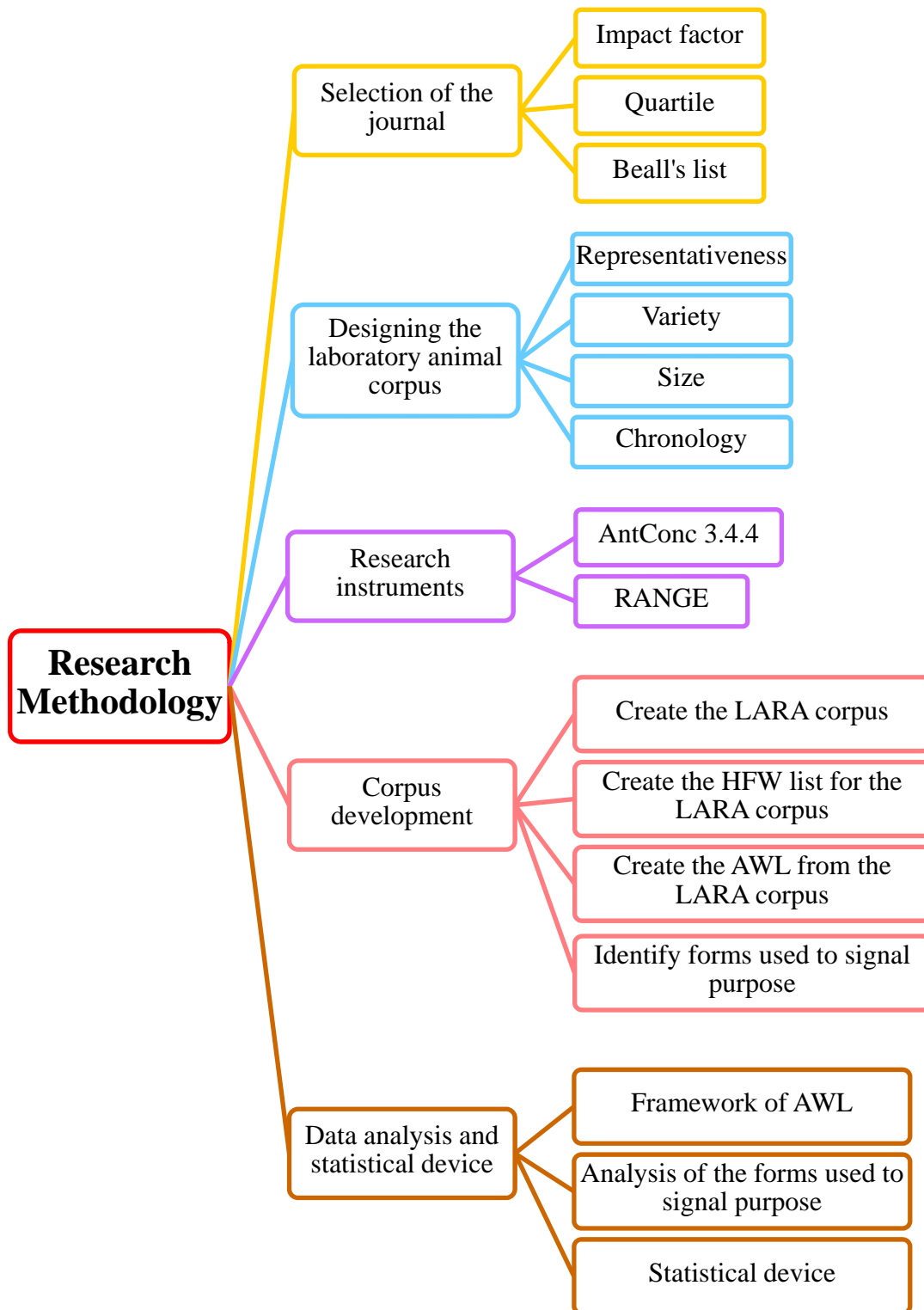


Figure 3.1 Road map of Chapter 3

3.1 Research Procedure

The procedures of a study support and inform the research (Miles & Huberman, 1994; Robson, 2011). The procedures are an important part of designing the study. Therefore, the all procedures used to accomplish the objectives of this study are described as follows.

As a first step, the language input for the laboratory animal review article (hereafter LARA) corpus was prepared. The researcher first interviewed the director of the National Laboratory Animal Center, Mahidol University, in order to select laboratory animal journals. The interview conducted on March 20, 2015.

The second step was to select the laboratory animal journal to be used. The laboratory animal journal was chosen by the ranking of journals in *Journal Citation Reports* and the quartile of the journal was determined by *SCImago Journal Rank*. The researcher selected the *ILAR Journal* and decided as individually saved into text files in Notepad (*.txt) in order to be able to process the text for corpus preparation.

The next step was to create a word list for the LARA corpus by using AntConc Version 3.4.4 (Anthony, 2014). Analyzing the word lists was accomplished by using AntConc Version 3.4.4. This program shows the rank, the frequency, the words, and lemmas. The number of word types, word tokens, and search hits was also indicated. The words shown in the frequency column were arranged according to frequency of appearance in this corpus.

Then the researcher classified words as HFW or AWL following the criteria of Coxhead (2000), including *specialized occurrence*, *range*, and *frequency*. Word families were counted.

Finally, the researcher analyzed language forms and functions. The language forms and functions used in this study were those used to signal purpose (Swales and Feak, 2012 & Soranastaporn, 2013).

All the above-mentioned stages are shown in Figure 3.2 below.

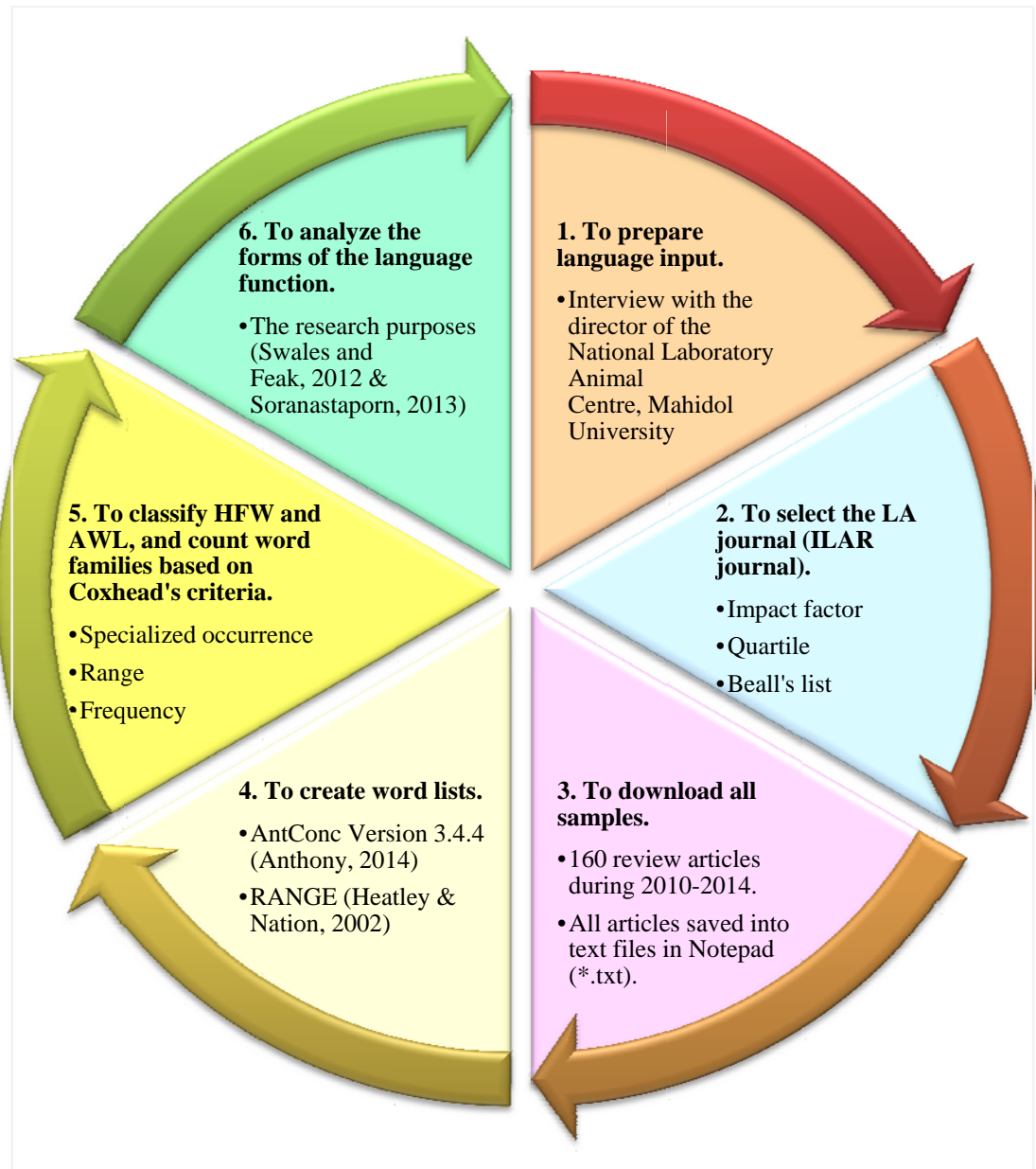


Figure 3.2 Procedures in the current study of the frequency and functions of words from various word lists in LARA corpus

To sum up, there were six main stages in the procedures for this study. The researcher started by preparing language input for the LARA corpus, selecting and a journal by using certain criteria, collecting articles, to classifying words from the articles into categories according to the different word lists used in the study, and analyzing the language forms and functions used to signal purpose.

3.2 Selection of the Journal

The choice of a journal is the most important decision to be made in a study of published scientific articles. The selection of the right journal helps make the results of the study effective. However, selecting a substandard journal can lead to ineffective results. Therefore, finding the appropriate journal for; the analysis of the results in a study can be a difficult process. In this study, impact factor, quartile, and predatory open-access publishers were all criteria to consider the *ILAR Journal* via sources such as *Journal Citation Reports (JCR)*, *SCImago Journal Rank* and *Open-Access Publishing*.

3.2.1 Impact Factor

The first criterion was **Impact Factor**, based on the criteria in *JCR*. The journal impact factor is “a measure of the frequency with which the “average article” in a journal has been cited in a particular year or period (Thomson, 2015).” The impact factor is computed by dividing the number of citations by the total number of articles published in that journal during the two previous years. The *Institute for Laboratory Animal Research (ILAR) Journal* is science journal dedicated to one scientific area. *JCR* as reported by *JCR Science Edition 2014* for the *ILAR Journal* are shown in Table 3.1. The total number of citations for the *ILAR Journal* is 1,180. The impact factor of the journal in 2014 is 2.393. This represents the average number of times the articles in the journal were in 2014. The 5-Year Impact Factor is 2.158. This number represents the average number of times that articles from the *ILAR Journal* published over the past five years have been cited. In short, the *ILAR* journal has in average been cited about more two times in the past two years and in the past five years; in other words, the impact factor of this journal in 2014 is 2.393.

Table 3.1 Data for the *ILAR Journal* as Reported by *JCR Science Edition 2014*

Abbreviated Journal Title (linked to journal information)	ISSN	JCR Data					Cited Half- life
		Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	
<i>ILAR J</i>	1084-2020	1180	<u>2.393</u>	<u>2.158</u>	0.391	46	8.2

3.2.2 The Quartile (Q)

Next, **the quartile (Q)** is based on the criteria in the *SCImago Journal Rank (SJR)*. *SJR* includes “the journals and country scientific indicators developed from the information contained in the Scopus® database” (SCImago, 2007). These indicators can be used to assess and analyze scientific domains. There are four data values for the quartile (Q), consisting of Q1 (highest), Q2 (high), Q3 (low), and Q4 (lowest). The quartile for *ILAR* is presented in *SJR*. The *ILAR Journal* from 1999 to 2014 was usually from the first quartile, or Q1, to the third quartile, or Q3, as shown in Figure 3.3. The data show that the ranked set of the value of the *ILAR Journal* ranges from the highest value to a low value. However, from 2010 to 2014, the ranked set for the value was Q1 to Q2; that is, the *ILAR Journal* was from the highest value to a high value, but the value was mostly high (Q2). In brief, the ranked set for the values for the *ILAR Journal* from Q1 to Q3; that is, the ranked set was from a low value to the highest value. Moreover, from 2010 to 2014, the *ILAR Journal* ranged from the highest value to a high value, but the value was mostly high (Q2). Therefore, the *ILAR Journal*, which is considered rather high-quality (Q2), was selected as the source for the data in the corpus.

ILAR Journal

Country: United States

Subject Area: Agricultural and Biological Sciences | Biochemistry, Genetics and Molecular Biology

Subject Category:

Category	Quartile (Q1 means highest values and Q4 lowest values)															
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Animal Science and Zoology	Q1	Q3	Q3	Q3	Q1	Q2	Q3	Q2	Q1	Q2	Q1	Q2	Q2	Q1	Q2	Q1
Biochemistry, Genetics and Molecular Biology (miscellaneous)	Q1	Q3	Q2	Q2	Q2	Q2	Q2	Q2	Q2	Q2	Q2	Q2	Q2	Q2	Q2	Q2

Publisher: Institute for Laboratory Animal Research. Publication type: Journals. ISSN: 19306180, 10842020

Figure 3.3 The quartile for the *ILAR journal* from 1999 to 2014

3.2.3 Beall's list

In academic publishing, **predatory open access publishing** describes an exploitative open-access publishing business model that involves charging publication fees to authors without providing the editorial and publishing services associated with legitimate journals. "Beall's List", a regularly-updated report by Jeffrey Beall, sets criteria for categorizing predatory publications and lists publishers and independent journals that reflect those criteria (Elliott, 2012). Newer scholars from developing countries are at particular risk of becoming victims of these practices (Kearney, 2015; Xia et al., 2014). When the *ILAR Journal* was searched for via Beall's List, it did not appear. Therefore, this journal is considered reliable for selection of articles for data analysis.

All in all, the *ILAR Journal*, according to the criteria used for selection of the journal, that is, such a journal must have a relatively high impact factor, be in a relatively high quartile (Q), and not be considered a predatory open access publication according to the rankings and criteria of *Journal Citation Reports (JCR)*, *SCImago Journal Rank*, and *Open-Access Publishing* respectively. The results showed that the impact fact of the *ILAR journal* in 2014 was 2.393. Secondly, the journal ranged from the highest quartile to the high quartile from 2010 - 2014. Importantly, the journal did not appear on Beall's list. As a result, the *ILAR Journal's* articles were appropriate for selection for the analysis of the linguistic data; namely, words frequently appearing on various word lists as well as the language forms and functions signaling the objective in this research study.

3.3 Designing the Laboratory Animal Review Article Corpus

The design of the LARA corpus entailed the saving of each text in electronic form, exclusion of references, appendices, footnotes, captions, and acknowledgements as well as the word count. The corpus for this study is the LARA corpus. It was used for analysis in this study. The corpus was chosen to reveal the stylistic characteristics of the corpus itself and to determine which words from which word lists and which language forms and functions inform the LARA were used to

signal purpose. The following criteria were used in the selection of articles forming the corpus: representativeness, variety, corpus size, and chronology.

3.3.1 Representativeness

In terms of representativeness in the field of laboratory animal science, this LARA corpus is the main required area in ESP course, especially, the course of English for veterinarians or English academic course. This provides timely information for all who use, care for, and oversee the use of animals in research. For these reasons, the findings of the research can offer practical results for study at the university level. In other words, the words and the functions are the representative of for the words and functions met in reading and writing academic texts or journal articles in the field of laboratory animal science.

3.3.2 Variety

A variety of authors and topics in LARA were fundamental sources to develop the corpus. In terms of topics, LARA articles includes a number of different articles such as “Progress in Genetics and Genomics of Nonhuman Primates”, “Animal Models of Peripheral Neuropathy”, or “Naturally Occurring Diseases in Animals” in each issue, as shown in Table 3.2. Each article makes use of different vocabulary and forms for signaling the purpose, which means that the current study had access to a diverse collection of vocabulary words and purposive forms. For this reason, it is significant to include language samples for the role model of English for veterinarians or English academic course.

3.3.3 Corpus Size

Because of the purpose of the study, the size of the corpus was not limited. Coxhead (2000) claims that a reasonable number of running words should be compiled to provide a better chance for the occurrence of particular items and to examine more lexical items in depth. Accordingly, 840,773 running words of text were adequate and sensible for the objectives of the recent study; the origin of the data is shown in Table 3.2.

3.3.4 Chronology

As stated by Pearson (1998), the samples of the text in the corpus should have been written or spoken in the ten years prior to the date of compilation. Thus, the Laboratory Animal Corpus in the recent study was compiled of review articles dating from the period of time from 2010 to 2014 in order to avoid inclusion of outdated research articles.

All in all, the criteria for inclusion in the academic corpus of the laboratory animal field firstly included representativeness in this field. The next criterion was that a variety of topics from the laboratory animal field be included. In addition, a corpus of 840,773 running words was selected. Finally, the time frame for articles selected for this corpus ranged from 2010 to 2014.

3.4 Research Instruments

Research instruments are an important factor in the process of gathering information for this current study. Two research instruments were used for analyzing the data in this study: AntConc Version 3.4.4 and RANGE.

3.4.1 AntConc Version 3.4.4

AntConc Version 3.4.4 (Anthony, 2014) was used for data analysis: to count and sort the words in the LARA corpus. This program analyzed a new corpus, which generated a list of all the words in the corpus (Anthony, 2014). This concordancing software was developed by Professor Laurence Anthony of Waseda University, Tokyo, Japan. This easy-to-use and multi-purpose program can run under any Windows environment. The most recent version is AntConc 3.4.4, released in 2014.

3.4.2 RANGE

RANGE analyzed the range of vocabulary in the texts. It is able to (a) determine the vocabulary size necessary to understand the vocabulary in a text, (b) create word lists based on the frequency of occurrence and range of use of vocabulary

in different types of discourse, (c) determine the number of encounters with words in a text, and (d) evaluate the vocabulary load of a text for teaching and learning language. In the recent research, this program was used to analyze the extent of the academic word list.

In summary, two research instruments, AntConc Version 3.4.4 and RANGE were used to create the word lists including the high frequency word list, the academic word list, and the words the outside high-frequency and academic word lists as well as to investigate the forms and functions used to signal purpose.

3.5 Procedures for Analyzing the Corpus

The corpus analysis phase identified words that met the criteria for inclusion in the word lists and the forms and functions used to signal purpose in the corpus. As a result, the phase of corpus analysis consists of the following steps:

- 3.5.1 Create the laboratory animal corpus
- 3.5.2 Create the high-frequency word list for the laboratory animal corpus
- 3.5.3 Create the academic word list from the laboratory animal corpus
- 3.5.4 Create a list of forms used to signal the purpose

3.5.1 Creating the Laboratory Animal Corpus

The review articles were downloaded in order to create the laboratory animal corpus according to the criteria set for corpus compilation. Each text was saved after removing its bibliography, references, appendices, footnotes, captions, and acknowledgements. All files were saved as text files (*.txt) in Notepad because the AntConc cannot read the text from PDF or Word files, only from text files. Five text files were created, each representing one of five years of research articles; they were saved with a file name of *year (volume_ issue)_number of article* such as 2014(55_30)_1. The corpus contained 160 review articles by more than 400 authors, containing 840,773 running words (tokens) and 22,434 word types. The corpus was divided into 16 sub-corpora followed by 16 topics published during 2010 and 2014. as shown in Table 3.2.

Table 3.2 Composition of the Laboratory Animal Review Article Corpus

Published Year	Word Tokens	Word Types	Articles	Topics
2010	32,999	4,581	6	Regenerative Medicine
	43,445	5,466	8	Disaster Planning and Management
	42,103	5,239	10	The Intersection of Humans, Animals, and the Environment
	49,760	5,771	9	Birds as Animal Models in the Behavioral and Neural Sciences
2011	54,478	6,219	13	Animal Models of Aging
	43,368	5,963	8	Spineless Wonders
	50,835	4,940	11	Animal Models of Drug Addictions
2012	69,854	6,026	15	Neurobiology of Addiction-like Behaviors
	62,458	5,953	12	Zebrafish Health and Husbandry
	63,066	5,574	11	Epigenetics
2013	36,508	4,483	7	Ethical and IACUC Considerations for Field Biology Studies
	74,009	5,950	12	Progress in Genetics and Genomics of Nonhuman Primates
2014	35,619	4,449	9	Animal Models of Peripheral Neuropathy
	89,556	7,669	14	Naturally Occurring Diseases in Animals
	76,347	5,762	11	Behavioral Assessment in Animal Models
	16,368	2,150	4	Experimental Design and Statistics
Total	840,773	22,434	160	16

3.5.2 Creating the High-Frequency Word List for the LARA Corpus by Using AntConc

Wordlist is a tool that counts all the words in a corpus and presents them in an ordered list. This tool allows a researcher to find quickly which words are the most frequent in a corpus (Anthony, 2014). In order to create a wordlist for the laboratory animal corpus, the following steps were performed.

3.5.2.1 The researcher clicked on the "Open File(s)" option in the "File" menu and added files for processing as shown in Figure 3.4.

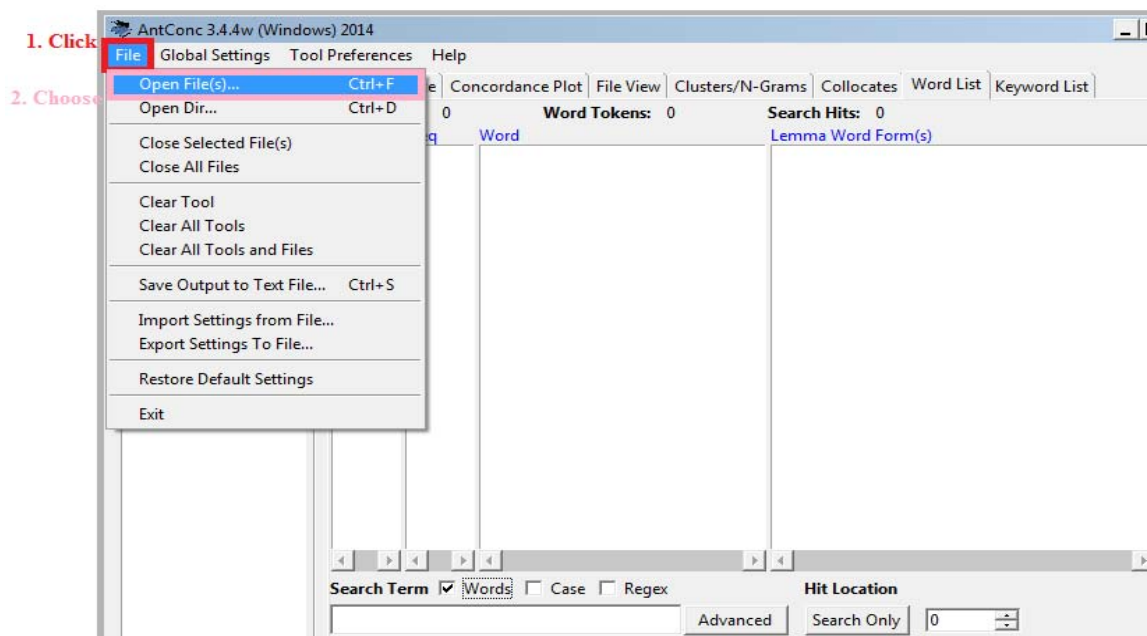


Figure 3.4 Example of adding files for processing

3.5.2.2 The researcher clicked on the "Tool Preferences" menu. Then he choose "Word list" as the "Category". He clicked to choose the "Treat all data as lowercase" option as shown in Figure 3.5. Using the "Treat all data as lowercase" option causes all words to be transformed to lower-case, which it is useful in getting an accurate count of words in certain cases (Anthony, 2014). That is, there is no difference in the count caused by the use of capital letters in some cases. For example, the word "Brain" at the beginning of a sentence and "brain" in the middle of a sentence would be the same. The aim of this study is to focus on lexical units, not syntactic analysis.

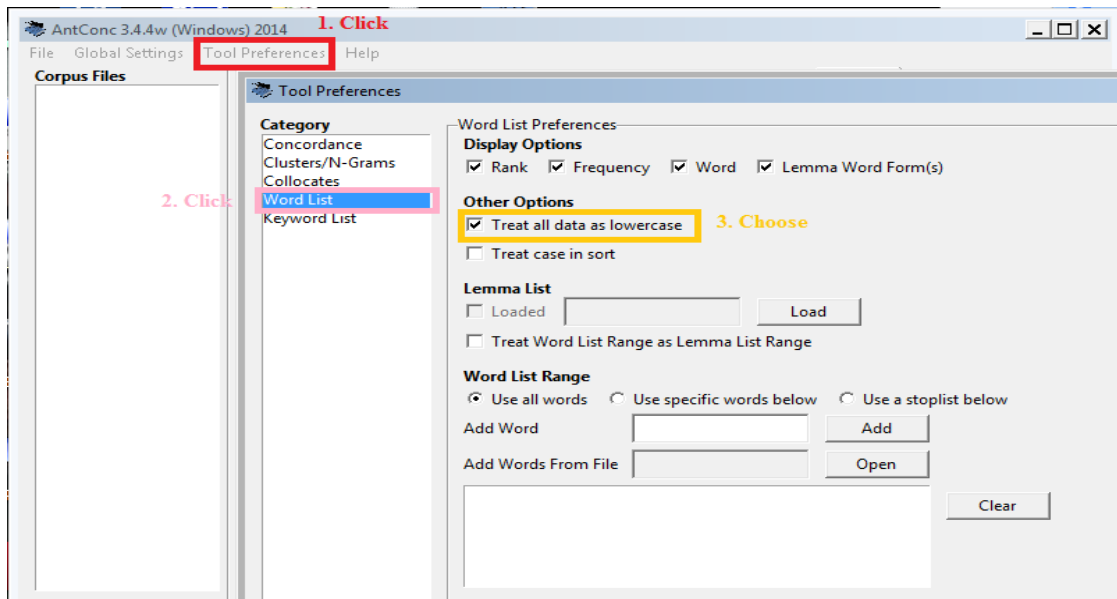


Figure 3.5 Example for clicking on the “Treat all data as lower case” option

3.5.2.3 The researcher pressed the ”start” button, then the word list was sorted by frequency or alphabetical ranking depending on the selection made by the researcher.

3.5.2.4 He clicked on the “Clone Results” button to create a copy of the results so that different sets of results could be compared.

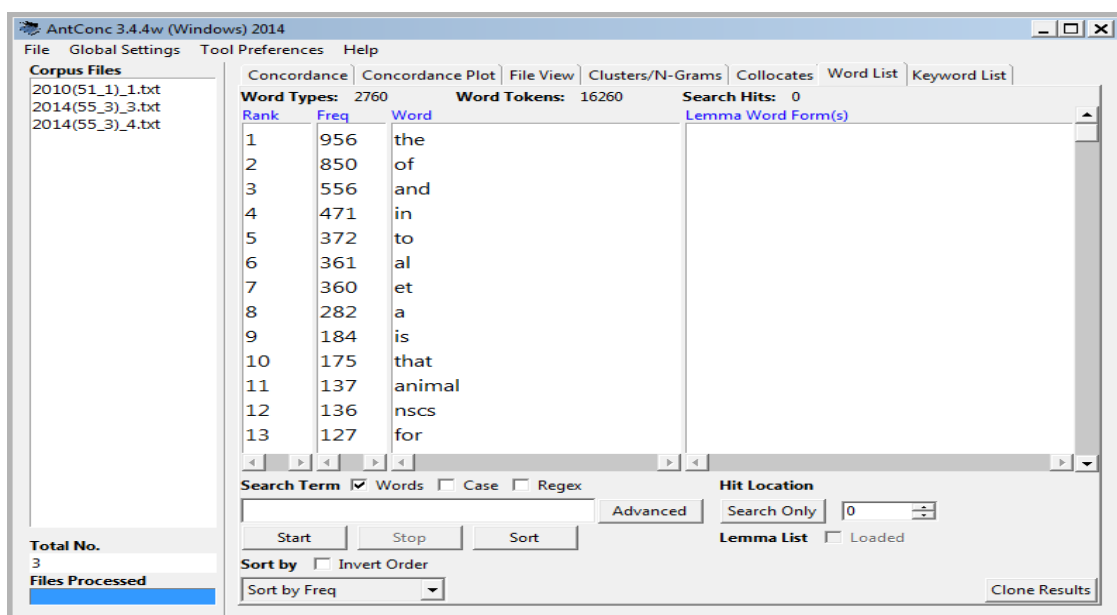


Figure 3.6 Example of word list sorted by frequency

3.5.3 Creating the Academic Word List from the LARA Corpus by Using RANGE

The corpus analysis program RANGE (Heatley & Nation, 2002) was used to count and sort the words in the laboratory animal corpus. This program counts the frequency of words (*frequency*) and records the number of files in which each word occurs in total (*range*). The following are the steps in data analysis.

3.5.3.1 All words in the ILAR articles were entered in .txt files. The files were then entered in RANGE. Then the number of files was selected and *Families* or *Types* was selected in the *Option* menu. Finally, files the files were uploaded. These steps are shown in Figure 3.7.

Range for Texts v.3 Capacity x3 July '15 (8 novels x 125k = 1,000,000 wds/40 secs); column sorting

Upload up to 25 of your own texts (e.g., chapters of a course book, works of an author, frequency lists created by by other Lextutor routines, etc) and find (1) the frequency of each word in the combined collection, (2) the *range* of each word across texts, e.g., the word appears in *these* 6 of your 10 texts, and (3) *NEW* precise grouping and counting with online sorting (by alpha, frequency, VP, or individual text).

[See Prefab Sample Output + Output Filter Demo](#)] **WORK COMPLETE ON THIS ROUTINE 24 JULY 2015** [What's New in v.3?](#)

(1) Save files as TEXT FILES (with *.txt extension) *Explore Range with chapters of [Call of the Wild](#) or [News in French, July '15](#)*

(2) Choose number of files **1**

(3) Browse desktop to load file names into upload spaces below

(4) Enter a title

(5) Options : Unit Families || Types **2**

Language/VP-Frame BNC-Coca || BNC || French v.5

(6) Stoplists (words to ignore) BNC Coca: 1k 2k] or BNC (Eng.): 1k 2k] or French v.5: 1k 2k]

(7) **Submit Files** **3**

01.

02.

03.

04.

Figure 3.7 Steps for uploading files to analyze using RANGE

3.5.3.2 The length of each article was counted in terms of tokens by RANGE

3.5.3.3 The program presented the *range* of the data as shown in Figure 3.8. The figure presents the *range* for each word, arranged by frequency. For example, the first word, 'the' occurred in three texts (T_1, T_2, and T_3), so the *range* of this word is three. Meanwhile, the *range* for the word 'meta' is two because it occurred in two texts (T_1 and T_2).

Home> Range Input [Use «Back to preserve inputs]> Range for Texts - Output
 Filenames, stoplists, offlist, output filtering (freq x range), & Excel-copiable output at bottom
 (new! SORT ROUTINES - SUMMER 2015) VP (tokens/coverage) VP (fams only)

INPUT FILES: 3 | FAMS: 1220 | CLASSIFIABLE TOKENS: 13,544

Click headings to sort ↕

000.	Fams	Freq	Range	BNCoCa	Found in these texts...	T_1	T_2	T_3
000.	the	1044	3	1	T_1	T_2	T_3	
001.	be	735	3	1	T_1	T_2	T_3	
002.	of	647	3	1	T_1	T_2	T_3	
003.	a	404	3	1	T_1	T_2	T_3	
004.	in	359	3	1	T_1	T_2	T_3	
005.	to	322	3	1	T_1	T_2	T_3	
006.	and	320	3	1	T_1	T_2	T_3	
007.	study	248	3	1	T_1	T_2	T_3	
008.	animal	211	3	1	T_1	T_2	T_3	
009.	that	185	3	1	T_1	T_2	T_3	
010.	analyse	152	3	3	T_1	T_2	T_3	
011.	this	143	3	1	T_1	T_2	T_3	
012.	for	136	3	1	T_1	T_2	T_3	
013.	experiment	130	3	3	T_1	T_2	T_3	
014.	effect	126	3	2	T_1	T_2	T_3	
015.	result	116	3	2	T_1	T_2	T_3	
016.	meta	109	2	17	T_1	T_2		
017.	use	91	3	1	T_1	T_2	T_3	
018.	or	90	3	1	T_1	T_2	T_3	
019.	research	86	3	2	T_1	T_2	T_3	
020.	as	82	3	1	T_1	T_2	T_3	

Figure 3.8 Example of the *range* as determined by RANGE

In brief, RANGE was used to analyze the *range* of words in the corpus following the steps mentioned above. In addition, this program shows the *frequency*.

3.5.4 Establishing the Forms and Functions for Research Purposes

To investigate the forms and functions of language, sentences used to the purpose of the research were collected and analyzed using Microsoft Office Excel 2007. Each sentence was analyzed according to words, verb tenses, and the forms. Therefore, words signaling research purpose were investigated and matched with the words on the word lists of the current corpus. Moreover, verb tenses were analyzed to establish tense usage in the purpose sentences. Finally, the purpose patterns were

studied, another useful step in studying purpose statements. The following steps were performed, as shown in the Figure 3.9.

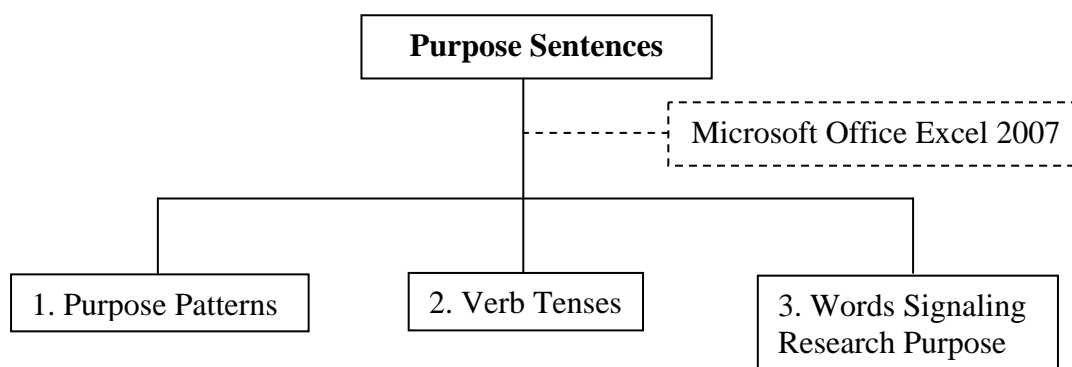


Figure 3.9 Chart detailing steps in the investigation of the forms and functions used to signal research purpose

Firstly, all sentences in the articles which described the objective of the research were collected. These purpose sentences were in Introduction or Abstract sections of the articles. After these sentences had been collected in and counted by Microsoft Office Excel 2007, the purpose sentences were coded according to categories such as *volume no. issue year* such as *51 1 1 2010* so that they could be retrieved and re-checked easily.

Second, the patterns used in the purpose sentences were analyzed. Then items exhibiting the same pattern were grouped and the number and percentage of use for each pattern were calculated.

Third, the verb tenses used in the purpose sentences were analyzed. Then the number and percentage of used for each verb tense were calculated.

Fourth, the words signaling research purposes were analyzed. Then these words were matched with words on the word lists of the current corpus. Then the number and percentage of use of words from each word list were calculated.

Finally, the results involving purpose sentences and the patterns they used were checked by a native speaker who is an expert in linguistics.

In short, the sentences presenting the research purposes in the LARA corpus were analyzed by forms, verb tenses, and word use. Therefore, forms, verb tenses, and word signaling the research purposes were analyzed, and the number and

the percentage of use for each were also calculated. Additionally, to ensure the correctness of the analysis, a native speaker carefully checked the purpose sentences and patterns.

3.6 Data Analysis and Statistical Devices

Two programs, AntConc 3.4.4 and RANGE were used to determine the number of, tokens in the raw data and to determine word families. Then the AWL and descriptors of the forms and functions used to signal research purposes were used to analyze the raw data. The AWL data were calculated using statistics. After that the data for the forms and functions used to signal purposes were checked by a native speaker to ensure the correctness of the analysis. The data analysis then comprises the framework for the AWL, analysis of the forms used to signal the purpose, and statistical devices.

3.6.1 Framework for the AWL

In framework for the AWL, three of Coxhead’s criteria (2000) including *specialized occurrence*, *range*, and *frequency* were used, as shown in Figure 3.10; in addition, the limitations imposed by division into word families were used in analyzing the words on the word lists.

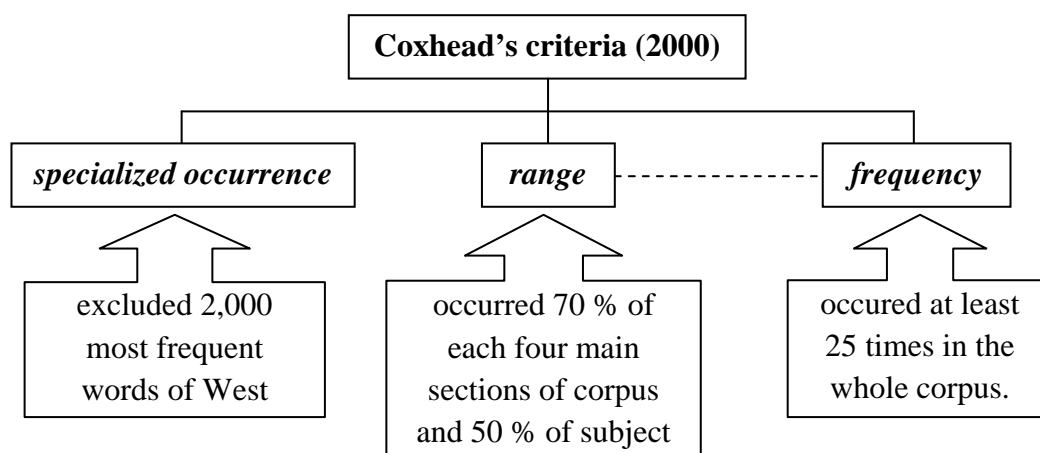


Figure 3.10 Framework for AWL according to Coxhead’s criteria

First, *specialized occurrence* means the 2,000 most frequent words (as determined by using word families) from West (1953) are excluded.

Secondly, the *range* followed by Coxhead (2000) is that a member of a word family occurs at least 10 times in each four disciplines of corpus and in 15 or more of the 28 subject areas. (See in Table 3.3). In other words, a member of a word family appears 70 % of each four main sections of corpus and 50 % of subject areas.

Table 3.3 Composition of Academic Corpus

	Discipline				
	Arts	Commerce	Law	Science	Total
running words	883,214	879,547	874,723	875,846	351,333
Texts	122	107	72	113	414
Subject area	Education	Accounting	Constitutional	Biology	
	History	Economics	Criminal	Chemistry	
	Linguistics	Finance	Family and	Computer	
	Philosophy	Industrial	medicolegal	science	
	Politics	relations	International	Geography	
	Psychology	Management	Pure	Geology	
	Sociology	Marketing	commercial	Mathematics	
		Public policy	Quasi-	Physics	
			commercial		
			Rights and		
			remedies		

(Coxhead, 2000: 220)

Consequently, in the LARA corpus, a member of a word family occurred at least three times in year 2010 sections of corpus, two times in years 2011 and 2012, one time in year 2013, and three times in year 2014 respectively, shown in Table 3.4 and in 8 or more of the 16 topics.

Table 3.4 Criteria for Word-Family Distribution in the LARA Corpus

Main sections (year / issues)	Hit (times)
2010 (4 issues)	at least 3
2011 (3 issues)	at least 2
2012 (3 issues)	at least 2
2013 (2 issues)	at least 1
2014 (4 issues)	at least 3

Thirdly, *frequency* requires that a member of a word family occur at least 100 times in the corpus. The *frequency* is considered secondary to the *range* because a word count based mainly on frequency would bias by longer texts and topic-related words. Furthermore, the frequency of 100 times is applied strictly for multiple-member word families because single-member families operate at a disadvantage in gaining a high frequency of occurrence. As a result, the *frequency* becomes almost 28.5 times in a million words. Nonetheless, the LARA corpus consists of 881,621 running words, so members of word families had to occur at least 25 times in the whole corpus in order to be ranked for frequency.

In the corpus, the *word family* with only one member which occurs the least frequently is forthcoming. A word family includes a headword, its inflected forms, and its closely related derived form. For example, analyze, analyzed, analyzes, analyzer, analyzable, analysis, and analyses are all in the same word family.

In brief, three criteria (*specialized occurrence*, *range*, and *frequency*) as well as word families were used for analyzing the words appearing on the academic word list using AntConc 3.4.4 and RANGE. Furthermore, with the three criteria, word lists consisting of words from the GSL and AWL were generated.

3.6.2 Analyzing the Forms of the Purpose Functions

To determine which words and forms indicated objectives in LAC, the frameworks of Swales and Feak (2012) and Soranastaporn (2013) were used. Each pattern contains a main verb, and these verbs appear in different tenses. Thus, the current research analyzed the verb tenses in the purpose sentences using the

description of Swales and Feak (2012). Importantly, words are considered as the key factor in forming the sentences, so it is essential to study the words presenting the research purposes using the word lists in the current research. The analysis of purpose forms and functions is shown in Figure 3.11.

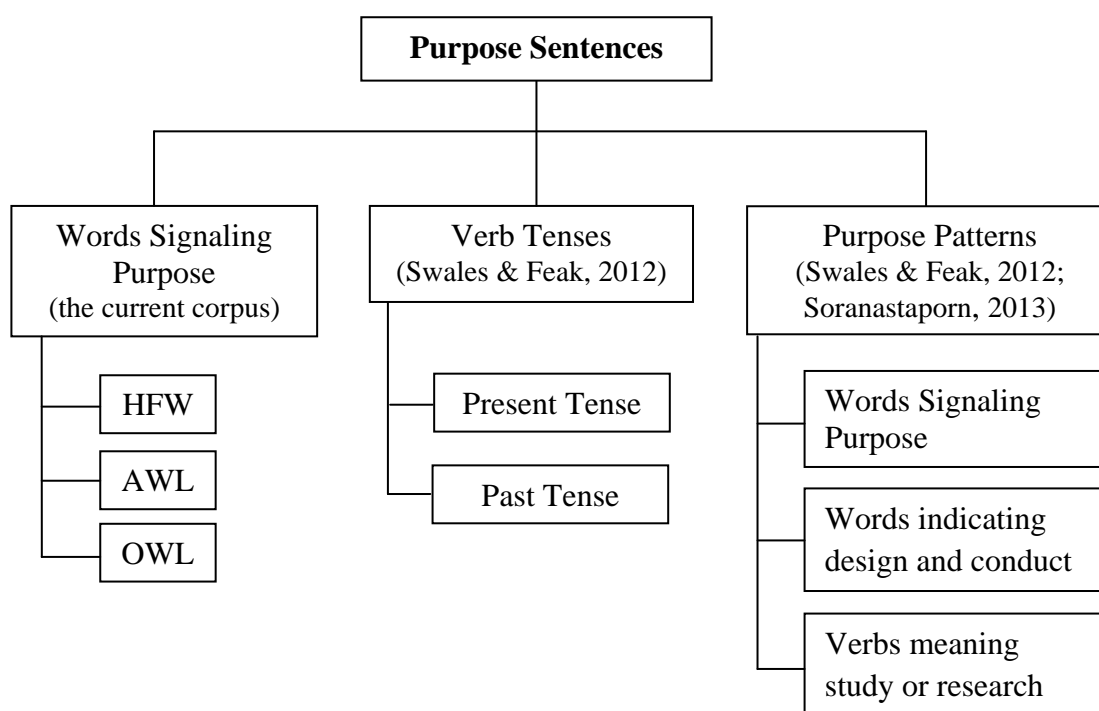


Figure 3.11 Studying the forms and functions used to signal research purposes

3.6.3 Statistical Devices

Descriptive statistics were used in data analysis. Descriptive statistics were used to determine frequency and percentages.

In short, the criteria of Coxhead (2000) were used to analyze words on the academic word list. Furthermore, a list of high-frequency words was generated by the criteria. Analysis of the purpose sentences determined words, verb tenses, and forms used to express research purposes using the description of Swales and Feak (2012) and Soranastaporn (2013). Finally, statistics were used to determine frequency and percentage.

Chapter Summary

To conclude, this research methodology section contains an explanation of the steps used in collection and analysis of the data. The first step in this study was the selection of the journal, which considered the impact factor, the quartile (Q), and predatory open access publishing. Moreover, the criteria were used in the selection of articles forming the corpus consisting of representativeness, variety, corpus size, and chronology. Also, research instruments used for analyzing the data in this study were AntConc Version 3.4.4, RANGE, and Microsoft Office Excel 2007. The laboratory animal review article corpus between 2010 and 2014 consists of 840,773 running words and 22,434 word types from 16 topics in 160 articles. The corpus created the high-frequent word list, the academic word list, and the words outside high-frequency word list and academic word list. AntConc 3.4.4 was used to create the high-frequency word list for LARA Corpus. Moreover, RANGE was used to analyze the *range* of the words in the corpus. In the LARA corpus, forms and functions of the research purposes were investigated via Microsoft Office Excel 2007. Finally, data of the word lists were analyzed by the criteria of Coxhead (2000), and the purpose sentences with the word use were analyzed by the framework of Swales and Feak (2012) and Soranastaporn (2013). Furthermore, statistical analysis of the data established frequency and percentage.

CHAPTER IV

RESULTS

In this chapter, the results and findings of the research are reported by data analysis and computation based on the research objectives and questions set at the beginning of the study. The results showed the top-100 most occurred words used in review articles of laboratory animal (LARA) corpus, the content words of AWL occurring with high frequency in the same corpus, and the words and the patterns which laboratory animal researchers use to describe their research purposes the most frequently with the following headings:

- 4.1 The high-frequency words
- 4.2 The content words of AWL
- 4.3 The words and patterns describing the research objectives

4.1 The High-Frequency Words

Research question 1: What hundred content words appear most frequently in laboratory animal review articles?

The statistical details of LARA corpus include the numbers and the percentages of word tokens, word types, word families, and word family types, as shown in Table 1. All of 160 research articles between 2010 and 2014 were published and written in English by one or more native and non-native speaker(s) of English as mentioned in Chapter III. The LARA corpus included 840,773 word tokens and 22,434 word types. The word tokens were the highest frequency in the statistical details since every word was counted in this corpus. Moreover, the token words were divided into two categories; namely, function words and content words. The function words containing 371,213 words (44.15%) occurred less than the content words

containing 469,560 words (55.85%). Additionally, there were 19,920 word family types (2.37%), which were less than word types about one time.

Table 4.1 Statistical Details of LARA Corpus

Statistics	Number	%¹
Word tokens (running words)	840,773	100
Word types	22,434	2.67
Tokens of function words	371,213	44.15
Tokens of families of content words	469,560	55.85
Types of families of content words	19,920	2.37

¹ % of Text Coverage

Moreover, Table 4.2 presents the statistics details of high-frequency words (HFW) in LARA corpus. The first 1,000 high-frequency words consisted of 623,583 word tokens (74.17%) in LARA corpus. In the meanwhile, the second 1,000 high-frequency words consisted of 79,713 word tokens (9.48%). That is to say, the number of word tokens of the first 1,000 words appeared more than the second 1,000 words.

Table 4.2 Statistical Details of High-Frequency Words in LARA Corpus

High-frequency words	Word tokens	%¹
The first 1,000 words	623,583	74.17
The second 1,000 words	79,713	9.48
Total	703,296	83.65

¹% of Text Coverage

According to Figure 4.1, the proportions of the highest-frequency words are shown. In other words, the bar graph displays the percentage of the content words, the function words, and the total of words in the high-frequency words. Firstly, the number of content words of the high-frequency words (40.13%) was lower than the function words (43.52%). That is to say, the content words in the first 1,000 high-frequency words (31.02%) were lower than the function words (43.15%). In contrast,

the content words of the second 1,000 high-frequency words (9.11%) were more than the function words (0.37%).

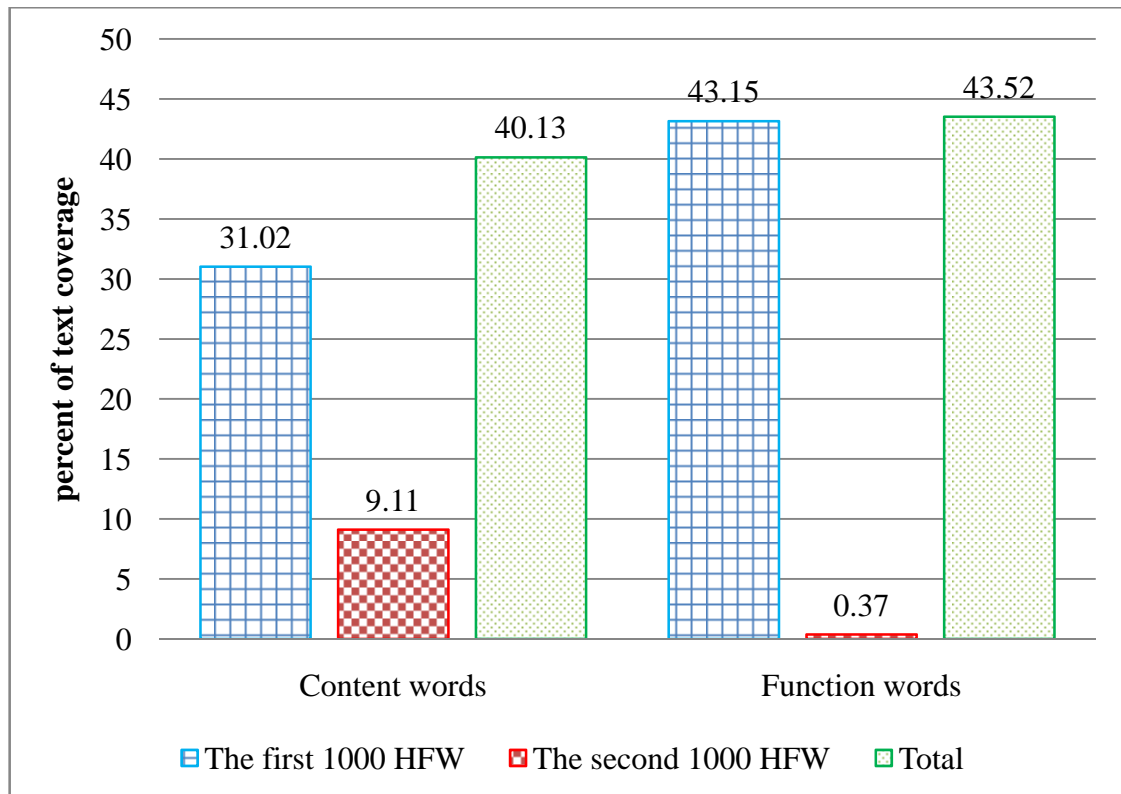


Figure 4.1 Proportions of function words and content words in the high-frequency words

According to the findings by AntConc, it was found that the function words occurred in the first twenty-three rank of the tokens, as shown in Figure 4.2. The word *'the'* was the most frequently found (5.41% of text coverage). The second function word was *'of'* (4.38% of text coverage). The third one was *'and'* (3.64% of text coverage). The others of the first twenty-three words were the words *'in'*, *'to'*, *'a'*, *'that'*, *'is'*, *'for'*, *'with'*, *'are'*, *'as'*, *'or'*, *'be'*, *'by'*, *'on'*, *'have'*, *'this'*, *'from'*, *'an'*, *'not'*, and *'these'* respectively. However, since the objective of the current research was to investigate the content words of the review articles of laboratory animal, the function words were eliminated from the word list in the corpus.

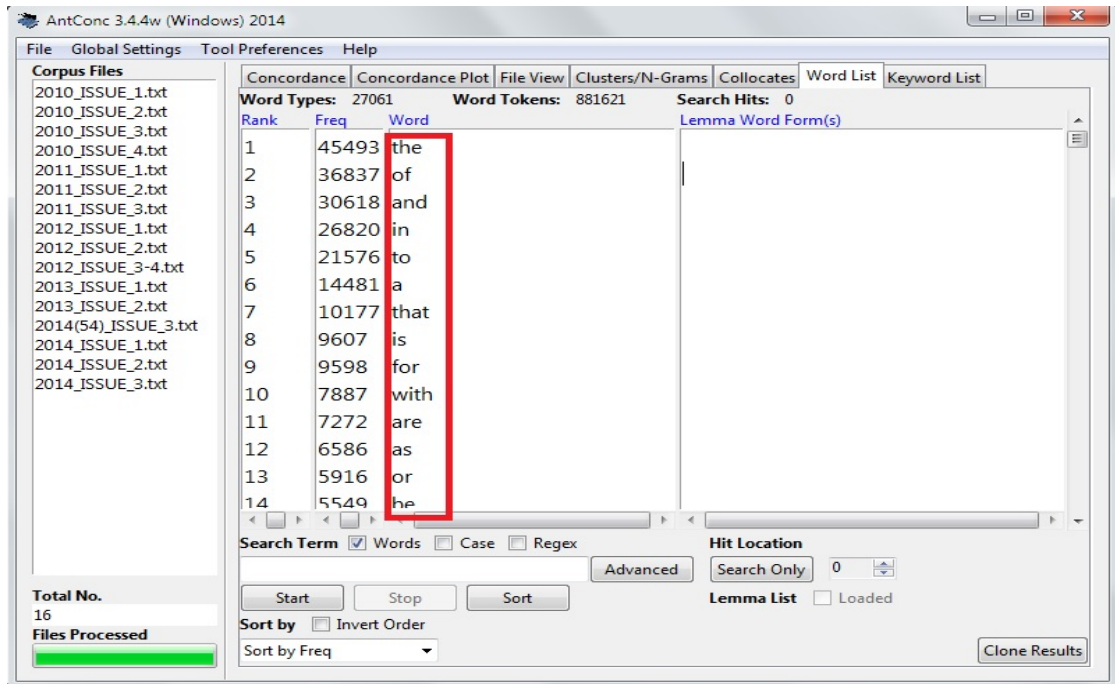


Figure 4.2 The most occurred function words in the high-frequency words

In addition to the first twenty-three tokens, the result appeared the first content word, which was the twenty-fourth rank of the corpus, as shown in Figure 4.3. The first content word was the word ‘*studies*’, which covered 0.36% of text coverage.

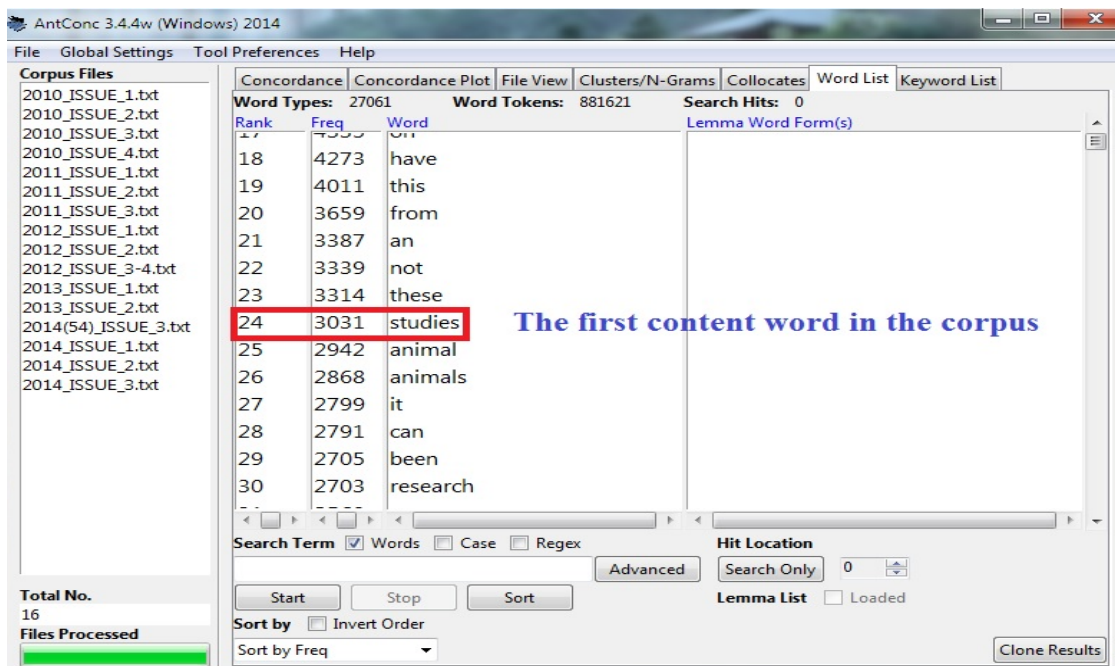


Figure 4.3 The first content words found in the LARA corpus

Moreover, Table 4.3 illustrates the top-100 most occurred content words used in LARA corpus, which calculated as 11.03% of text coverage.

Table 4.3 Top-100 Most Occurred Content Words Used in LARA Corpus

Rank	Words	f	% ¹	Rank	Words	f	% ¹
1	studies	3031	0.36	51	brain	761	0.09
2	animal	2942	0.35	52	behavioral	754	0.09
3	animals	2868	0.34	53	similar	754	0.09
4	research	2703	0.32	54	drug	748	0.09
5	human	2252	0.27	55	related	743	0.09
6	species	1954	0.23	56	data	742	0.09
7	used	1646	0.20	57	new	735	0.09
8	models	1596	0.19	58	control	729	0.09
9	study	1571	0.19	59	based	711	0.08
10	use	1536	0.18	60	risk	707	0.08
11	humans	1529	0.18	61	found	702	0.08
12	model	1507	0.18	62	social	696	0.08
13	disease	1488	0.18	63	significant	694	0.08
14	cells	1355	0.16	64	early	693	0.08
15	effects	1318	0.16	65	factors	685	0.08
16	mice	1195	0.14	66	differences	665	0.08
17	development	1177	0.14	67	group	662	0.08
18	cell	1139	0.14	68	often	658	0.08
19	associated	1117	0.13	69	reported	657	0.08
20	genetic	1054	0.13	70	potential	646	0.08
21	well	1041	0.12	71	cancer	635	0.08
22	gene	1038	0.12	72	long	613	0.07
23	specific	1030	0.12	73	years	602	0.07
24	behavior	994	0.12	74	age	598	0.07
25	levels	989	0.12	75	infection	595	0.07
26	expression	978	0.12	76	fish	594	0.07
27	response	942	0.11	77	food	591	0.07
28	stress	934	0.11	78	effect	590	0.07
29	time	923	0.11	79	include	578	0.07
30	important	910	0.11	80	large	575	0.07
31	using	907	0.11	81	role	574	0.07
32	rats	906	0.11	82	likely	571	0.07
33	exposure	902	0.11	83	same	569	0.07
34	dogs	894	0.11	84	common	561	0.07
35	including	879	0.10	85	monkeys	561	0.07
36	genes	867	0.10	86	macaques	558	0.07
37	water	865	0.10	87	addition	557	0.07
38	zebrafish	860	0.10	88	activity	556	0.07
39	different	857	0.10	89	diseases	555	0.07

Table 4.3 Top-100 Most Occurred Content Words Used in LARA corpus(Cont.)

Rank	Words	f	% ¹	Rank	Words	f	% ¹
40	health	855	0.10	90	provide	554	0.07
41	increased	840	0.10	91	responses	552	0.07
42	results	825	0.10	92	experimental	550	0.07
43	changes	824	0.10	93	methylation	547	0.07
44	high	802	0.10	94	information	544	0.06
45	laboratory	802	0.10	95	induced	538	0.06
46	system	802	0.10	96	pain	536	0.06
47	clinical	798	0.09	97	environmental	535	0.06
48	number	791	0.09	98	compared	534	0.06
49	example	782	0.09	99	analysis	524	0.06
50	treatment	781	0.09	100	systems	522	0.06
TOTAL						92,707	11.03

¹% of Text Coverage

According to Table 4.3, the content word '*studies*' was found the most in the corpus, which occurred 3,031 times calculated as 0.36 % of text coverage. The 100th occurred word '*systems*', occurring 535 times calculated as 0.06 % of text coverage. That is to say, the first content word was five times as frequent as the 100th one. The top-5 high-frequency content word families were '*studies*', '*animal*', '*animals*', '*research*', and '*human*' which occurred 3,031 (0.36%), 2,942 (0.35%), 2,868 (0.34%), 2,703 (0.32%), and 2,252 (0.27%) times respectively.

In short, the current study found 840,773 word tokens and 22,434 word types in LARAC. All of the word tokens consisted of 371,213 function words and 469,560 content words. Moreover, there were 703,296 high-frequency words (83.65% of text coverage), which was composed of 365,934 function words (43.52% of text coverage) and 337,362 content words (40.13% of text coverage). The function word occurring the most was the word '*the*' (5.41% of text coverage). Nevertheless, because of the objective of the study, the function words were gotten rid of. In terms of the content words, the word '*studies*' was found the most in this corpus (0.36% of text coverage), The word '*studies*' appeared at the twenty-fourth rank of the token words, and the 100th content word was the word '*system*' occurring 522 times (0.06% of the text coverage).

4.2 The Content Words of AWL

Research question 2: What content words from the AWL occurred with high frequency in the same corpus of laboratory animal review articles?

With the criteria of Coxhead (2000), 1) *specialized occurrence* is the exclusion of the 2,000 most frequent word families in the corpus. 2) *Range* is that the member of a word family occurs at least three times in year 2010 sections of corpus, two times in years 2011 and 2012, one time in year 2013, and three times in year 2014 respectively, as mentioned in Chapter III and in eight or more of the 16 topics of the corpus. Finally, 3) *Frequency* is that the members of word families had to occur at least 25 times in the whole corpus. Therefore, the detailed research results of the content words of AW are reported with the statistical details of AWL and the headwords of content words of AWL occurred with high frequency in LARA corpus in Table 4.4 and 4.5 below. Furthermore, the 100-top highest-frequency academic words in the corpus are displayed in Table 4.6 as follows.

Firstly, the statistical details of AWL in LARA corpus are shown in Table 4.4, which consists of the frequency and the percent of word tokens, headwords of content word families, and word tokens of headwords of content word families. The number of academic words was 36,658 word tokens. In other words, the text coverage (4.36%) was lower than word families (7.81 %) about a half. The number of academic word types was 1,026 words. That is to say, the text coverage (0.12 %) was lower than word families (0.22 %) about a half.

Table 4.4 Statistical Details of AWL in LARA Corpus

Academic Words List (AWL)	<i>f</i>	% of text coverage	% of word families
Word Tokens	36,658	4.36	7.81
Word Types	1,026	0.12	0.22

In addition to the statistical details of AWL, the statistical information of headwords of academic words is displayed in Table 4.5. The headwords of the families of content words consisted of 27,656 token words. They covered 3.29 %, and the word families covered 5.89%. The LARA corpus also comprised 768 types of

headwords of content word families (seen in Appendix C). In other words, the text coverage (0.09 %) was lower than word families (0.16 %) about a half.

Table 4.5 Statistical Details of Headwords of AWL in LARA Corpus

Headwords of Academic Words	f	% of text coverage	% of word families
Word Tokens	27,656	3.29	5.89
Word Types	768	0.09	0.16

Furthermore, the top-100 most frequent headwords of academic words in the corpus displays in the Table 4.6. The total of the top-100 highest headwords comprised 5,055 tokens (0.601%). The words '*regardless*' and '*surrounding*' were the highest-frequency of the academic words (53 times), while the last three of the first 100 academic words were '*actually*', '*agriculture*', and '*amounts*' (47 times).

Table 4.6 Top-100 Most Occurred Headwords of AWL in LARA Corpus

Rank	Headwords of Academic Words	f	%¹	Rank	Headwords of Academic Words	f	%¹
1	regardless	53	0.006	51	profound	50	0.006
2	surrounding	53	0.006	52	publication	50	0.006
3	adjacent	52	0.006	53	reasonable	50	0.006
4	consider	52	0.006	54	rise	50	0.006
5	corresponding	52	0.006	55	separated	50	0.006
6	diagnose	52	0.006	56	threshold	50	0.006
7	dopaminergic	52	0.006	57	timing	50	0.006
8	drive	52	0.006	58	underlie	50	0.006
9	ethics	52	0.006	59	wall	50	0.006
10	explore	52	0.006	60	amino	49	0.006
11	eye	52	0.006	61	aware	49	0.006
12	face	52	0.006	62	beneficial	49	0.006
13	feces	52	0.006	63	careful	49	0.006
14	institute	52	0.006	64	cognition	49	0.006
15	intravenous	52	0.006	65	combinations	49	0.006
16	livestock	52	0.006	66	conducting	49	0.006
17	mate	52	0.006	67	contributes	49	0.006

Table 4.6 Top-100 Most Occurred Headwords of AWL in LARA Corpus (Cont.)

Rank	Headwords of Academic Words	f	% ¹	Rank	Headwords of Academic Words	f	% ¹
18	meet	52	0.006	68	depth	49	0.006
19	newly	52	0.006	69	dosing	49	0.006
20	<u>predicted</u>	52	0.006	70	ear	49	0.006
21	remodeling	52	0.006	71	<u>eventually</u>	49	0.006
22	step	52	0.006	72	genus	49	0.006
23	advance	51	0.006	73	illness	49	0.006
24	begin	51	0.006	74	improving	49	0.006
25	brain	51	0.006	75	lives	49	0.006
26	certainly	51	0.006	76	opposite	49	0.006
27	change	51	0.006	77	parent	49	0.006
28	class	51	0.006	78	permanent	49	0.006
29	<u>consist</u>	51	0.006	79	presents	49	0.006
30	<u>domesticated</u>	51	0.006	80	reductions	49	0.006
31	exercise	51	0.006	81	storage	49	0.006
32	influence	51	0.006	82	tract	49	0.006
33	lacking	51	0.006	83	<u>analogous</u>	48	0.006
34	marker	51	0.006	84	<u>creation</u>	48	0.006
35	mechanistic	51	0.006	85	department	48	0.006
36	morbidity	51	0.006	86	<u>error</u>	48	0.006
37	office	51	0.006	87	industry	48	0.006
38	seem	51	0.006	88	<u>inhibit</u>	48	0.006
39	sodium	51	0.006	89	lipid	48	0.006
40	subset	51	0.006	90	maturation	48	0.006
41	temporal	51	0.006	91	modern	48	0.006
42	chain	50	0.006	92	obese	48	0.006
43	greatest	50	0.006	93	paper	48	0.006
44	<u>input</u>	50	0.006	94	<u>professional</u>	48	0.006
45	likelihood	50	0.006	95	prominent	48	0.006
46	<u>link</u>	50	0.006	96	thermal	48	0.006
47	mimic	50	0.006	97	threats	48	0.006
48	<u>minor</u>	50	0.006	98	actually	47	0.006
49	painful	50	0.006	99	agriculture	47	0.006
50	product	50	0.006	100	amounts	47	0.006
TOTAL						5,055	0.601

¹% of text coverage

Bold and underlined words are AWL of Coxhead (2000)

4.3 The Words and Patterns Describing the Research Purposes

Research question 3: What words and forms signal the function of indicating research purpose in laboratory animal studies?

According 160 review papers of laboratory animal, 133 purpose sentences was found. Each sentence which presents the purposes consisted of difference words and patterns. In other words, each pattern was formed by the different words in order to make the sentences convey the meaning of the purposes. Theoretically, verbs, one word class of English, are central to sentences. Verb tenses are one of the grammatical categories for verbs. The section of results; therefore, displays the results of content words, verb tenses, and patterns describing the research purposes in the corpus as followings:

4.3.1 Content Words

In terms of the content words, the laboratory animal researchers consisted of 68 content words describing the research purposes. The list of content words which occurred the most frequently to describe the research purposes contained 45 content words in high-frequency word list (79.46% of words describing the research purposes), while AWL included 15 content words (16.67% of words describing the research purposes), which occurred lower than in high-frequency word list. However, words outside high-frequency word lists and academic word list (OWL) were used to describe the research purposes the least. It contained eight content words (3.87 % of words describing the research purposes), as shown in Figure 4.4.

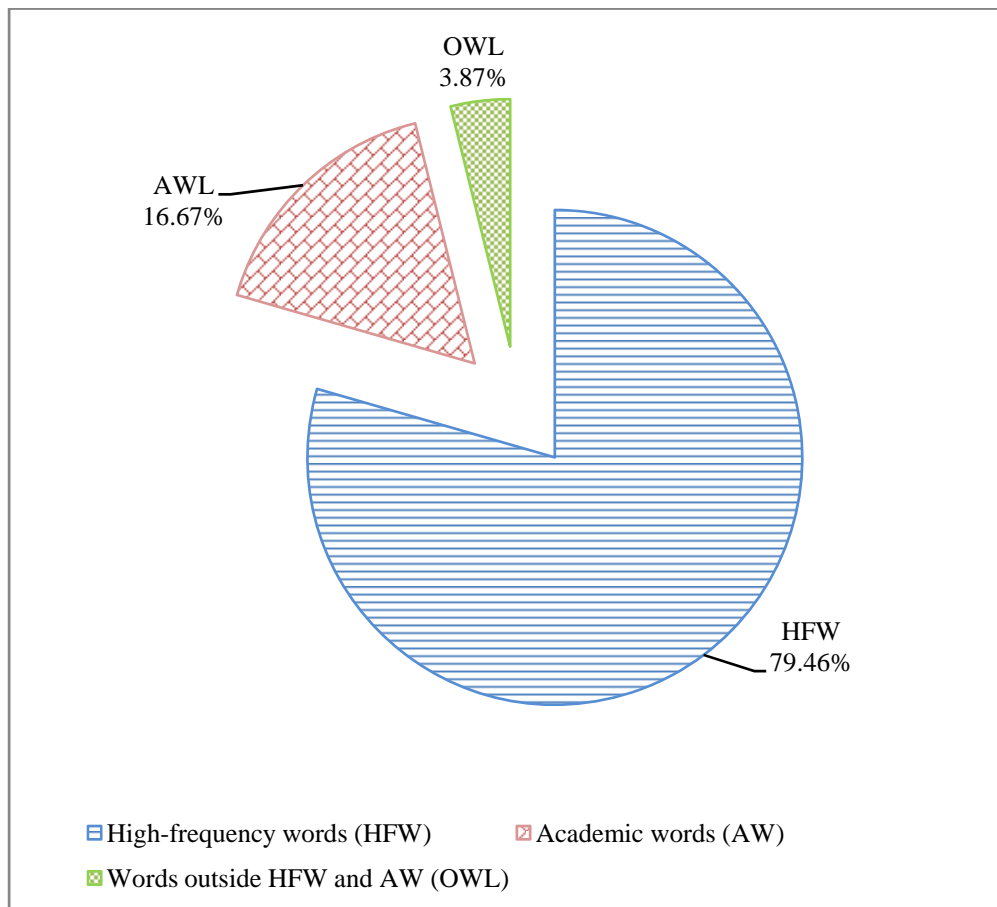


Figure 4.4 Percentage of word lists describing the research purposes

The content words in high-frequency word list which were used to describe the research purposes included 45 content words (79.46%), as shown in Table 4.7. The word 'review' occurred the first most frequently (22.02%). The word occurring the second most frequently was 'article' (13.69%). The third most frequent word was 'discuss' (6.25%). Interestingly, the first most frequent word occurred higher than the second one about 1.6 times, while the third most frequent word occurred about 2.2 times as high as the second word.

Table 4.7 Content Words Describing the Research Purposes in High-Frequency Word List

Rank	HFW	f	% ¹	Rank	HFW	f	% ¹	Rank	HFW	f	% ¹
1	<i>review</i>	74	22.02	16	<i>show</i>	3	0.89	31	<i>gain</i>	1	0.30
2	<i>article</i>	46	13.69	17	<i>use</i>	3	0.89	32	<i>give</i>	1	0.30
3	<u>discuss</u>	21	6.25	18	<i>address</i>	2	0.60	33	<i>important</i>	1	0.30
4	<u>provide</u>	17	5.06	19	<u>investigate</u>	2	0.60	34	<i>literature</i>	1	0.30
5	<u>focus</u>	14	4.17	20	<i>propose</i>	2	0.60	35	<i>main</i>	1	0.30
6	<u>present</u>	14	4.17	21	<i>act</i>	1	0.30	36	<i>mean</i>	1	0.30
7	<i>describe</i>	8	2.38	22	<i>allow</i>	1	0.30	37	method	1	0.30
8	<u>goal</u>	7	2.08	23	assess	1	0.30	38	<i>offer</i>	1	0.30
9	<u>identify</u>	5	1.49	24	brief	1	0.30	39	procedure	1	0.30
10	<u>purpose</u>	5	1.49	25	<i>broad</i>	1	0.30	40	promote	1	0.30
11	<u>study</u>	5	1.49	26	<u>compared</u>	1	0.30	41	<i>reason</i>	1	0.30
12	overview	4	1.19	27	<u>design</u>	1	0.30	42	<u>research</u>	1	0.30
13	<u>examine</u>	4	1.19	28	<u>determine</u>	1	0.30	43	resource	1	0.30
14	<i>intend</i>	3	0.89	29	<i>end</i>	1	0.30	44	specially	1	0.30
15	<u>report</u>	3	0.89	30	<u>evaluate</u>	1	0.30	45	<i>understand</i>	1	0.30
TOTAL										267	79.46

¹% of words describing the research purposes in LARA corpus

bold and italic words = words consistent with West's GSL (1953)

underlined words = words consistent with examples in textbooks

Regarding the content words in AWL, academic words which were used to describe the research purposes consisted of 15 content words, as shown in Figure 4.5. The words, 'aim' and 'highlight', were used the most frequently (3.27%). The word used the second most frequently was 'summarize'. The third most frequent word was 'briefly'. The percentage of the rest of content words gradually decreased. Interestingly, the word 'briefly' was only one adverb occurring in this list. This word was applied to modify the main verb.

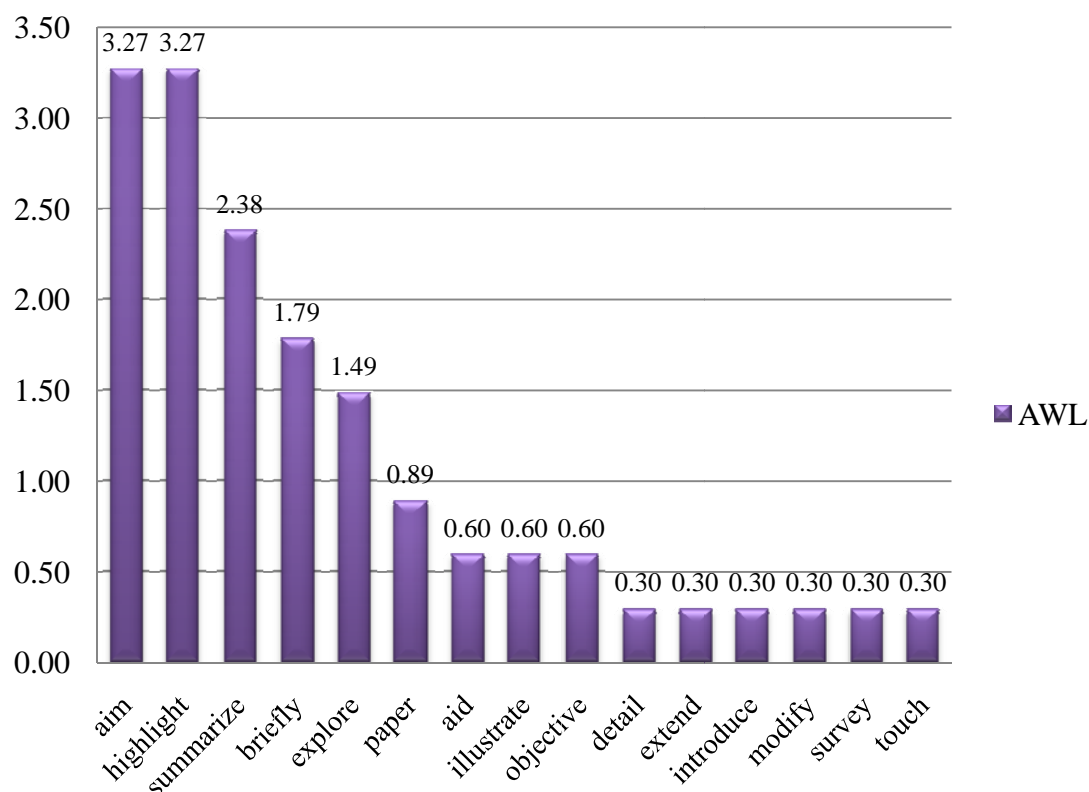


Figure 4.5 Content words describing the research purposes in AWL

Finally, the content words describing the research purpose outside the HFW and AWL comprised eight content words (3.86% of words describing the research purposes); that is to say, *essay*, *text*, *concentrate*, *optimize*, *outline*, *sketch*, *underscore*, and *workshop*. The word ‘*essay*’ was used the most (1.49% of words describing the research purposes).

In short, 68 content words describing the research purposes covered the three lists of words; namely, HFW, AWL, and the list of words outside the HFW and AWL. The content words in high-frequency word list occurred the most (79.46% of words describing the research purposes). Purpose words in high-frequency word list of the corpus were also accounted for 66.67% of words describing the research purposes in high-frequency word list, which were consistent with West’s GSL (1953). The second was AWL (16.67% of words describing the research purposes). Furthermore, purpose content words in AWL were accounted for 46.67% of purpose words in AWL, consistent with Coxhead’s (2000). The last was OWL (3.87% of purpose words).

4.3.2 Verb Tenses

In addition to the lists of content words describing the research purposes, according to the collection of 133 purpose sentences in the LARA corpus, it was found that there were four tenses which were used in the purpose sentences, as shown in Figure 4.6.

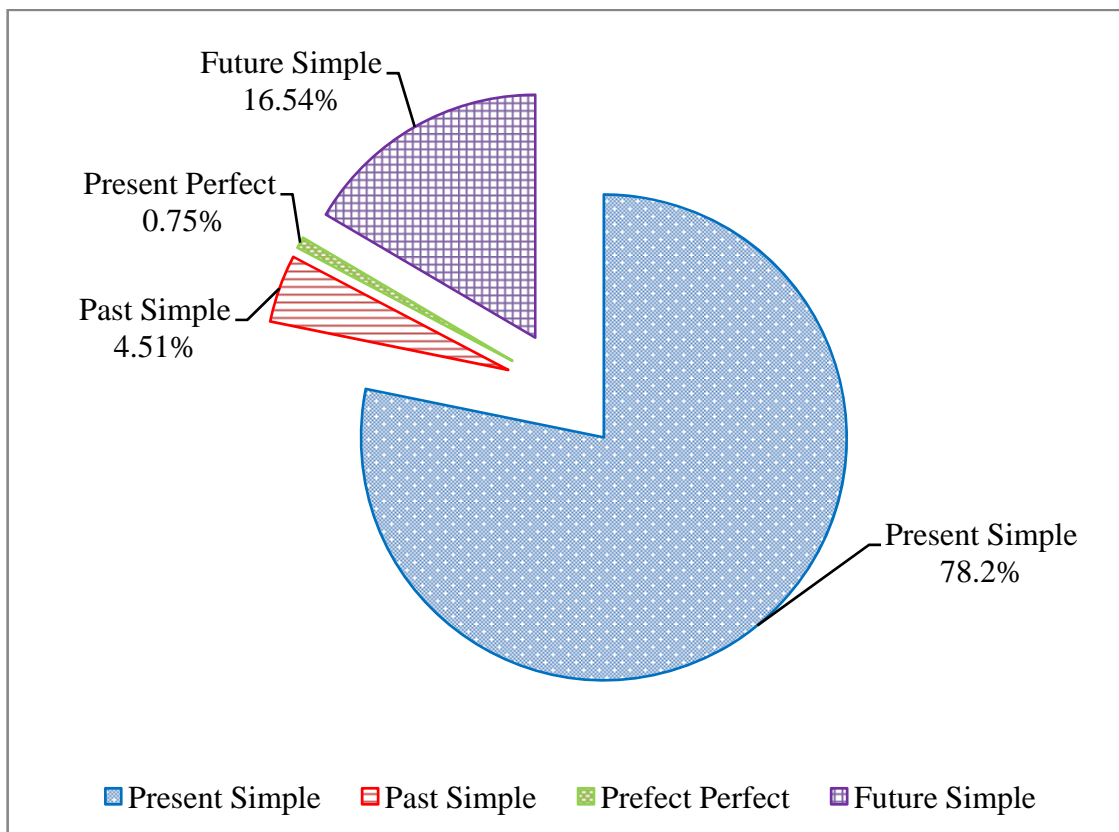


Figure 4.6 Tenses in purpose sentences

According to Figure 4.6, present simple tense was used the most (78.20% of the total tenses), while present perfect tense was used the least (0.75%). Future simple tense was found the second most frequently (16.54%). Finally, past simple tense was the third (4.51%). The examples are shown below.

4.3.2.1. Purpose Sentences in Present Simple Tense

The present simple tense in the sentences presenting the research purposes occurred 104 sentences. For example,

(1) In this article we **show** that it is possible to create a reproducible ovine model of postinfarct DCM by inducing an anterolateral infarct of approximately 25% of the LV mass based on the visual selection of the arteries that require ligation.

(Locatelli et al., 2011)

(2) In this review we **highlight** some salient contributions of the mouse in aging research: lifespan intervention studies in the Interventions Testing Program of the National Institute on Aging; identification of the genetic underpinnings of the effects of calorie restriction on lifespan; the Aging Phenome Project at the Jackson Laboratory, which has submitted multiple large, freely available phenotyping datasets to the Mouse Phenome Database; insights from spontaneous and engineered mouse mutants; and complex traits analyses identifying quantitative trait loci that affect lifespan. in the laboratory.

(Yuan, Peters & Paigen, 2011: 4)

(3) The objective of this review **is** to describe the features of the normal neurogenic niche in the adult mammalian subventricular zone (SVZ1) as well as the factors that promote (1) NSC homing and recruitment to a site of disease and (2) the integration of NSCs into functional tissues in different animal models of human neurological disease.

(Garzón-Muvdi & Quiñones-Hinojosa, 2010: 3)

(4) This article **discusses** the variety of neoplasia observed in zebrafish submitted to the ZIRC diagnostic service, neoplasia and related lesions in sentinel fish from selected colonies, results of our prospective tumor studies with WT and mutant lines, our studies of neoplasia in retired broodstock from

various colonies, and the diversity of neoplasia documented in carcinogen and genetic research using zebrafish.

(Spitsbergen, Buhler & Peterson, 2012: 115)

(5) *This review aims to show that noninvasive brain imaging strategies such as small animal positron emission tomography offer significant potential and promise for modeling motivational disorders such as drug addiction and obesity in humans.*

(Michaelides, Thanos, Volkow & Wang, 2012: 59)

(6) *This review is meant to briefly describe the utility of specific preparations that can be used in the fly to explore injury-induced degeneration and functional loss of axons, dendrites, and synapses.*

(Rooney & Freeman, 2014: 291)

(7) *Because animal models may not be exact in all features of the disorder being studied, it is important to investigate the strengths and weaknesses of using a nonhuman primate model for stereotyped behavior in people with....*

(Lutz, 2014: 284)

According to (1) – (7), the verb tenses were the present simple tense. That is to say, the verbs of this tense were the base form such as the verbs in (1) and (2). However, in (3) – (7), the subjects were the third person singular, so the verb form added an ‘s’ or ‘es’. Interestingly, the verb in (6) was the passive form of the present verb.

4.3.2.2. Purpose Sentences in Future Simple Tense

The future simple tense in the sentences presenting the research purposes was composed of 22 sentences. For example,

(8) *In this article, the focus **will be** briefly on the human and canine spinal anatomy, after which this review will particularly focus on spondylosis and DISH in both humans and dogs.*

(Kranenburg, Hazewinkel & Meij, 2014: 150)

(9) *In this review we **will explore** the comparative aspects of human and equine tendon pathophysiology in these specific structures and the practical considerations in obtaining clinically relevant research data for both species.*

(Patterson-Kane & Rich, 2014: 87)

(10) *This review **will discuss** the most widely used mouse models of DPN, including the main phenotypic characteristics of each model, how to choose the appropriate model, and important considerations when working with selected strains.*

(O'Brien, Sakowski & Feldman, 2014: 259)

(11) *This article **will detail** some of the issues that must be considered as institutional animal care and use committees (IACUCs) review the use of NHPs in research.*

(Tardif, Coleman, Hobbs & Lutz, 2013: 234)

(12) *A variety of standard histopathological and immunopathological procedures including Epidermal Nerve Fiber Density (ENFD) analysis, Schwann Cell Density, and analysis for other conserved cellular markers can be used effectively with armadillos and **will be briefly reviewed** in this text.*

(Truman et al., 2014: 304)

(13) This review **will focus on** the mechanisms through which genetic and environmental factors (specifically dietary) can act alone or in combination to alter DNA methylation patterns, and we **will review** the evidence supporting a critical role for these mechanisms in the etiology of ASDs.

(Schaevitz & Berger-Sweeney, 2012:323)

The examples (8) – (13) used the verb forms of the future simple tense (*will* + V_{inf}). Interestingly, the sentence (12) was formed into the passive of the future tense (*will be* + *past participle*).

4.3.2.3. Purpose Sentences in Past Simple Tense

The past simple tense in the sentences presenting the research purposes included six sentences. For example,

(14) The aim of the present study **was** therefore to modify, optimize, and extend the *in vitro* system of bovine respiratory mucosal explants in order to study the BoHV-1 dissemination kinetics in respiratory mucosa.

(Steukers et al., 2012)

(15) An objective **was** to promote communication among research groups with the hope that lessons learned in one species might help inform and advance understanding in others.

(Mansfield, Riley & Kent, 2010: 171)

(16) The aim of the study **was** to design a protocol to train alpacas successfully to remain in a specially designed metabolism pen that would be used for future nutritional studies.

(Lund, Maloney, Milton & Blache, 2012)

(17) We ***aimed*** to (1) study the anatomy of the brain vasculature in our gerbil population and identify heritable characteristics of CoWs and (2) use the results of this study to selectively breed an ischemia-sensitive gerbil group.

(Du et al., 2011)

The verb forms in the examples (14) – (17) were the past simple tense (V_2). The examples (14) – (16) used the irregular verbs of ‘be’ (*was*), while the example (17) used the regular verb (*aimed*).

4.3.2.4. Purpose Sentence in Present Perfect Tense

The present perfect tense in the sentences presenting the research purposes included one sentence. For example,

(18) In this review, we ***have examined*** the current literature and expert opinions on environmental epigenetics, a term that is narrowly defined as how epigenetics explains the variability in the risk and severity of environmental disease.

(Ho et al., 2012: 289)

According to the example (18), the present perfect tense was used to show the research purposes. The verb tense was formed with the present tense of have and the past participle (V_3).

In conclusion, according to the corpus of the purpose sentences, the verb forms of each sentence were composed of the different tenses. In other words, there were four verb tenses in the purpose sentences. The present simple tense occurred the most (104 sentences). The tense used the second most frequently was the future simple tense (22 sentences). Six of the past simple tense occurred the third highest. Finally, the present perfect tense was used the least (one sentence).

4.3.3 Patterns

In terms of the patterns of the research purposes, laboratory animal researchers used difference patterns to describe the research purposes. According to the 160 review articles, there were 133 sentences which described the research purposes. The purpose sentences were grouped into eight patterns shown in Table 4.8.

Table 4.8 Eight Patterns Describing the Research Purposes in the LARA Corpus

	Patterns	Number	%
I	Signal words for research purposes and to-infinitive	20	15.04
II	Signal words for research purposes as the subject and phrases explaining the details of research purposes	4	3.00
III	Personal subjects	50	37.60
IV	Impersonal subjects	45	33.84
V	Main verbs and to-infinitive	5	3.76
VI	Main verbs in the passive and to-infinitive	7	5.26
VII	It is + adjective + to-infinitive	1	0.75
VIII	Mixed pattern	1	0.75
	TOTAL	133	100

% of the total of patterns

According to Table 4.8, the Pattern III was found the most (50 sentences). The second most frequent pattern was the Pattern IV (45 sentences). The third most frequent pattern was the Pattern I (20 sentences). Finally, only one purpose sentence was presented with two patterns.

4.3.3.1 Pattern I: Signal Words for Research Purposes and To-Infinitive

In relation to Pattern I, there were 20 sentences describing the research purposes. This pattern used the signal words of purposes as a subject of the sentence. The main verb of this pattern was 'verb to be' (*is, are, was, and were*) and 'to-infinitive' (*to + VI*) was followed by the *verb to be*. The function of *to-infinitive* in

this pattern is to express the purpose; namely, 'in order to'. The pattern is shown in Table 4.9 and the examples are shown under the table.

Table 4.9 Pattern I: Signal Words for Research Purposes and To-Infinitive

Subject (signal words of purposes)			Verb to be	To-infinitive (to + V1)	Noun phrase	
The objective (s)	of	this research	is (are) was (were)	to	provide investigate give identify describe design present modify introduce optimize gain extend use report review summarize discuss	Noun details of purpose(s)
The goal (s)		the (present) study				
The (main) aim (s)		this review				
The purpose (s)		this article the workshop				
The objective (s)						
The goal (s)						
The purpose (s)						

e.g.

Signal words for research purposes

(1) *The objective of this review is to describe the features of the normal neurogenic niche in the adult mammalian subventricular zone (SVZ1) as well as the To-infinitive note (1) NSC homing and recruitment to a site of disease and (2) the integration of NSCs into functional tissues in different animal models of human neurological disease.*

(Garzón-Muvdi & Quiñones-Hinojosa, 2010: 3)

Signal words for research purposes

(2) *The goal of this article is to present lessons learned from the devastating effects of two specific natural disasters in Texas: Tropical Storm Allison, which fl To-infinitive June 2001, and Hurricane Ike, which caused severe damage in Galveston in September 2008.*

(Goodwin & Donaho, 2010: 104)

Signal words for research purposes

(3) *The aim of this article is to provide scientific information concerning laboratory confinement of Psittaciformes, with the intention of refining arrangements for housing, and maintenance.*

(Kalmar, Janssens, & Moons, 2010: 410)

Signal words for research purposes

(4) *The purpose of this article is to review the existing literature on anesthetic protocols adopted in mice for molecular imaging studies.*

(Gargiulo et al, 2012)

Signal words for research purposes

(5) *An objective was to promote communication among research groups with the hope that the knowledge learned in one species might help inform and advance understanding in others.*

(Mansfield, Riley & Kent, 2010: 171)

Signal words for research purposes

(6) *Our goal is to survey the results of in vivo experimentation in order to overcome the challenges to successful clinical application of SC-based treatments for neurological disorders.*

(Joers & Emborg, 2010: 24)

Signal words for research purposes

(7) *Our purpose is to underscore the importance of up-to-date crisis management planning for knowledgeable, realistic assessment of risks.*

(Bailey, Rich & Bennett, 2010: 139)

According to examples (1) – (4), the subjects of each sentence were modified by the prepositional phrase (*of + noun*). However, the subject in (5) – (7) was only the signal word for research purposes. Moreover, the *to-infinitive* of each sentence contained the verbs which describe the research purposes.

4.3.3.2 Pattern II: Signal Words for Research Purposes as the Subject and Phrases Explaining the Details of Research Purposes

Pattern II was almost similar to Pattern I. In other words, the subject of this pattern was the signal words for research purposes, but the subject complement was phrases explaining the details of purposes. The main verb was also ‘verb to be’. This pattern contained four sentences, as shown in Table 4.10 and three examples below.

Table 4.10 Pattern II: Signal Words for Research Purposes as the Subject and Phrases Explaining the Details of Research Purposes

Subject (signal words for research purposes)			Verb to be	Phrase
The focus (es)			is (are) will be	Prepositional phrase
The focus (es)	of	this review this article		Noun details of purpose(s)

e.g.

Signal words for research purposes

(1) *In this article, the focus will be briefly on the human and canine spinal anatomy, after which this review will particularly focus on sp and DISH in both humans and dogs.*

Adverb

Prepositional phrase

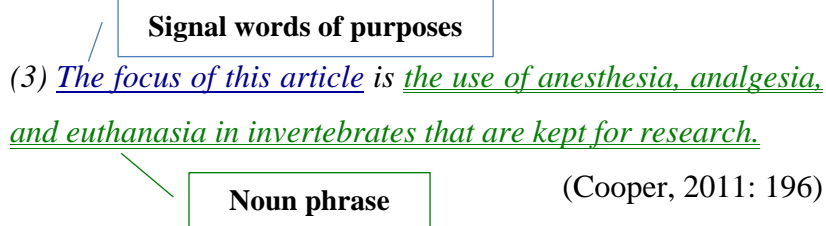
(Kranenburg, Hazewinkel & Meij, 2014: 150)

Signal words of purposes

(2) *The focus of this review is the concerted action of these signals as well as the regulatory and/or stabilizing control circuits exhibited by a class of small RNAs, designated microRNAs.*

Noun phrase

(Gordeladze et al, 2010: 42)



The example (1) used the main verb of the future tense and the prepositional phrase (*on the human and...*). Moreover, the adverb '*briefly*' was used to modify the prepositional phrase as the subject complement. However, the subjects of the sentence in (2) and (3) were modified by the prepositional phrase (*of + noun*). Furthermore, the subject complements were the noun phrase showing the details of the research purposes.

4.3.3.3 Pattern III: Personal Subjects

Pattern III occurred the most (50 sentences). Interestingly, the subject of the pattern was the personal noun which was the first person pronouns or the noun '*author(s)*'. Moreover, this pattern started with the references to the text as the prepositional phrase (*in + noun*) or the adverb of place (*here*). Furthermore, the phrasal verbs were found in the main verbs of this pattern such as *focus on* and *touch upon*. Also, the researchers used the adverb of manner '*briefly*' to modify the main verb in some sentences of this pattern. Additionally, the objects of the main verbs were both noun phrases and noun clauses showing the details of the purposes, as shown in Table 4.11 and the examples below.

Table 4.11 Pattern III: Personal Subjects

References to the text		Personal subject	Main verb (V1 / has, have + V3 / will + V. inf.)		Phrase / Clause
Prep.	Noun phrase	we I author(s)	present(s)	show(s)	Noun phrase / Noun clause
In	this review		examine(s)	provide(s)	
	this essay		discuss(es)	focus(es) on	
Here	this article	describe(s)	assess(es)		
	the (present) report	(briefly) review(s)	summarize(s)		
Here	this paper	identify(ies)	propose(s)		
		evaluate(s)	sketch(es)		
		highlight(s)	touch(es) upon		
		explore(s)	determine(s)		
		illustrate(s)			

e.g.

First personal pronoun

(1) In this article I describe an example of progress in a nontraditional species, through studies of the role of sleep in birdsong learning.

Main verb

(Margoliash, 2010: 378)

First personal pronoun

(2) In this article, we briefly review research that contributes to understanding how One Health messages could affect support for wildlife health and conservation.

Adverb

Main verb

(Decker et al, 2010: 256)

Personal noun

(3) In this article the authors (one DVM and one MD) present examples of species “cross talk” based on their parallel careers in musculoskeletal oncology sur

main verb

(Withrow & Wilkins, 2010: 208)

(4) In this article **we** describe [how clinical and basic research can substantiate the claims of AAT and we identify specific components for successful AAT.]

First personal pronoun

Main verb

Noun clause

(Palley, O'Rourke & Niemi, 2010: 199)

(5) **Here**, **we** briefly review the state of knowledge of human cancer genetics to elaborate on the need for different types of mammalian models, highlighting the strengths of the dog.

Adverb of place

First personal pronoun

Main verb

(Alvarez, 2010: 16)

(6) In this review, **we** have examined the current literature and expert opinions on environmental epigenetics, a term that is narrowly defined as how epigenetics explains the variability in the risk and severity of environmental disease.

First personal pronoun

Main verb

(Ho et al, 2012: 289)

(7) In this review **we** will explore the comparative aspects of human and equine tendon pathophysiology in these specific structures and the practical considerations in obtaining clinically relevant research data for both species.

First personal pronoun

Main verb

(Patterson-Kane & Rich, 2014: 87)

(8) **We** examine the ethical and policy-level aspects of research and conservation activities that involve captive wildlife in zoos and aquariums, focusing on some of the implications of accelerating biodiversity decline and rapid environmental change.

First personal pronoun

Main verb

(Minteer & Collins, 2013: 42)

According to examples, the subjects of all examples were the first person pronoun ‘*I*’ and ‘*we*’, except that the example (3) was ‘*the authors*’, the plural personal noun. All of the examples began with the references to the text as the prepositional phrases, but in (5), it started with the adverb of place ‘*here*’. However, in (8), the sentence started without the reference of the text. In (2), the adverb ‘*briefly*’ was used to modify the main verb. Finally, the objects of the main verb were noun phrases, except that the object of (4) was the noun clause, showing the details of the research purposes.

4.3.3.4 Pattern IV: Impersonal Subjects

Pattern IV was found the second most frequent (45 sentences). It resembled the pattern III. In other words, the pattern included a subject, a main verb, and a direct object. However, the subject of this pattern was different from the pattern III. That is to say, the subject was an impersonal noun such as ‘*review*’, ‘*article*’, ‘*essay*’, ‘*text*’, ‘*overview*’, ‘*report*’, or ‘*paper*’. Likewise, the adverb ‘*briefly*’ was also used to modify the main verb. Finally, the direct object of the main verb was the noun phrase and the noun clause giving the details of the research purposes shown in Table 4.12.

Table 4.12 Pattern IV: Impersonal Subjects

Impersonal subjects	Main verb (V1 / will + V. inf.)		Phrase / Clause
This paper	explore(s)	(briefly) discuss(es)	Noun phrase / Noun clause
This (present) review	provide(s)	address(es)	
The article	allow(s)	outline(s)	
This essay	highlight(s)	detail(s)	
The text	purpose(s)	illustrate(s)	
The review article	review(s)	show(s)	
This overview	present(s)	identify(ies)	
This report	summarize(s)	concentrate(s) on	
The study of ...	focus(es) on	end(s)	
A review of ...			

e.g. | **Impersonal noun** | **Main verb**
 (1) *This review highlights the usefulness of investigations using Japanese quail (*Coturnix japonica*) to illustrate general principles about the hormonal control of reproductive and related social behaviors.*

(Ball & Balthazart, 2010: 310)

| **Impersonal noun** | **Adverb**
 (2) *The present review focuses specifically on the transmission and control of microsporidia in zebrafish facilities.*

Main verb

(Sanders, Watral & Kent, 2012: 106)

| **Impersonal noun**
 (3) *This essay will discuss some of the ethical issues raised by African great ape research in the hopes of generating a greater dialogue about best practices.*

Main verb

(Gruen, Fultz & Pruetz, 2013: 24)

| **Impersonal noun** | **Noun clause**
 (4) *This review highlights [how recent comparative studies have enhanced the knowledge of human genetics and human evolution.]*

Main verb

(Wall, 2013: 82)

| **Impersonal noun** | **Main verb**
 (5) *A review of the literature shows [that the toxicity and pathogeneses of chemicals adversely affecting the peripheral nervous system have been studied using animal models.]*

(Rao, Jortner & Sills, 2014: 315)

According to the five examples, the subjects of each sentence were the impersonal nouns such as *review* or *essay*. Notably, the subject in (2) was modified by the adjective ‘*present*’ in front of the subject, and the subject in (5) was also modified by the prepositional phrase ‘*of the literature*’ after the subject. In terms of the main verb, the main verbs were not only one word but also a phrasal verb such

as ‘concentrate on’ or ‘focus on’. Noticeably, the main verb in (2) was modified by the adverb ‘specially’ to emphasize the main verb. In relation to the direct object of the sentences, the objects of (1) – (3) were noun phrases, while the examples (4) and (5) were noun clauses. Nevertheless, both types of the object explained the details of the research purposes.

4.3.3.5 Pattern V: Main Verbs and To-Infinitive

There were five sentences which were found in this Pattern V. This pattern used the main verbs and to-infinitive in order to present the objectives. The subjects were the first person pronoun (*I* and *we*) and the impersonal noun. The objects of the pattern were noun phrases and noun clauses which explained the details of the research purposes. The pattern is displayed in Table 4.13.

Table 4.13 Pattern V: Main Verbs and To-Infinitive

Subject	Main verb	To-infinitive (to + V1)	Phrase / Clause
I	aim (s/ed) act (s) as (+ object)	to	Noun phrase / Noun clause
We			
This review			
This article			
		address	
		study	
		use	
		show	
		provide	
		understand	

e.g.

(1) In this article we aim to address the above deficits by providing a broad overview of the most salient aspects of passerine biology and their implications for the husbandry and welfare of these species in the laboratory.

To-infinitive

Main verb

(Bateson & Feenders, 2010: 395)

Main verb

(2) We aimed to (1) study the anatomy of the brain vasculature in our gerbil population and identify heritabilities of CoWs and (2) use the results of this study to selectively breed an ischemia-sensitive gerbil group.

To-infinitive

(Du, 2011)

Main verb

To-infinitive

(3) This article aims to provide insights into the usefulness of the mouse model for epigenetic studies involving nutrition as well as the inherent limitations when compared with epigenetic phenomena in humans.

(Niculescu, 2012: 270)

Main verb

(4) This review aims to show that noninvasive brain imaging strategies such as small animal positron emission tomography offer significant potential and promise for modeling motivational disorders such as drug addiction and obesity in humans.

To-infinitive

(Michaelides et al., 2012: 59)

Main verb

(5) This review acts as a succinct resource and provides references for investigators supported by the PHS to understand the main expectations and requirements when using animals in research.

To-infinitive

(Silk, Hampton & Brown, 2014: 324)

According to all of five examples, the subject of (1) and (2) was the first person pronoun (*I* and *We*), whereas the examples (3) – (5) was the impersonal noun (*article* and *review*). The main verbs to present the research purposes were ‘*aim*’ and ‘*act as*’ and followed the *to-infinitive*. The sentence (2) presented two objectives; that is, the former one was ‘*to study*’, and the later one was ‘*use*’. Notably, the later used the *bare infinitive*. For example (5), the main verb ‘*acts as*’ followed a

noun phrase, and after the noun phrase the *to-infinitive* was used to present the research purposes.

4.3.3.6 Pattern VI: Main verbs in the Passive and To-Infinitive

The Pattern VI was composed of seven purpose sentences. The pattern was formed into the structure of passive voice (be + V3) and followed *to-infinitive* to express the research purposes. The pattern is displayed in Table 4.14.

Table 4.14 Pattern VI: Main verbs in the Passive and To-Infinitive

Subject	Main Verb (Passive voice)		To-infinitive (to + V1)	
	be	Past participle (V3)		
The (brief) overview of ...	is (are)	intended	to	aid (briefly) describe provide
This review	was (were)	focused on		
This article	will be	meant		
A variety of procedure	can be			
Several methods				

e.g.

Passive verb

(1) *This overview of Internet resources is intended to aid both beginning and experienced individuals in the search for such information.*

To-infinitive

(Smith, 2012: 165)

Passive verb

To-infinitive

(3) *This review is meant to briefly describe the utility of specific preparations that can be used in the fly to explore injury-induced degeneration and functional loss of axons, dendrites, and synapses.*

(Rooney & Freeman, 2014: 291)

(3) A variety of standard histopathological and immunopathological procedures ...*can be used* effectively with armadillos and *will be* briefly *reviewed* in this text.

Passive verb

(Truman et al., 2014: 304)

(4) In this article, several allocentric assessment methods for rodents *are reviewed and compared with* the MWM. MWM advantages (little training required, no food deprivation, ease of testing, rapid and reliable learning, insensitivity to differences in body weight, absence of nonperformers, control methods for proximal cue learning, and performance effects) and disadvantages (concern about stress, perhaps not as sensitive for working memory) are discussed.

Passive verb

(Vorhees & Williams, 2014: 310)

All of four examples used the verb from of the passive voice such as 'is intended', 'is meant', 'can be used', 'will be reviewed', and 'are reviewed and compared'. In (3), the modal verbs, 'can' and 'will' were used and formed into the passive form. The examples (1) and (2) used *to-infinitive* to present the research purposes, while the examples (3) and (4) did not use it.

4.3.3.7 Pattern VII: It is + Adjective + To-Infinitive

The last pattern is shown in the box below. The pattern starts with the anticipatory 'It' and followed by the 'postponed subject'. The pattern 'It is + *to-infinitive*' emphasizes the information of the research purposes.

It is + adjective + to-infinitive

e.g.

(1) *Because animal models may not be exact in all features of the disorder being studied, it is important to investigate the strengths and weaknesses of using a nonhuman primate model for stereotyped behavior in people with psychological disorders.*

To-infinitive

(Lutz, 2014: 284)

According to the example, it stated the importance of investigating the strengths and weakness in the study. *To-infinitive phrase* was the expression of the details of the research purposes.

4.3.3.8 Pattern VIII: Mixed Pattern

Additionally, one sentence presenting the research purposes mixed the pattern III and IV. In other words, the sentence used the pattern with the personal noun as the subject and the pattern with the impersonal noun as the subject, shown in the below example.

(1) [*This review will focus on the mechanisms through which genetic and environmental factors (specifically dietary) can act alone or in combination to alter DNA methylation patterns,*][PATTERN 4] and [*we will review the evidence supporting a critical role for these mechanisms in the etiology of ASDs.*][PATTERN 3]

Impersonal noun

Main verb

First person pronoun

(Schaevitz & Berger-Sweeney, 2012: 323)

More importantly, each pattern was differently formed by the verb tenses. The numbers and percents of each tense occurring in each pattern presenting the research purposes are shown in Table 4.15.

According to Table 4.15, the present simple tense was used the most in seven of the patterns (78.20%). Particularly, in the pattern III, the verb tenses of the present simple tense occurred the most (32.33%), whereas the pattern VIII did not

appear. Moreover, the pattern IV occurred the second highest (24.06%). The second most frequent verb tenses in the corpus was the future simple tense (16.54%). The pattern IV consisted of 13 sentences with the verb form of the future simple tense (9.77%). The verb form of the past simple tense occurred the third most frequently (4.51%). In the pattern I, the verb tense of the past simple tense was used the most (3.76%), and the other used in the pattern V (0.75%), which occurred five times as frequently as the pattern I. Finally, the present perfect tense was used the least (0.75%) and occurred in the pattern III only.

Table 4.15 Numbers and Percentage of Verb tenses in the Patterns Presenting the Research Purposes

Tense	Present Simple		Past Simple		Present Perfect		Future Simple		TOTAL	
	N ¹	% ²	N ¹	% ²	N ¹	% ²	N ¹	% ²	N ¹	% ²
I	15	11.28	5	3.76	0	0.00	0	0.00	20	15.04
II	3	2.26	0	0.00	0	0.00	1	0.75	4	3.00
III	43	32.33	0	0.00	1	0.75	6	4.15	50	37.60
IV	32	24.06	0	0.00	0	0.00	13	9.78	45	33.84
V	4	3.01	1	0.75	0	0.00	0	0.00	5	3.76
VI	6	4.51	0	0.00	0	0.00	1	0.75	7	5.26
VII	1	0.75	0	0.00	0	0.00	0	0.00	1	0.75
VIII	0	0.00	0	0.00	0	0.00	1	0.75	1	0.75
TOTAL	104	78.20	6	4.51	1	0.75	22	16.55	133	100

¹Number of verb tenses in each pattern

²% of the total of verb tenses

To conclude, this LARA corpus contained eight patterns which presented the research purposes. The pattern with the personal subject was found the most. Moreover, the pattern VII, “*It is + adjective + to-infinitive*” and Pattern VIII were found the least. Interestingly, the pattern VIII was formed with mixing the pattern III and IV. Additionally, each pattern was composed of different verb tenses. In other words, the present simple was used the most, while the present perfect tense was used

the least. The present simple tense occurred in the pattern III the most, and the present perfect tense also occurred in the pattern III.

Chapter Summary

All in all, in the present research, the word tokens comprised 840,773 words and 22,434 word types which include 337,362 function words (44.15%) and 469,560 content words (55.85%). Moreover, the LARA corpus consisted of 703,296 high-frequency words (83.65% of text coverage), comprising 337,362 content words (40.13%) and 365,934 function words (43.52%). The first most occurred content word of this corpus was the word *'studies'*, whereas the 100th content word was the word *'system'*. In terms of academic words, the frequency of word tokens of academic words was 36,658 (4.36% of text coverage and 7.81 % of word families). Furthermore, the headwords of content words of AWL included 768 words. The first two academic words (53 times) were *'regardless'* and *'surrounding'*. Most importantly, the LARA corpus was found that there were 21 word families of the academic words which were consistent with word families of Coxhead's AWL. Finally, the findings presented 68 content words presenting the research purposes, which covered in the three word lists of the corpus. In other words, 45 content words covered in the HFW; 15 in AWL; and eight in the list of words outside HFW and AWL. Moreover, eight patterns presented the research purposes in the LARA corpus. The third pattern *'The Personal Subject'* was found the most frequently, while the pattern, *"It is + adjective + to-infinitive"* was the lower. There was one sentence structured by mixing the pattern III and IV. More importantly, all the eight patterns used the verb form of the present simple tense the highest, whereas the present perfect tenses was used the lowest.

CHAPTER V

DISCUSSION

This chapter is to discuss the results, give a summary, and make recommendations for future studies. The discussion of the results carried out according to the objectives of the study. Moreover, the interpretation of the results is derived from the research problems and previous studies. This section discusses similarities and differences in the results. Finally, implications for future research are also identified in the section. This chapter discusses the following:

- 5.1 High-frequency words
- 5.2 Content words from the academic word list
- 5.3 Words and patterns for presenting the research purposes
- 5.4 Implications of the study

5.1. High-Frequency Words

Research question 1: What hundred content words appear most frequently in laboratory animal review articles?

The current study identified the 100 highest frequency content words in the corpus. The results, shown in Figure 5.1, reveal that high-frequency words were accounted for 83.65% of text coverage. Moreover, the high frequency words consisted of function words (43.52% of text coverage) and content words (40.13% of text coverage). The thousand highest frequency words covered 74.17%, while the next thousand covered 9.48%. Importantly, the first content word in the corpus was the words *studies* and the 100th was *systems*. Therefore, the results for high frequency words are discussed in terms of high frequency words, the proportions for high frequency words, and the proportions for word categories: function words and content words.

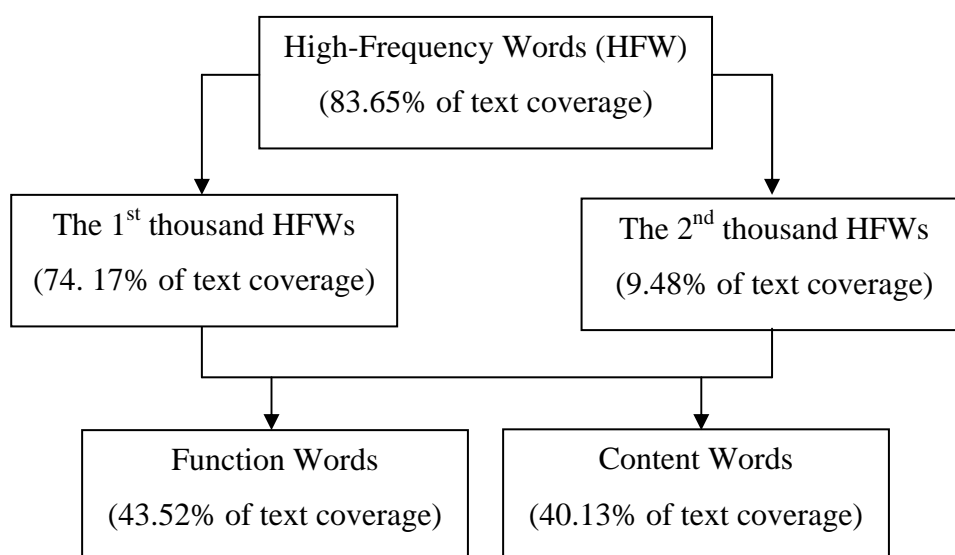


Figure 5.1 Summary of the findings for high frequency words in the LARA corpus

Firstly, the occurrence of high-frequency words in the corpus was higher than the 80% found by West (1953) or the 76% found by Coxhead (2,000). A comparison with previous studies reveals that the high frequency words in the current study are also used more frequently than in previous studies, as shown in Figure 5.2; that is, the academic text corpus (Coxhead, 2000), a psychology corpus (Kaewphanngam, 2002), an engineering corpus (Para, 2004), an academic corpus (Hyland & Tse, 2007), an agriculture corpus (Martinez et al., 2009), a financial services corpus (Li & Qian, 2010), and an environmental science corpus (Liu & Han, 2015). However, the current study is lower than the 85.05% found in the corpus of the Thailand university admission tests (Cherngchawano & Jaturapitakkul, 2014) because the admission tests include the general knowledge that is more appropriate for the knowledge level of the students than is specific knowledge. Likewise, Nation (2001) stated that 80% of text coverage in a corpus is suitable for learners to be able to study. That is, the vocabulary size of the high-frequency words in the current corpus is suitable for a member in the discourse community of laboratory animal research. Furthermore, differences in high frequency words among the various corpora are due to changes in English language and culture (Paquot, 2010). In other words, the words from the high frequency word list are regarded as the limitations of each field.

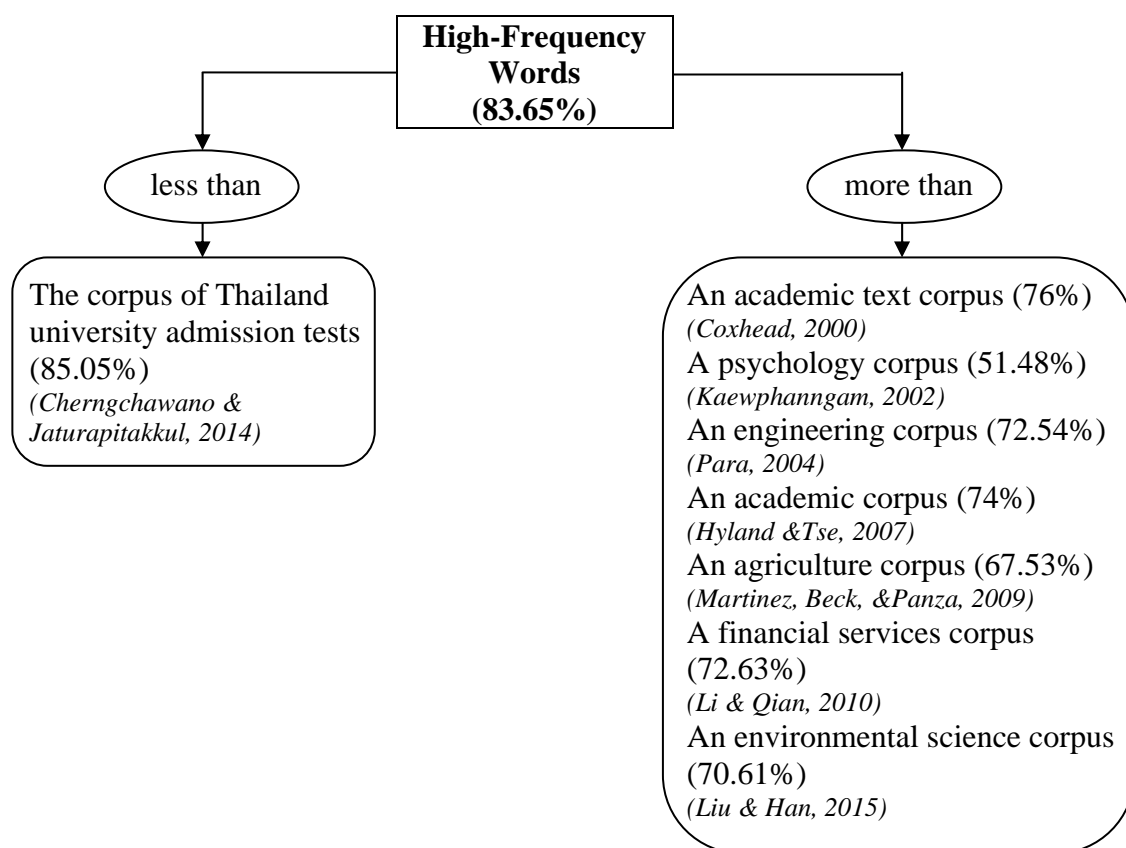


Figure 5.2 Comparison of high-frequency words in this study with previous studies

Given the ratio for high frequency words in the corpus, the first thousand high frequency words occurred more frequently than the second thousand because the words with the highest frequency are words such as *the* (5.41%), *of* (4.38%), and (3.64%), *in* (3.19%), and *to* (2.57%), all of which are among the top five words in the recent corpus. Similarly, in the first ten words, six words are the same as those found by West (1953): *the*, *of*, *and*, *in*, *to*, and *a*. However, the ranking is different, except for the first word *the*, which also received the top rank from West (1953). The 74.17% coverage for the first thousand words is lower than that found by West (1953) (77% of text coverage), but higher than for Coxhead (2000) (71.4% of text coverage). Another ratio for high frequency words is the second thousand high frequency words. They cover less than the first thousand because the first word *healthy* in the second thousand covers only 0.014%. However, the 9.48% coverage for the second thousand higher than that found by West (1953) or Coxhead (2000).

Finally, word classes in the high frequency words consisted of content words and function words. The proportion of function words (43.52% of text coverage) was higher than that for content words (40.13% of text coverage) because the first twenty-three words among the first thousand high frequency words were function words such as *in, to, a, that, is, for, with, are, as, or, be, by, on, 'have, this, from, an, not, and these* respectively. Function words provide grammatical information. In other words, function words may be used beside the lexical words to provide grammatical information (Tallerman, 2005). Likewise, the coverage of the function words is consistent with that found in previous studies (43-44% of the text coverage) (Nation, 2001). The other word class is content words. In this current corpus, the first content word was ranked 24; it was the word *studies* (0.36%). The first content word in West's study (1953) was the word *study* at rank 35.

To summarize, it is evident that the value of high-frequency words in the LARA corpus (83.65% of text coverage) is lower than that of the corpus of Thailand university admission tests because the vocabulary words in the corpus are secondary core vocabulary needed for their admission tests. The coverage of high-frequency words is higher than that for several previous studies such as the corpus of academic texts (Coxhead, 2000; Hyland & Tse, 2007), psychology, engineering, agriculture science, financial services, and environmental science. This is due to limitations of each field; the high frequency words are different for each corpus. Moreover, the proportion for the use of the first thousand high frequency words in the corpus is higher than that of the second thousand because of the highest frequency words in the first list. This is consistent with previous studies. Nevertheless, the first thousand words are in accord with the findings of previous studies. They cover less than what was found by West (1953), but more than Coxhead (2000). Last, function words in the high-frequency words are more frequent than content words because the first twenty-three words in the first thousand high frequency words are function words. They also provide grammatical information, and may be adjacent to lexical words. Importantly, high-frequency words are considered the best decision for learners continuing to academic study of laboratory animal science.

5.2 Content Words from the Academic Word List

Research question 2: What content words from the AWL occurred with high frequency in the same corpus of laboratory animal review articles?

The second objective of the recent study was to investigate words from the academic word list (AWL) occurring in the LARA corpus. A total of 36,658 academic word tokens were found, calculated as 4.36% of text coverage. In addition, 768 content headwords occurred in the LARA corpus. Finally, the hundred most frequent academic words in the corpus included twenty-one word families, which is consistent with the word families on Coxhead's AWL. The following discussion provides details in terms of comparing words and numbers for the LARA corpus with previous research.

First of all, when the results for the list of AW word families (4.36% of text coverage) are compared with the text coverage for Coxhead's academic corpus (2000), AWL of LARA corpus covers less than Coxhead (approximately 10% of text coverage), and also science sub-corpus of the academic corpus (Coxhead, 2000) (9.1%). This is because the science sub-corpus of Coxhead (2000) covers seven subject areas, which is more than for the recent study. Moreover, the coverage in this corpus is lower than other previous studies, as shown in Table 5.1; namely, a psychology corpus (Kaewphanngam, 2002), civil engineering corpus (Para, 2004), an academic corpus with a science sub-corpus (Hyland & Tse, 2007), a medical corpus (Chen & Ge, 2007), a medical corpus (Wang et al., 2008), an applied linguistics corpus (Vongpumivitch et al., 2009), an agricultural sciences corpus (Martinez, et al., 2009), a financial services corpus (Li & Qian, 2010), a chemistry corpus (Valipouri & Nassaji, 2013), an environmental science corpus (Liu & Han, 2015), and a nursing corpus (Yang, 2015). However, the coverage in the LARA corpus is similar to that for the corpus of the Thailand University Admission test (4.58%) (Cherngchawano & Jaturapitakkul, 2014). This is because the corpus for this current study is only the *ILAR Journal*, which is different from previous studies in which the corpus came from five or more related journals. However, the results indicate that words from the AWL are equally useful if members in the discourse community of the LARA field are to read and write academic works.

Table 5.1 Comparison between the AWL in the LARA Corpus and Previous Studies

Corpus	% of text coverage
Laboratory animal review articles corpus	4.36%
Academic (Coxhead, 2000)	8.5%
Science sub-corpus (Coxhead, 2000)	9.1%
Psychology (Kaewphanngam, 2002)	7.78%
Civil engineering (Para, 2004)	12.46%
Academic (Hyland & Tse, 2007)	10.6%
Science sub-corpus (Hyland & Tse, 2007)	9.3%
Medical (Chen & Ge, 2007)	10.07%
Medical (Wang et al., 2008)	12.24%
Applied linguistics (Vongpumivitch et al., 2009)	11.17%
Agricultural sciences (Martinez, et al., 2009)	9.06%
Financial services (Li & Qian, 2010)	10.46%
Chemistry (Valipouri & Nassaji, 2013)	9.60%
Thailand University Admission test (Cherngchawano & Jaturapitakkul, 2014)	4.58%
Environmental science (Liu & Han, 2015)	12.82%
Nursing (Yang, 2015)	13.64%

Additionally, when words in the current corpus are compared with Coxhead's AWL, 21 headwords in the top100 of the AWL in the LARA corpus (bold and underlined words in Table 4.6, Chapter 4 (pp. 99-100)) coincided with Coxhead's: such as *adjacent*, *corresponding*, *ethics*, *institute*, *predicted*, *consist*, *domesticated*, *input*, *link*, *minor*, *publication*, *underlie*, *aware*, *beneficial*, *contributes*, *eventually*, *analogous*, *creation*, *error*, *inhibit*, and *professional*. As some AWL words in the LARA corpus appear with high-frequency, the frequency order of those words differs from that in Coxhead's AWL. This supports the argument for the development of the word lists in specific fields based on the target genres and texts that learners have to read and write in their academic branch (Hyland & Tse, 2007; Martinez et al., 2009; Wang et al., 2008; Valipouri & Nassaji, 2013). That is, knowledge of about 20% of

words from the AWL in the LARA corpus is consistent with that found by Coxhead (2000) is necessary for a member in the discourse community of laboratory animals to be able to read and write the academic words of this field.

Moreover, the combination of word families for high-frequency words and word families for the AWL accounts for 88.01%; this is lower than what was found by Coxhead (2000), where the 2000 high frequency words and the AWL words provide 86.1% of text coverage. These results indicate that members in the discourse community of laboratory animal researchers have to take time to learn these further words recognized in the AWL of the corpus, which is very useful for those members.

Furthermore, when the coverage of AWL is compared with the second 1000 high-frequency words in this corpus, it can be seen that the coverage for the AWL (9.48%) is less than half that of the second 1000 high-frequency words (4.36%). This result is different from Coxhead's study. This implies that learners should essentially learn high-frequency words before AWL because high-frequency words are the core words that learners need to learn first (Nation, 2001; Paquot, 2010) and provide the beginnings for the knowledge of other vocabulary including academic words. In other words, the high-frequency words are also valuable for the forms and functions of language in the reading and writing of review or research scientific articles in English.

In brief, 768 content words of AWL were found in the present corpus, computed as 4.36% of text coverage. The AWL for the LARA corpus covers less than Coxhead's and other previous studies discussed above. Nevertheless, the coverage for the LARA corpus is similar to that of the corpus of the Thailand University Admission test, covering 4.58%. The difference is because of the variety of this current study. In other words, the corpus is only the *ILAR Journal*, which differs from previous studies which were comprised of over five related journals. Words from the AWL are also considered practical for learners in this field. In addition, words from Coxhead's AWL are not as frequently used in the laboratory animal field as are the words from the current corpus. The words on the AWL of the laboratory animal field are necessary if members in the discourse community of laboratory animal researchers are to read and write the academic words of this field. Also, the total of 88.01% high-frequency and AWL words is lower than Coxhead's (2000); this shows that it takes time for learners

of laboratory animal science to learn the AWL words in the corpus, which is very useful for them. Finally, a comparison of the coverage of AWL with the second thousand high-frequency words in this corpus demonstrates that the coverage of AWL is more than two times lower than the second thousand high-frequency words, which is unlike what was found for Coxhead's corpus. It is essential for learners to learn high-frequency words first because high-frequency words are the basic words to learn and offer the beginnings of knowledge of other vocabulary including academic words. Importantly, these high-frequency words are used to read and write scientific articles in English in terms of the forms and functions used for the writing of review or research articles.

5.3 Words and Patterns Presenting the Research Purposes

Research question 3: What words and forms signal the function of indicating research purpose in laboratory animal studies?

The investigation of words and patterns used to present the research purposes in the LARA corpus reveals that there were 133 sentences which presented the research purposes in the 160 review articles. The high-frequency content words occurred with the greatest frequency (79.46%). Second was AWL (16.67%). The last was words outside the high-frequency list and the AWL (3.87%). Moreover, the four verb tenses used in the sentences presenting the objectives are the present simple tense, the future simple tense, the past simple tense, and the present perfect tense respectively. Finally, this LARA corpus revealed eight patterns which present the research purposes. Thus, the content words, verb tenses, and patterns presenting the research purposes all are discussed in this section.

5.3.1 Words Signaling Purposes

Words used to signal purpose came from all three word lists for the LARA corpus. In other words, high-frequency content words appeared more often than words from the other lists. Moreover, the finding of 66.67% of words signaling purposes from the high-frequency words list is consistent with West's results (1953), shown in

Table 4.7, Chapter 4 (p. 103); these words include *review, article, discuss, provide, present, describe, purpose, study, examine, intend, report, show, use, address, propose, act, allow, broad, compared, determine, end, gain, give, important, literature, main, mean, offer, reason, and understand*. Furthermore, the underlined words in Table 4.7, Chapter 4 (p. 103) are also consistent with examples in textbooks (Pyrzszak & Bruce, 1998; Swales & Feak, 2012; McMillan, 2012; Soranastaporn, 2013). More than a quarter of purpose words in the HFW (37.78%) are also consistent with examples in the textbooks of Pyrczak and Bruce (1998), Swales and Feak (2012), McMillan (2012), and Soranastaporn (2013). More importantly, the word *review*, the most frequently used word, represents as the objective sentences in review articles. This word can be a noun or a verb, functioning as the subject and the verb in sentences. Researchers use this word to indicate clearly that they want to review the articles. Therefore, these purpose words from the high-frequency word list are considered as the basic words for the purpose sentences in every article. In other words, this finding confirms that it is necessary for L2 learners to recognize these high-frequency words to be able to read and write the purpose sentences in the articles, which provide an initial understanding of the research. Likewise, high-frequency content words are worthy of prudent attention in academic writing (Paquat, 2010). After high-frequency words, academic words are second highest in presenting the objectives; the use of 46.67% of AWL content words for presenting the research purpose is consistent with Coxhead's (2000). These words include *highlight, summarize, briefly, aid, objective, modify, and survey*. The use of five purpose words from the AWL (33.33%) is consistent with examples in the four above-mentioned textbooks; these words are *aim, explore, objective, and extend*. These words are more academic than the words on the high-frequency list. These purpose words from the AWL are useful for evaluating the difficulty of the texts and the targeting words, which is necessary to understand when reading and writing academic articles. Finally, the low-frequency words from words outside high-frequency words and AWL are the least used. Furthermore, when some words from the word lists such as *the present, this, reported, and here* (Swales and Feak, 2012) are used to form purpose sentences, they become reference to the present text. In short, it is very common that words from the high-frequency word list are used the most frequently to present the research

purposes in review articles because they are the basic words used to present the research purposes in every article, no matter whether review or research articles.

5.3.2 Purpose Sentences

Eight patterns were used in the LARA corpus to present the purpose in 133 sentences. The three most frequent patterns were Pattern III, *personal subjects*, Pattern IV, *impersonal subjects*, and Pattern I, *words signaling objectives and to-infinitive*, respectively. Each pattern is consistent with different examples found in the textbooks, as shown in Table 5.2.

Table 5.2 Comparison of the Purpose Patterns Used in the Textbooks of Each Author

	Patterns	Number	%	Consistent with
I	Words signaling purpose and to-infinitive	20	15.04	1) Pyrczak & Bruce (1998) 2) Swales & Feak (2012) 3) McMillan, (2012) 4) Soranastaporn, (2013)
II	Words signaling purpose as the subject and phrases explaining the details of purposes	4	3.00	1) Swales & Feak (2012)
III	Personal subjects	50	37.60	1) Swales & Feak (2012) 2) McMillan, (2012)
IV	Impersonal subjects	45	33.84	1) Swales & Feak (2012) 2) Soranastaporn, (2013)
V	The main verb and to-infinitive	5	3.76	1) Swales & Feak (2012)
VI	The main verb in the passive and to-infinitive	7	5.26	1) Soranastaporn, (2013)
VII	It is + adjective + to-infinitive	1	0.75	-
VIII	Mixed pattern	1	0.75	-
	TOTAL	133	100	

Pattern III occurred the most, as is consistent with examples in the textbooks of Swales and Feak (2012) and McMillan (2012), while the second most used pattern, Pattern IV, is consistent with Swales and Feak (2012) and Soranastaporn (2013). Moreover, Pattern II and Pattern V are consistent with Swales and Feak (2012). Also, Pattern VI is consistent with Soranastaporn (2013). Pattern I, *words signaling objectives and to-infinitive*, is used the third most often and is consistent with examples in all four textbooks; namely, Pyrczak and Bruce (1998), Swales and Feak (2012), McMillan (2012), and Soranastaporn (2013). This confirms that it is the most popular pattern that can be used as a model for writing the purpose of the study. However, the research purposes in review articles make more frequent use of Patterns III and IV (37.60% and 33.84% respectively) in the ILAR journal, which makes up this corpus. Noticeably, four out of eight patterns, Patterns I, V, VI and VII, used the *to-infinitive* to present the purposes, while the rest used phrases or clauses presenting objectives. Furthermore, each pattern includes spatial deixis. The spatial deixis for each pattern is *this...* or *here...*, which refers to the current article. When the pattern includes spatial deixis, personal subjects frequently replace impersonal subjects; the pronoun *we* is frequently used (Swales & Feak, 2012). The personal subject used the most was the first person plural pronoun (*we*). It was in Pattern III, the most used pattern in the LARA corpus. The pronoun *we* refers to the writers as a team. That is to say, the pronoun *we* emphasizes the doers (the people who did the action), while the focus for pattern IV is different. Moreover, all have exclusive semantic reference. Namely, the writers themselves state the purposes of their review articles. The pronoun *we* is also used to indicate the writers' intentions and provide an obvious structure for the texts. Another type of subject in purpose sentences was impersonal subjects which were used more than personal subjects. Writers use impersonal subjects to refer to the research purposes of the articles as a whole. That is, the purpose sentences indicate what is stated or done in the whole article. This subject-type was found in Pattern IV the most often. However, impersonal subjects as words signaling research purpose focus on the objectives of the study. They were found most often in Pattern I. In summary, the two alternatives for expressing the purpose are writing purpose sentences using Pattern III or IV.

5.3.3 Verb-Tenses in Purpose Sentences

In the LARA corpus, the present simple was used the most frequently. These findings are opposite to the examples in the books *Writing Empirical Research Reports* (Pyrzczak & Bruce, 1998) and *Effective Reading & Writing English Text* (Soranastaporn, 2013). Both books state that past simple tense was used in all purpose sentences, but the past tense in the current study was used the third most often. Moreover, in the book, *Educational Research Fundamentals for the Consumer* by McMillion (2012), purpose sentences using the present simple tense, past simple tense, and future simple tense are given as examples. That is to say, the three above-mentioned tenses for purpose sentences in that book are consistent with in the LARA corpus. However, Swales and Feak (2012) explain that the purpose sentences use present or past verbs dependent upon different situations, referring to the type of *text* and *investigation*. In other words, when referring to the type of *text*, purpose sentences use the present tense. The *ILAR Journal* is a peer-reviewed publication, which is the type of *text*, so the purpose sentences in the LARA corpus use the present tense the most. Remarkably, the purpose sentences in the corpus include spatial deixis (*this* or *here*), which refers to the recent articles, so the present simple was used the most. If the sentences refer to an original purpose that has now changed, the past tense is used. That is, some sentences used the past simple tense to denote the original purpose. Importantly, no evidence shows the use of the present perfect to present the research purposes. In brief, the present simple tense should be used to present the research purposes in review articles.

To conclude, purpose sentences are formed by words, verb tenses, and patterns. Words used for the purpose statement should be words from the high-frequency word list because these basic words are used in every article, both review and research articles. The present simple tense is used to describe the purpose sentences in the major of review articles because the results of the current study are the representativeness of the review articles and the evidence according to Swales and Feak (2012). That is, purpose sentences mentioning the type of *text* use the present tense. Finally, Pattern III, *personal subjects*, and Pattern IV, *impersonal subjects*, are appropriate for forming purpose sentences in review articles.

5.4 Implications of the Study

In the current study of high-frequency words, academic words, and the purpose words and patterns in the review articles of laboratory animal field, the results provide empirical inputs for learning and teaching English for academic courses or ESP programs for English for laboratory animal science. Because of the problems of science learners in terms of reading and writing articles or texts in English, these English courses help L2 learners solve the problems. Moreover, researchers in laboratory animal science can benefit from this study. The implications for the use of word lists and the forms of purpose functions can be extended as follows.

1. The findings of the current study are helpful in teaching ESP courses. Teachers can select the words from each word list to teach their students in the field of laboratory animal science. Teachers can teach each word separately or in a relevant context. The pre-teaching of vocabulary is also an effective way to learn English. Moreover, teachers can use the purpose patterns with the words signaling the research purposes in the corpus to teach their students to read and write academic articles or texts. The results can also help English teachers to select or design ESP materials. In other words, English teachers can design the materials based on the word lists.

2. The learners of English for laboratory animal science can also use the words from the word lists to read and write academic articles and texts in laboratory animal science. Especially when they write academic articles, they can also use the purpose forms as well as the words signaling purpose to write purpose sentences. That is, learners can form purpose sentences with the words signaling purpose and purposive patterns correctly; especially, Patterns III and IV, which are suitable for writing the purpose sentences in the review articles.

3. The implications for researchers who are interested in laboratory animal science are that the word lists and forms used for the research purposes in the academic articles of laboratory animal science can be used to read and write academic articles; especially, the purpose function which describes what to be done in the article. In the writing section, forms and functions are necessary for researchers to write well-structured academic texts with the purpose patterns shown in Chapter IV including the words signaling the research purpose such as *review*, *article*, *discuss*, *provide*, and *focus* in the present study.

4. Educators can also take the findings of the current study to produce syllabi. Namely, words from the corpus, both high-frequency and academic words, and the purpose patterns can be included in the syllabus. Moreover, the syllabus expresses what is communicated through language such as presenting the research purposes, research methodology, discussion, or research results. When designing a syllabus, the designer has to consider the words and forms presenting the functions in order to meet the objectives for each unit in the courses.

5. For material designers, the word lists in the current study can be used in developing the materials by considering the frequency of words. That is to say, academic materials can be developed by including high-frequency words plus academic words. Moreover, the purpose patterns can be included in the materials as a model for writing purpose sentences.

6. The results of this study can be used as a guideline for writing purpose sentences with words signaling research purposes in other fields.

In brief, the implications of the recent study are useful for English teachers, learners in related fields, researchers in laboratory animal science, educators, and material designers for ESP courses and as a guideline for writing.

Chapter Summary

To summarize, high-frequency words in this study are responsible for greater coverage than what was found in several previous studies such as the corpora of academic texts (Coxhead, 2000; Hyland & Tse, 2007), psychology, engineering, agriculture science, financial services, and environmental science, but lower than that for the corpus of the Thailand University admission tests. In other words, high frequency words differ according to each corpus due to changes in English language and culture. Moreover, the first thousand words have lower coverage than was found by West (1953), but more than what was found by Coxhead (2000). The proportions for the first and second thousand highest frequency words in the corpus are consistent with previous studies; that is, the first thousand are more frequent than the second. Furthermore, function words in the high-frequency words appear more frequently than

content words because of the high occurrence of the first twenty-three function words in the first thousand. Function words may also be close to lexical words. Therefore, high-frequency words are the best choice for academic study in laboratory animal science. Text coverage provided by words from the AWL is lower than what was found by Coxhead and other previous studies, but nearly equal to the corpus for the Thailand University Admission test covering because the corpus is only the *ILAR Journal*. This corpus is less varied than the corpora in previous studies, which included over five related journals. The words, verb tenses, and patterns used to construct sentences were studied in terms of the function of expressing the research purpose. Words signaling purpose ought to be words from the HFW list because they are the core words used in every type of article. The present simple tense is suitable to use for purpose sentences in review articles because of the evidence of Swales and Feak (2012). Finally, Pattern III *personal subjects*, and Pattern IV, *impersonal subjects*, can be models for purpose sentences in review articles, especially for novices or beginning learners. Eventually, the implications of the study of word lists are useful in selecting materials for English teachers and for materials development for course designers. Moreover, the study supports the use of words and language functions in the design of syllabi for teaching ESP.

CHAPTER VI

CONCLUSION

This recent study focused on vocabulary words and the function of signaling purpose as occurring in the corpus of articles on laboratory animal science. This is because there was no model for such articles in the field of laboratory animal science and no studies of words and language functions in the laboratory animal field. Moreover, there is no academic English course for L2 learners at Mahidol University. This corpus-based study was conducted to fill this gap, to gain knowledge in the field and to help L2 learners in terms of reading and writing scientific academic works.

The current study involved a number of steps. Firstly, the researcher selected the journal, according to the impact factor, the quartile (Q), and Beall's list. The journal provided the data for the laboratory animal corpus. The laboratory animal corpus, consisting of articles published between 2010 and 2014, includes 160 articles on 16 topics. The corpus consists of 840,773 tokens and 22,434 word types. Coxhead's (2000) criteria, including *specialized occurrence*, *range*, and *frequency*, were used to determine which words appeared on the high-frequency word list and the academic word list. Two research instruments were also used to analyze the data: AntConc Version 3.4.4 was used for *frequency* and RANGE was used for *range*. After the word lists had been determined, forms used to signal purpose were analyzed using Microsoft Office Excel 2007. Finally, the framework of AWL by Coxhead (2000) was applied for analysis of the word lists, and analysis of the sentences signaling purpose used the framework of Swales and Feak (2012) and Soranastaporn (2013). Furthermore, frequency and percentage were calculated using Microsoft Office Excel 2007. The results of the study follow:

Research question 1: What hundred content words appear most frequently in laboratory animal review articles?

Among the 840,773 word tokens and 22,434 word types of the present study, there were 703,296 high-frequency words (83.65% of text coverage) including 337,362 content words (40.13% of text coverage). The top-ten content words in this corpus were *studies*, *animal*, *animals*, *research*, *human*, *species*, *used*, *models*, *study*, and *use*, whereas the 100th content word was the word *system*. The number of high-frequency words of the current study was more frequent than of other previous studies due to the change of English language and culture. In other words, each principle provides the different findings.

Research question 2: What content words from the AWL occurred with high frequency in the same corpus of laboratory animal review articles?

The number of content words of AWL covered 4.36% (36,658 content words), which is lower than what was found by Coxhead and other previous studies because of the corpus which was not as varied as the corpora used in previous studies. That is to say, the current corpus comes from only one journal, ILAR, while other corpora included five or more related journals. Moreover, 768 content headwords occurred in the AWL. The top-ten academic words were *regardless*, *surrounding*, *adjacent*, *consider*, *corresponding*, *diagnose*, *dopaminergic*, *drive*, *ethics*, and *explore*. Most importantly, twenty-one content words from the LARA corpus were found to come from the academic word list, which is consistent with word families on Coxhead's AWL such as *adjacent*, *corresponding*, *ethics*, *institute*, *predicted*, *consist*, *domesticated*, *input*, *link*, *minor*, *publication*, *underlie*, *aware*, *beneficial*, *contributes*, *eventually*, *analogous*, *creation*, *error*, *inhibit*, and *professional*.

Research question 3: What words and forms signal the function of indicating research purpose in laboratory animal studies?

The findings for the last question concern the content words and forms signaling purpose found in the corpus. The conclusion is divided into two categories: content words and forms signaling purpose.

Sixty-eight content words were found to signal the purpose. Forty-five content words can be found in the high-frequency word list; 15 in the AWL; and eight are words outside the high-frequency word list and AWL. Most words signifying the purpose are found in the high-frequency word list because such words form the nucleus words for every type of article. The content words, the ten most frequently appearing in the corpus, are found in the high-frequency word list; e.g., *review, article, discuss, provide, focus, present, describe, goal, identify, and purpose*. The ten most frequently appearing content words from the corpus that are found on the AWL are *aim, highlight, summarize, briefly, explore, paper, aid, illustrate objective, and detail*. The content words in the list of words outside the high-frequency word list and the AWL were *essay, text, concentrate, optimize, outline, sketch, underscore, and workshop*.

In addition, eight patterns were found for signaling the purpose in the LARA corpus. The third pattern, *the personal subject*, was found the most frequently, while patterns VII, *It is + adjective + to-infinitive*, was the least-used pattern. Pattern III is the best to use to form sentences signaling purpose in review articles because the corpus is representative of the review articles. Additionally, one purpose sentence consisted of a mixture of Patterns III and IV. More importantly, the present simple was the most frequently used tense for each pattern, whereas the present perfect tenses was used the least. It is very likely that the present simple tense is used for sentences signaling purpose in review articles because of the representativeness of review articles and the evidence of Swales and Feak (2012).

Recommendations for Further Studies

As a follow-up to the findings from the analyses of words and forms signaling purpose in the laboratory animal corpus, further corpus studies should be conducted so as to achieve a more profound understanding and find more beneficial applications as recommended below:

1. The variety of the corpus used in this current study is not wide enough; that is, it consisted of articles from only one journal in the laboratory animal field.

Hence, the sources for the corpus should be expanded; for example, extended to other sources of laboratory animal academic writing such as textbooks, theses, or websites.

2. Differences in the use of language among sub-fields of laboratory animal science should be taken into account.

3. The meanings and parts of speech of the words related to the laboratory animal science from all word lists should be investigated in order to help establish a selection of words suitable for teaching to L2 learners in this field.

4. This recent study was conducted to analyze the words and verb tenses used in statements signaling purpose. Further studies should be conducted regarding other language functions such as presenting the research methodology, data collection, statistics used in the research, the discussion, or research results.

5. Further studies can be conducted in other areas such as political science, communication arts, and architecture.

REFERENCES

- Ahmad, J. (2012). Stylistic features of scientific English: A study of scientific research articles. *English Language and Literature Studies*, (2)1, 47-55.
- Anthony, L. (2006). Developing a freeware, multiplatform corpus analysis toolkit for the technical writing classroom. *IEEE transactions on professional communication*, 49(3), 275-286.
- Anthony, L. (2013). Developing AntConc for a new generation of corpus linguists. *Proceedings of the Corpus Linguistics Conference (CL 2013)*, July 22-26, 2013. Lancaster University, UK, 14-16.
- Anthony, L. (2014). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software/antconc/>.
- Armer, T. (2011). *Cambridge English for scientists*. Cambridge: Cambridge University Press.
- Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: an analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language*, 4(2), 91-105.
- Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Barber, C.L. (1962). Some measurable characteristics of modern scientific prose. In *Contributions to English syntax and philology* (21-43), Goteburg: Acta Universitatis Gothoburgensis.
- Biber, D. (2010). 'What can a corpus tell us about registers and genres?' in A.O'Keeffe and M. McCarthy (Eds.), *The Routledge handbooks of corpus linguistics*. Abingdon: Routledge, 241-54.
- Brotherton, M. (2008). *Science and English*. Retrieved from <http://www.mikebrotherton.com/2008/02/19/science-and-english/>
- Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English*. Cambridge: Cambridge University Press

- Chawwang, N. (2008). *An investigation of English reading problems of Thai 12 th-grade students in Nakhonratchasima educational regions 1, 2, 3, and 7*. M.A. Thesis in English. Srinakharinwirot University.
- Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, 26(4), 502-514.
- Cheng, W. (2012). *Exploring corpus linguistics language in action*. London and New York: Routledge.
- Cherngchawano, W., & Jaturapitakkul, N. (2014). Lexical profiles of Thailand university admission tests. *PASAA*, 48, 1-27.
- Clancy, B. (2010). 'Building a corpus to represent a variety of a language' in A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbooks of corpus linguistics*, Abingdon: Routledge, 80-92.
- Cook, G. (2003). *Applied Linguistics*. Oxford: Oxford University Press.
- Conrad, S. (2010). 'What can a corpus tell us about grammar?' in A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbooks of corpus linguistics*, Abingdon: Routledge, 227-40.
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), 213-238.
- Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4), 389-400.
- Crystal, D. (1991). *A dictionary of linguistics and phonetics*. Oxford, UK; Cambridge, Mass., USA: B. Blackwell.
- Donley, K.M., & Reppen, R. (2001). Using corpus tools to highlight academic vocabulary in SCLT. *TESOL Journal*, 10 (2/3), 7-12.
- Drubin, D.G., & Kellogg, D. R. (2012). English as the universal language of science: opportunities and challenges. *Molecular Biology of the Cell*, 1399-1399.
- Dudley-Evans, T., & St John, M. (1998). *Developments in English for Specific purposes: A multi-disciplinary approach*. Cambridge: Cambridge University Press.
- Elliott, C. (2012). *On predatory publishers: a Q&A with Jeffrey Beall*. Retrieved from <http://chronicle.com/blogs/brainstorm/on-predatory-publishers-a-qa-with-jeffrey-beall/47667>.

- Engber, D. (2013). How did English get to be the international language of science? *Popular Science*. Retrieved from <http://www.popsci.com/article/science/fyi-how-did-english-get-be-international-language-science>.
- Farrell, P. (1990). *Vocabulary in ESP: A lexical analysis of the English of electronics and a study of semi-technical vocabulary*. CLCS Occasional Paper No. 25 Trinity College.
- Flowerdew, J. (2002). *Academic discourse*. Great Britain: Pearson Education Limited.
- Flowerdew, J., & Peacock, M. (2001). *Research perspectives on English for academic purposes*. Cambridge: Cambridge University Press.
- Fukushima, S., Watanabe, Y., Kinjo, Y., Yoshihara, S., & Suzuki, C. (2012). Development of a web-based concordance search system based on a corpus of English papers written by Japanese University students. *Social and Behavioral Sciences*, 34, 54-8.
- Gass, S., & Selinker, L. (2008). *Second language acquisition an introductory course*. (3 nd. ed.). New York: Routledge.
- Ghadessy, M. (1979). Frequency counts, word lists, and materials preparation: a new approach. *English Teaching Forum*, 17, 24-27.
- Glasman-Deal, H. (2010). *Science research writing for non-native speakers of English*. London: Imperial College Press.
- Groom, N., & Littlemore, J. (2011). *Doing applied linguistics a guide for students*. New York: Routledge.
- Halliday, M.A.K. (1993). Some grammatical problems in scientific English. In M.A.K. Halliday and J.R. Martin (Eds.), *Writing Science*. 69-85. London: The Falmer Press.
- Hamel, R. E. (2007). The dominance of English in the international scientific periodical literature and the future of language use in science. *AILA Review*, 20, 53-71.
- Harris, R.P., Helfand, M., Wolff, S.H., et al. (2001). Methods Work Group, Third U.S. Preventive Services Task Force. Current methods of the U.S. Preventive Services Task Force. A review of the process. *Am J Prev Med.*;20 (3 suppl), 21-35.

- Heatley, A., & Nation, P. (1996). *Range [Computer software]*. Wellington, New Zealand: Victoria University of Wellington. Available from <http://www.vuw.ac.nz/lals>
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes: a learning centered approach*. Cambridge: Cambridge University Press.
- Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: John Benjamins.
- Hyland, K. (2006). *English for academic purposes An Advanced resource book*. New York: Routledge.
- Hyland, K. (2007). English for specific purposes: Some influences and impacts. In J. Cummins and C. Davison (Eds.), *International Handbooks of English Language Teaching*. 391-402. New York: Springer.
- Hyland, K., & Tse, P. (2007). Is there an academic vocabulary? *TESOL Quarterly* 41: 235-253.
- Jordan, R.R. (1997). *English for academic purposes a guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Kaewphanngam, C. (2002). *A corpus analysis of psychology texts as a basis for the development of teaching materials in English for academic purposes*. M.A. Dissertation. Mahidol University. Thailand.
- Kearney, M. H. (2015). Predatory publishing: what authors need to know. *Research in Nursing & Health*, 38, 1–3.
- Kennedy, C., & Bolitho, R. (1984). *English for specific purposes*. London: Macmillan.
- Kilgarriff, A. (1997). Putting frequencies in the dictionary. *International Journal of Lexicography*, 10, 135-155.
- Lamy M-N., & Klarskov Mortensen H. J. (2012) Using concordance programs in the modern foreign languages classroom. Module 2.4 in Davies G. (Ed.), *Information and Communications Technology for Language Teachers (ICT4LT)*, Slough, Thames Valley University [Online]. Available at: http://www.ict4lt.org/en/en_mod2-4.htm [Accessed 24 July 2015].

- Li, Y., & Qian, D. (2010). Profiling the academic word list (AWL) in a financial corpus. *SYSTEM*, 38, 402-411.
- Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, 39, 1-11.
- Lynn, R. W. (1973). Preparing word lists: a suggested method. *RELC Journal*. 4, 25-32.
- Martin, A. V. (1976). Teaching academic vocabulary to foreign graduate students. *TESOL Quarterly*, 10, 91-97.
- Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, 28 (3), 183-198.
- Master, P. (1991). Active verbs with inanimate subjects in scientific pros. *English for specific purposes*, 10, 15-33.
- Master, P. (2003). Noun compounds and compressed definitions. *English Teaching Forum*, 41(3), 2-9.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd. ed.). Edinburgh: Edinburgh University Press.
- McMillian, J.H. (2012). *Educational research: fundamentals for the consumer* (6th. ed.). Boston, MA: Pearson Education
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage Publications, Inc.
- Montgomery, S., & Crystal, D. (2013). *Does science need a global language?: English and the future of research*. Chicago, IL: University Of Chicago Press.
- Moon, R. (2010). 'What can a corpus tell us about lexis?' in A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbooks of corpus linguistics*, Abingdon: Routledge, 197-211.
- Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25, 235-256.
- Murphy, M., L. (2010). *Lexicon meaning*. Cambridge: Cambridge University Press.
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

- Nation, I. S. P., & Coxhead, A. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew and M. Peacock (Eds.), *Research perspectives on English for academic purposes* (252-267). Cambridge: Cambridge University Press.
- Nation, I. S. P., & Hwang, K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23, 35-41.
- Oakes, M.P., & Netliberary Inc. (1998). Statistics for corpus linguistics. *Edinburgh textbooks in empirical linguistics*. Edinburgh: Edinburgh University Press.
- Paltridge, B., & Starfield, S. (2013). *The handbook of English for specific purposes*. Malaysia. John Wiley&Sons.
- Paquot, M. (2010). *Academic vocabulary in learn writing from extraction to analysis*. London: Continuum International Publishing Group.
- Para, C. (2004). *A corpus study of high-frequency words in civil engineering research article: Sub-disciplinary difference between structure and transportation*. M.A. Dissertation. Mahidol University. Thailand.
- Parkinson, J. (2000). Acquiring scientific literacy through content and genre: a theme-based language course for science students. *English for Specific Purposes*, 19,369-387
- Pawapatcharandom, R. (2007). *An investigation of Thai students' English language problems and their learning strategies in the international program at Mahidol University*. M.A. Thesis in English for Business and Industry. King Mongkut's Institute of Technology North Bangkok.
- Pearson, J. (1998). *Terms in context*. Amsterdam: John Benjamins Publishing Company.
- Ponmanee, S., & Sinsuwan, P. (2001). *Needs and problems in English usage of graduate students in teaching Thai and social science compared with students in teaching English*. Unpublished master's thesis, Chiang Mai University, Bangkok, Thailand.
- Pueyo, I.G., & Val, S. (1996). The construction of technicality in the field of plastics: A functional approach towards teaching technical terminology. *English for Specific Purposes*, 15, 251-78.

- Pyrzczak, F., & Bruce, R.R. (1998). *Writing empirical research reports a basic guide for students of social and behavioral sciences*. (2nd. ed.). CA: Pyrczak Publishing.
- Rethlefsen, M.L. (2014). Engaging medical liberians to improve the quality of review article. *JAMA*, 312(10), 999-1,000.
- Robson, C. (2011). *Real world research: A resource for users of social research methods in applied settings* (3rd. ed.). Chichester, West Sussex: Wiley.
- Rühlemann, C. (2010). 'What can a corpus tell us about pragmatics?' in A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbooks of corpus linguistics*, Abingdon: Routledge, 281-301.
- Russell, W.M.S., & Burch, R.L., (1959). *The principles of humane experimental technique*. Methuen: London.
- Salager-Meyer, F. (1992). A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes*, 11, 93-113.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- SCImago. (2007). *SJR — SCImago journal & country rank*. Retrieved July 04, 2015, from <http://www.scimagojr.com>
- Sinclair, J. McH. (1991). *Corpus concordance collocation*. Oxford: Oxford University Press.
- Sinclair, J. McH. (2005). 'Corpus and text-basic principles', in M. Wynne (Ed.), *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow Book.
- Siwek, J., Gourlay, M.L., Slawson D.C., & Shaughnessy, A.F. (2002). How to write an evidence-based clinical review article. *Am Fam Phys.*, 65(2), 251-258.
- Soranastaporn, S. (2013). *Effective reading and writing English texts*. Bangkok, Wongaugorn Press.
- Sprenger, M. (2013). *Teaching the critical vocabulary of the common core*. Alexandria, Virginia: ASCD.

- Stubbs, M. (1986). Language development, lexical competence and nuclear vocabulary, in K. Durkin (Ed.), *Language Development in the School Years*. London: Croom Helm.
- Swales, J. M. (2010). 'Meta-talk in American academic talk: The cases of "point" and "thing"', *Journal of English Linguistics*, 29: 34-54.
- Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: essential tasks and skills*. (3 rd. ed.)Ann Arbor: University of Michigan Press.
- Tallerman, M. (2005). *Understanding syntax*. (2nd. ed.). London: Malta for Hooder Arnold.
- Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1981). On the use of the passive in two astrophysics papers. In Swales, J. (Ed.). *Episode in ESP*. Prentice Hall.
- Taylor, R.B. (2011). *Medical writing: a guide for clinicians, educators, and researchers*.doi: 10.1007/978-1-4419-8234-6_6
- The National Academic of Sciences. (2011). *The guide for the care and use of laboratory animals*. (8th. ed.). Washington, D.C.: The National Academies Press.
- Thomson, R. (2015). *Journal citation reports*. Available on line: <http://thomsonreuters.com/en/products-services/scholarly-scientific-research/research-management-and-evaluation/journal-citation-reports.html>.
- Tribble, C., & Jones, G. (1990). *Concordances in the classroom*. London: Longman.
- Tribble, C. (2012). *Teaching and language corpora: quo vadis?* 10th Teaching and Language Corpora Conference (TALC). Warsaw, 11th-14th July 2012.
- Trimble, L. (1985). *English for science and technology: A discourse approach*. Cambridge: Cambridge University Press.
- Valipouri, L., & Nassaji, H. (2013). A Corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, 12(4), 248-263.
- Vongpumivitch, V., Huang, J., & Chang, Y. (2009). Frequency analysis of the words in the academic word list (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33-41.

- Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27(4), 442-458.
- Watt, R. J.C. (2015). Concordance: Software for concordancing and text analysis. (n.d.). Retrieved July 25, 2015.
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28, 46-65.
- Webb, S., & Nation, P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, 16, 1-10.
- West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.
- Wood, A. (2001). International scientific English: The language of research scientists around the world. In J. Flowerdew & M. Peacock (eds.), *Research perspectives on English for academic purposes*. Cambridge, Cambridge University Press.
- Xia, J., Harmon, J. L., Connolly, K. G., Donnelly, R. M., Anderson, M. R., & Howard, H. A. (2014). Who publishes in “predatory” journals?. *Journal of the Association for Information Science and Technology*: n/a.
- Yang, M. (2015). A nursing academic word list. *English for Specific Purposes*, 37, 27-38.

Article References

- Alvarez, C.E. (2010). Naturally occurring cancers in dogs: insights for translational genetics and medicine. *ILAR Journal*, 2(1), 16-35.
- Bailey, M.R., Rich, B. A., & Bennett, B.T. (2010). Crisis planning to manage risks posed by animal rights extremists. *ILAR Journal*, 4(2), 138-148.
- Ball, G.F., & Balthazart, J. (2010). Japanese quail as a model system for studying the neuroendocrine control of reproductive and social behaviors. *ILAR Journal*, 1(4), 310-325.
- Bateson, M., & Feenders, G. (2010). The use of passerine bird species in laboratory research: implications of basic biology for husbandry and welfare. *ILAR Journal*, 8(4), 394-408

- Cooper, J.E.. (2011). Anesthesia, Analgesia, and Euthanasia of Invertebrates. *ILAR Journal*, 6(2), 196-204.
- Decker, J.D. et al. (2010). Understanding Risk Perceptions to Enhance Communication about Human-Wildlife Interactions and the Impacts of Zoonotic Disease. *ILAR Journal*, 7(3), 255-261.
- Du, X.Y. et al. (2011). Characteristics of Circle of Willis Variations in the Mongolian Gerbil and a Newly Established Ischemia-Prone Gerbil Group. *ILAR Journal*, 12, e1-e7.
- Gargiulo, S. et al. (2012). Mice Anesthesia, Analgesia, and Care, Part I: Anesthetic Considerations in Preclinical Research. *ILAR Journal*, 11.
- Garzón-Muvdi, T., & Quiñones-Hinojosa, A. (2010). Neural Stem Cell Niches and Homing: Recruitment and Integration into Functional Tissues. *ILAR Journal*, 3-23.
- Goodwin, B.S., & Donaho, I.C. (2010). Tropical Storm and Hurricane Recovery and Preparedness Strategies. *ILAR Journal*, 1(2), 104-119.
- Gordeladze, J.O. et al. (2010). From Stem Cells to Bone: Phenotype Acquisition, Stabilization, and Tissue Engineering in Animal Models. *ILAR Journal*, 3, 42-61.
- Gruen, L., Fultz, A., & Pruetz, J. (2013). Ethical Issues in African Great Ape Field Studies. *ILAR Journal*, 3(1), 24-32.
- Ho, S. et al. (2012). Environmental Epigenetics and Its Implication on Disease Risk and Health Outcomes. *ILAR Journal*, 6(3), 289-305.
- Joers, V.L., & Emborg, M.E. (2010). Preclinical Assessment of Stem Cell Therapies for Neurological Diseases. *ILAR Journal*, 2, 24-41.
- Kalmar, I.D., Janssens, G.P.J., & Moons, C.P.H. (2010). Guidelines and Ethical Considerations for Housing and Management of Psittacine Birds Used in Research. *ILAR Journal*, 9(4), 409-423.
- Kranenburg, H.C., Hazewinkel, H.A.W., & Meij, B.P. (2014). Naturally Occurring Spinal Hyperostosis in Dogs as a Model for Human Spinal Disorders. *ILAR Journal*, 9, 150-163.

- Locatelli, P. et al. (2011). An Ovine Model of Postinfarction Dilated Cardiomyopathy in Animals with Highly Variable Coronary Anatomy. *ILAR Journal*, 14, 16-21.
- Lund, K.E., Maloney, S.K., Milton, J.T.B., & Blache, D. (2012). Gradual Training of Alpacas to the Confinement of Metabolism Pens Reduces Stress When Normal Excretion Behavior Is Accommodated. *ILAR Journal*, 8.
- Lutz, C.K. (2014). Stereotypic Behavior in Nonhuman Primates as a Model for the Human Condition. *ILAR Journal*, 6(2), 284-296.
- Mansfield, K.G., Riley, L.K., & Kent, M.L. (2010). Workshop Summary: Detection, Impact, and Control of Specific Pathogens in Animal Resource Facilities. *ILAR Journal*, 8(2), 171-179.
- Margoliash, P. (2010). Sleep, Learning, and Birdsong. *ILAR Journal*, 6(4), 378-386.
- Michaelides, M., Thanos, P.K., Volkow, N. D., & Wang, G. (2012). Translational Neuroimaging in Drug Addiction and Obesity. *ILAR Journal*, 5, 59-68.
- Minteer, B.A., & Collins, J.P. (2010). Ecological Ethics in Captivity: Balancing Values and Responsibilities in Zoo and Aquarium Research under Rapid Global Change. *ILAR Journal*, 5(1), 41-51.
- Niculescu, M.D. (2012). Nutritional Epigenetics. *ILAR Journal*, 4(3), 270-278.
- O'Brien, P.D., Sakowski, S.A., & Feldman, E.L. (2014). Mouse Models of Diabetic Neuropathy. *ILAR Journal*, 259-272.
- Palley, L.S., O'Rourke, P.P., & Niemi, S.M. (2010). Mainstreaming Animal-Assisted Therapy. *ILAR Journal*, 1(3), 199-207.
- Patterson-Kane, J.C., & Rich, T. (2014). Achilles Tendon Injuries in Elite Athletes: Lessons in Pathophysiology from Their Equine Counterparts. *ILAR Journal*, 6, 86-99.
- Rao, D.B., Jortner, B.S., & Sills, R.C. (2014). Animal Models of Peripheral Neuropathy Due to Environmental Toxicants. *ILAR Journal*, 7(3), 315-323.
- Rooney, T.M., & Freeman, M.R. (2014). Drosophila Models of Neuronal Injury. *ILAR Journal*, 4(3), 291-295.
- Sanders, J.L., Watral, V., & Kent, M.L. (2012). Microsporidiosis in Zebrafish Research Facilities. *ILAR Journal*, 2(2), 106-113.

- Schaevitz, L.R., & Berger-Sweeney, J.E. (2012). Gene–Environment Interactions and Epigenetic Pathways in Autism: The Importance of One-Carbon Metabolism. *ILAR Journal*, 8(3), 322-340.
- Silk, S.B., Hampton, L.L. & Brown, P.A. (2014). What Investigators Need to Know About the Use of Animals. *ILAR Journal*, 8(3), 324-328.
- Smith, S.A. (2012). Zebrafish Resources on the Internet. *ILAR Journal*, 12(2), 208-214.
- Spitsbergen, J. M., Buhler, D.R., & Peterson, T.S. (2012). Neoplasia and Neoplasm-Associated Lesions in Laboratory Colonies of Zebrafish Emphasizing Key Influences of Diet and Aquaculture System Design. *ILAR Journal*, 3, 114-125.
- Steukers, L. et al. (2012). Kinetics of BoHV-1 Dissemination in an in Vitro Culture of Bovine Upper Respiratory Tract Mucosa Explants. *ILAR Journal*.
- Tardif, S.D., Coleman, K., Hobbs, T.R., & Lutz, C. (2013). IACUC Review of Nonhuman Primate Research. *ILAR Journal*, 12(2), 234-245.
- Truman, R.W. et al. (2014). The Armadillo as a Model for Peripheral Neuropathy in Leprosy. *ILAR Journal*, 6(3), 304-314.
- Vorhees, C.V., & Williams, M.T. (2014). Assessing Spatial Learning and Memory in Rodents. *ILAR Journal*, 8(2), 310-332.
- Wall, J.D. (2013). Great Ape Genomics. *ILAR Journal*, 1(2), 82-90.
- Withrow, S.J., & Wilkins, R.M. (2010). Cross Talk from Pets to People: Translational Osteosarcoma Treatments. *ILAR Journal*, 2(3), 208-213.
- Yuan, R. Peters, L.L., & Paigen, B. (2011). Mice as a Mammalian Model for Research on the Genetics of Aging. *ILAR Journal*, 4-15.

APPENDICES

APPENDIX A
THE FIRST-1,000 HIGH-FREQUENCY WORD LIST IN LARA
CORPUS

Rank	HFW	f	%	Rank	HFW	f	%
1	the	45493	5.411	32	has	2506	0.298
2	of	36837	4.381	33	which	2414	0.287
3	and	30618	3.642	34	was	2346	0.279
4	in	26820	3.190	35	also	2340	0.278
5	to	21576	2.566	36	other	2253	0.268
6	a	14481	1.722	37	human	2252	0.268
7	that	10177	1.210	38	but	2187	0.260
8	is	9607	1.143	39	such	2180	0.259
9	for	9598	1.142	40	more	2152	0.256
10	with	7887	0.938	41	were	2150	0.256
11	are	7272	0.865	42	their	2024	0.241
12	as	6586	0.783	43	species	1954	0.232
13	or	5916	0.704	44	used	1646	0.196
14	be	5549	0.660	45	we	1619	0.193
15	by	4451	0.529	46	models	1596	0.190
16	on	4335	0.516	47	study	1571	0.187
17	have	4273	0.508	48	than	1547	0.184
18	this	4011	0.477	49	use	1536	0.183
19	from	3659	0.435	50	humans	1529	0.182
20	an	3387	0.403	51	one	1521	0.181
21	not	3339	0.397	52	model	1507	0.179
22	these	3314	0.394	53	they	1491	0.177
23	studies	3031	0.361	54	disease	1488	0.177
24	animal	2942	0.350	55	both	1438	0.171
25	animals	2868	0.341	56	there	1423	0.169
26	it	2799	0.333	57	most	1398	0.166
27	can	2791	0.332	58	cells	1355	0.161
28	been	2705	0.322	59	between	1328	0.158
29	research	2703	0.321	60	effects	1318	0.157
30	at	2569	0.306	61	after	1284	0.153
31	may	2527	0.301	62	some	1273	0.151

Rank	HFW	f	%	Rank	HFW	f	%
63	all	1258	0.150	103	changes	824	0.098
64	when	1236	0.147	104	high	802	0.095
65	because	1215	0.145	105	laboratory	802	0.095
66	mice	1195	0.142	106	system	802	0.095
67	development	1177	0.140	107	clinical	798	0.095
68	many	1158	0.138	108	number	791	0.094
69	cell	1139	0.135	109	if	789	0.094
70	during	1119	0.133	110	through	785	0.093
71	associated	1117	0.133	111	example	782	0.093
72	although	1102	0.131	112	treatment	781	0.093
73	will	1101	0.131	113	those	770	0.092
74	however	1080	0.128	114	brain	761	0.091
75	genetic	1054	0.125	115	behavioral	754	0.090
76	well	1041	0.124	116	its	754	0.090
77	gene	1038	0.123	117	similar	754	0.090
78	specific	1030	0.123	118	drug	748	0.089
79	only	1022	0.122	119	related	743	0.088
80	behavior	994	0.118	120	data	742	0.088
81	levels	989	0.118	121	new	735	0.087
82	expression	978	0.116	122	about	733	0.087
83	into	953	0.113	123	thus	733	0.087
84	response	942	0.112	124	control	729	0.087
85	stress	934	0.111	125	based	711	0.085
86	time	923	0.110	126	risk	707	0.084
87	important	910	0.108	127	found	702	0.083
88	using	907	0.108	128	social	696	0.083
89	rats	906	0.108	129	significant	694	0.083
90	exposure	902	0.107	130	early	693	0.082
91	dogs	894	0.106	131	could	692	0.082
92	two	886	0.105	132	factors	685	0.081
93	should	882	0.105	133	differences	665	0.079
94	including	879	0.105	134	group	662	0.079
95	genes	867	0.103	135	several	662	0.079
96	water	865	0.103	136	often	658	0.078
97	no	864	0.103	137	reported	657	0.078
98	zebrafish	860	0.102	138	each	654	0.078
99	different	857	0.102	139	first	647	0.077
100	health	855	0.102	140	potential	646	0.077
101	increased	840	0.100	141	cancer	635	0.076
102	results	825	0.098	142	long	613	0.073

Rank	HFW	f	%	Rank	HFW	f	%
143	within	603	0.072	183	shown	505	0.060
144	years	602	0.072	184	mouse	503	0.060
145	age	598	0.071	185	evidence	502	0.060
146	infection	595	0.071	186	available	497	0.059
147	fish	594	0.071	187	possible	497	0.059
148	even	593	0.071	188	small	497	0.059
149	food	591	0.070	189	do	496	0.059
150	effect	590	0.070	190	known	494	0.059
151	would	580	0.069	191	individual	488	0.058
152	include	578	0.069	192	days	487	0.058
153	large	575	0.068	193	so	485	0.058
154	role	574	0.068	194	body	484	0.058
155	likely	571	0.068	195	genome	483	0.057
156	same	569	0.068	196	recent	483	0.057
157	had	566	0.067	197	how	477	0.057
158	common	561	0.067	198	conditions	476	0.057
159	monkeys	561	0.067	199	individuals	475	0.056
160	macaques	558	0.066	200	administration	473	0.056
161	addition	557	0.066	201	among	472	0.056
162	any	557	0.066	202	level	471	0.056
163	activity	556	0.066	203	complex	469	0.056
164	diseases	555	0.066	204	must	469	0.056
165	provide	554	0.066	205	epigenetic	463	0.055
166	responses	552	0.066	206	our	463	0.055
167	experimental	550	0.065	207	observed	462	0.055
168	less	547	0.065	208	patients	460	0.055
169	methylation	547	0.065	209	groups	458	0.054
170	information	544	0.065	210	life	455	0.054
171	induced	538	0.064	211	showed	455	0.054
172	pain	536	0.064	212	cocaine	454	0.054
173	environmental	535	0.064	213	further	449	0.053
174	compared	534	0.064	214	higher	447	0.053
175	analysis	524	0.062	215	increase	447	0.053
176	over	524	0.062	216	either	437	0.052
177	systems	522	0.062	217	learning	434	0.052
178	colleagues	521	0.062	218	intake	432	0.051
179	mechanisms	518	0.062	219	male	432	0.051
180	care	515	0.061	220	review	431	0.051
181	function	513	0.061	221	result	428	0.051
182	rhesus	513	0.061	222	under	426	0.051

Rank	HFW	f	%	Rank	HFW	f	%
223	very	426	0.051	263	ability	379	0.045
224	before	424	0.050	264	females	378	0.045
225	three	424	0.050	265	normal	375	0.045
226	understanding	424	0.050	266	reduced	375	0.045
227	being	421	0.050	267	might	374	0.044
228	single	420	0.050	268	wild	374	0.044
229	another	419	0.050	269	work	374	0.044
230	environment	419	0.050	270	withdrawal	373	0.044
231	protein	416	0.049	271	greater	372	0.044
232	involved	413	0.049	272	much	371	0.044
233	where	413	0.049	273	tumor	371	0.044
234	tissue	410	0.049	274	wildlife	371	0.044
235	low	407	0.048	275	molecular	368	0.044
236	day	404	0.048	276	procedures	368	0.044
237	behaviors	402	0.048	277	blood	365	0.043
238	variation	402	0.048	278	cases	363	0.043
239	birds	400	0.048	279	them	363	0.043
240	then	400	0.048	280	due	360	0.043
241	therefore	399	0.047	281	experience	360	0.043
242	strains	398	0.047	282	given	360	0.043
243	whereas	398	0.047	283	effective	359	0.043
244	regions	396	0.047	284	diet	358	0.043
245	female	395	0.047	285	methods	357	0.042
246	show	395	0.047	286	self	356	0.042
247	multiple	394	0.047	287	weight	356	0.042
248	whether	394	0.047	288	aging	353	0.042
249	identified	393	0.047	289	bone	350	0.042
250	population	393	0.047	290	state	349	0.042
251	researchers	392	0.047	291	various	348	0.041
252	type	392	0.047	292	males	347	0.041
253	described	389	0.046	293	neurons	347	0.041
254	primates	389	0.046	294	period	347	0.041
255	approach	388	0.046	295	term	347	0.041
256	like	388	0.046	296	without	346	0.041
257	virus	387	0.046	297	does	342	0.041
258	adult	386	0.046	298	field	342	0.041
259	major	386	0.046	299	primate	340	0.040
260	populations	386	0.046	300	significantly	340	0.040
261	growth	384	0.046	301	infections	339	0.040
262	size	382	0.045	302	demonstrated	337	0.040

Rank	HFW	f	%	Rank	HFW	f	%
303	rodents	337	0.040	343	range	311	0.037
304	consumption	336	0.040	344	did	309	0.037
305	experiments	336	0.040	345	provided	306	0.036
306	loss	336	0.040	346	veterinary	306	0.036
307	present	336	0.040	347	quality	305	0.036
308	functional	335	0.040	348	up	305	0.036
309	process	335	0.040	349	addiction	304	0.036
310	who	335	0.040	350	knowledge	304	0.036
311	affect	333	0.040	351	sequence	303	0.036
312	welfare	333	0.040	352	need	302	0.036
313	findings	332	0.039	353	while	301	0.036
314	limited	332	0.039	354	agents	297	0.035
315	across	331	0.039	355	access	296	0.035
316	chronic	330	0.039	356	infected	296	0.035
317	highly	329	0.039	357	lower	295	0.035
318	natural	329	0.039	358	positive	295	0.035
319	nonhuman	329	0.039	359	sex	295	0.035
320	alcohol	328	0.039	360	useful	294	0.035
321	housing	328	0.039	361	patterns	292	0.035
322	particular	328	0.039	362	critical	291	0.035
323	particularly	328	0.039	363	weeks	291	0.035
324	test	328	0.039	364	additional	290	0.034
325	maternal	327	0.039	365	least	290	0.034
326	cause	326	0.039	366	impact	289	0.034
327	factor	325	0.039	367	production	289	0.034
328	drugs	322	0.038	368	area	288	0.034
329	immune	322	0.038	369	occur	288	0.034
330	support	322	0.038	370	developed	287	0.034
331	types	322	0.038	371	trials	287	0.034
332	disorders	320	0.038	372	change	286	0.034
333	receptor	319	0.038	373	nerve	286	0.034
334	rodent	316	0.038	374	typically	286	0.034
335	anxiety	315	0.037	375	canine	285	0.034
336	areas	314	0.037	376	recently	284	0.034
337	testing	314	0.037	377	contrast	282	0.034
338	considered	313	0.037	378	exposed	282	0.034
339	breeding	312	0.037	379	states	282	0.034
340	develop	312	0.037	380	case	281	0.033
341	required	312	0.037	381	few	280	0.033
342	necessary	311	0.037	382	produce	280	0.033

Rank	HFW	f	%	Rank	HFW	f	%
383	decreased	279	0.033	423	form	259	0.031
384	pet	279	0.033	424	infectious	259	0.031
385	proteins	279	0.033	425	peripheral	259	0.031
386	investigators	278	0.033	426	later	258	0.031
387	training	278	0.033	427	suggest	258	0.031
388	general	277	0.033	428	following	257	0.031
389	iacuc	277	0.033	429	nhps	257	0.031
390	identify	277	0.033	430	biology	256	0.030
391	invertebrates	276	0.033	431	reduce	256	0.030
392	approximately	275	0.033	432	established	255	0.030
393	facilities	275	0.033	433	part	255	0.030
394	therapy	275	0.033	434	primary	255	0.030
395	lines	274	0.033	435	become	253	0.030
396	symptoms	274	0.033	436	depression	252	0.030
397	biomedical	272	0.032	437	novel	252	0.030
398	difficult	272	0.032	438	months	250	0.030
399	rate	272	0.032	439	people	250	0.030
400	survival	272	0.032	440	imaging	249	0.030
401	offspring	271	0.032	441	second	249	0.030
402	tissues	271	0.032	442	interactions	247	0.029
403	appropriate	270	0.032	443	neural	247	0.029
404	fat	270	0.032	444	great	246	0.029
405	others	270	0.032	445	consistent	245	0.029
406	memory	269	0.032	446	dose	244	0.029
407	require	268	0.032	447	injury	244	0.029
408	lead	267	0.032	448	management	244	0.029
409	nicotine	267	0.032	449	processes	244	0.029
410	out	267	0.032	450	sequencing	244	0.029
411	scientific	267	0.032	451	anesthesia	243	0.029
412	relatively	266	0.032	452	damage	243	0.029
413	ethanol	265	0.032	453	dependent	243	0.029
414	genomic	265	0.032	454	determine	243	0.029
415	better	264	0.031	455	ketamine	243	0.029
416	presence	264	0.031	456	relevant	243	0.029
417	per	263	0.031	457	seen	243	0.029
418	macaque	261	0.031	458	fear	242	0.029
419	subjects	261	0.031	459	made	242	0.029
420	especially	260	0.031	460	hours	240	0.029
421	what	260	0.031	461	variety	240	0.029
422	above	259	0.031	462	host	239	0.028

Rank	HFW	f	%	Rank	HFW	f	%
463	lack	238	0.028	503	now	221	0.026
464	song	238	0.028	504	make	220	0.026
465	design	237	0.028	505	regulation	220	0.026
466	exhibit	237	0.028	506	value	220	0.026
467	future	237	0.028	507	genetics	219	0.026
468	studied	236	0.028	508	assess	218	0.026
469	four	235	0.028	509	performed	218	0.026
470	acute	234	0.028	510	wide	218	0.026
471	generally	234	0.028	511	approaches	217	0.026
472	biological	233	0.028	512	facility	217	0.026
473	commonly	233	0.028	513	neuropathy	217	0.026
474	importance	232	0.028	514	staff	217	0.026
475	negative	232	0.028	515	stem	217	0.026
476	yet	232	0.028	516	therapeutic	217	0.026
477	affected	231	0.027	517	association	216	0.026
478	lesions	231	0.027	518	us	216	0.026
479	physiological	231	0.027	519	aspects	215	0.026
480	rather	230	0.027	520	conservation	215	0.026
481	influence	229	0.027	521	essential	215	0.026
482	muscle	229	0.027	522	expressed	215	0.026
483	rates	229	0.027	523	focus	215	0.026
484	characterized	228	0.027	524	national	215	0.026
485	short	227	0.027	525	context	213	0.025
486	tests	227	0.027	526	place	213	0.025
487	traits	227	0.027	527	rat	213	0.025
488	certain	226	0.027	528	samples	213	0.025
489	standard	226	0.027	529	subsequent	213	0.025
490	still	226	0.027	530	dog	211	0.025
491	measures	225	0.027	531	method	211	0.025
492	requires	225	0.027	532	united	211	0.025
493	transmission	224	0.027	533	cues	210	0.025
494	developmental	222	0.026	534	techniques	210	0.025
495	old	222	0.026	535	world	210	0.025
496	signaling	222	0.026	536	pathways	209	0.025
497	viral	222	0.026	537	interest	208	0.025
498	activation	221	0.026	538	target	208	0.025
499	cognitive	221	0.026	539	transgenic	208	0.025
500	furthermore	221	0.026	540	assessment	207	0.025
501	histone	221	0.026	541	cage	207	0.025
502	motor	221	0.026	542	neuronal	207	0.025

Rank	HFW	f	%	Rank	HFW	f	%
543	acid	206	0.025	583	despite	196	0.023
544	direct	206	0.025	584	features	196	0.023
545	efforts	206	0.025	585	platform	196	0.023
546	key	206	0.025	586	recovery	196	0.023
547	occurs	206	0.025	587	sensory	195	0.023
548	region	206	0.025	588	stimuli	195	0.023
549	source	206	0.025	589	larger	194	0.023
550	strategies	206	0.025	590	liver	194	0.023
551	current	205	0.024	591	defined	193	0.023
552	derived	205	0.024	592	metabolic	193	0.023
553	ethical	205	0.024	593	rapid	193	0.023
554	increases	205	0.024	594	way	193	0.023
555	needed	205	0.024	595	children	192	0.023
556	phenotype	205	0.024	596	provides	192	0.023
557	public	205	0.024	597	swine	192	0.023
558	reports	205	0.024	598	little	191	0.023
559	fact	204	0.024	599	metabolism	191	0.023
560	site	204	0.024	600	vivo	191	0.023
561	strain	203	0.024	601	appear	190	0.023
562	suggesting	203	0.024	602	contribute	190	0.023
563	clear	202	0.024	603	light	190	0.023
564	complete	202	0.024	604	central	189	0.022
565	lifespan	202	0.024	605	since	189	0.022
566	outcomes	202	0.024	606	tested	189	0.022
567	similarly	202	0.024	607	events	188	0.022
568	tumors	202	0.024	608	members	188	0.022
569	against	201	0.024	609	resources	188	0.022
570	indicate	201	0.024	610	binding	187	0.022
571	play	201	0.024	611	death	187	0.022
572	culture	200	0.024	612	extended	187	0.022
573	housed	200	0.024	613	obesity	187	0.022
574	usually	200	0.024	614	power	187	0.022
575	protocol	199	0.024	615	treatments	187	0.022
576	selection	199	0.024	616	viruses	187	0.022
577	developing	198	0.024	617	abuse	185	0.022
578	order	198	0.024	618	combination	185	0.022
579	physical	198	0.024	619	preclinical	185	0.022
580	medical	197	0.023	620	reduction	185	0.022
581	potentially	197	0.023	621	sequences	185	0.022
582	resulting	197	0.023	622	cellular	184	0.022

Rank	HFW	f	%	Rank	HFW	f	%
623	mammalian	184	0.022	663	induce	174	0.021
624	mediated	184	0.022	664	procedure	174	0.021
625	increasing	183	0.022	665	specifically	174	0.021
626	numerous	183	0.022	666	basis	173	0.021
627	protocols	183	0.022	667	frequency	173	0.021
628	concerns	182	0.022	668	influenza	173	0.021
629	involving	182	0.022	669	naturally	173	0.021
630	subject	182	0.022	670	numbers	173	0.021
631	underlying	182	0.022	671	characteristics	172	0.020
632	criteria	181	0.022	672	help	172	0.020
633	signs	181	0.022	673	insulin	172	0.020
634	vitro	181	0.022	674	pattern	172	0.020
635	genetically	180	0.021	675	toward	172	0.020
636	initial	180	0.021	676	appears	171	0.020
637	local	180	0.021	677	parameters	171	0.020
638	occurring	180	0.021	678	sites	171	0.020
639	severe	180	0.021	679	alterations	170	0.020
640	extinction	179	0.021	680	phase	170	0.020
641	mammals	179	0.021	681	report	170	0.020
642	requirements	179	0.021	682	suggests	170	0.020
643	throughout	179	0.021	683	best	169	0.020
644	analyses	178	0.021	684	diets	169	0.020
645	issues	178	0.021	685	live	169	0.020
646	relative	178	0.021	686	noted	169	0.020
647	act	177	0.021	687	temperature	169	0.020
648	controls	177	0.021	688	anesthetic	168	0.020
649	unique	177	0.021	689	conducted	168	0.020
650	activities	176	0.021	690	set	168	0.020
651	overall	176	0.021	691	cannot	167	0.020
652	contact	175	0.021	692	pathogens	167	0.020
653	directly	175	0.021	693	attention	166	0.020
654	distress	175	0.021	694	course	166	0.020
655	end	175	0.021	695	currently	166	0.020
656	moreover	175	0.021	696	means	166	0.020
657	outcome	175	0.021	697	point	166	0.020
658	produced	175	0.021	698	promoter	166	0.020
659	receptors	175	0.021	699	sample	166	0.020
660	until	175	0.021	700	susceptibility	166	0.020
661	variants	175	0.021	701	take	166	0.020
662	ensure	174	0.021	702	allow	165	0.020

Rank	HFW	f	%	Rank	HFW	f	%
703	caused	165	0.020	743	choice	157	0.019
704	good	165	0.020	744	diabetes	157	0.019
705	injection	165	0.020	745	infants	157	0.019
706	personnel	165	0.020	746	plasma	157	0.019
707	total	165	0.020	747	resulted	157	0.019
708	monitoring	164	0.020	748	capacity	156	0.019
709	whole	164	0.020	749	consequences	156	0.019
710	clearly	163	0.019	750	evaluation	156	0.019
711	enhanced	163	0.019	751	sexual	156	0.019
712	led	163	0.019	752	suggested	156	0.019
713	performance	163	0.019	753	together	156	0.019
714	proposed	163	0.019	754	widely	156	0.019
715	trial	163	0.019	755	class	155	0.018
716	amount	162	0.019	756	relationship	155	0.018
717	daily	162	0.019	757	status	155	0.018
718	degree	162	0.019	758	variability	155	0.018
719	differentiation	162	0.019	759	challenge	154	0.018
720	identification	162	0.019	760	difference	154	0.018
721	serve	162	0.019	761	discussed	154	0.018
722	treated	162	0.019	762	once	154	0.018
723	cages	161	0.019	763	plan	154	0.018
724	decrease	161	0.019	764	prevent	154	0.018
725	indeed	161	0.019	765	center	153	0.018
726	longer	161	0.019	766	controlled	153	0.018
727	vervet	161	0.019	767	heart	153	0.018
728	bartonella	160	0.019	768	mutations	153	0.018
729	designed	160	0.019	769	properties	153	0.018
730	measured	160	0.019	770	stimulus	153	0.018
731	structure	160	0.019	771	strong	153	0.018
732	efficacy	159	0.019	772	times	153	0.018
733	mean	159	0.019	773	around	152	0.018
734	nature	159	0.019	774	feeding	152	0.018
735	prevalence	159	0.019	775	baboons	150	0.018
736	safety	159	0.019	776	examples	150	0.018
737	sensitivity	159	0.019	777	generation	150	0.018
738	basic	158	0.019	778	agent	149	0.018
739	determined	158	0.019	779	application	149	0.018
740	included	158	0.019	780	previously	149	0.018
741	medicine	158	0.019	781	questions	149	0.018
742	regulatory	158	0.019	782	remain	149	0.018

Rank	HFW	f	%	Rank	HFW	f	%
783	stage	149	0.018	823	terms	144	0.017
784	values	149	0.018	824	auditory	143	0.017
785	year	149	0.018	825	components	143	0.017
786	concern	148	0.018	826	every	143	0.017
787	elevated	148	0.018	827	followed	143	0.017
788	husbandry	148	0.018	828	inflammatory	143	0.017
789	improve	148	0.018	829	pigs	143	0.017
790	surgical	148	0.018	830	successful	143	0.017
791	according	147	0.017	831	article	142	0.017
792	free	147	0.017	832	authors	142	0.017
793	interaction	147	0.017	833	challenges	142	0.017
794	line	147	0.017	834	comparative	142	0.017
795	literature	147	0.017	835	hand	142	0.017
796	revealed	147	0.017	836	pathway	142	0.017
797	room	147	0.017	837	program	142	0.017
798	stages	147	0.017	838	assessed	141	0.017
799	transcription	147	0.017	839	captive	141	0.017
800	active	146	0.017	840	causes	141	0.017
801	closely	146	0.017	841	drosophila	141	0.017
802	community	146	0.017	842	formation	141	0.017
803	improved	146	0.017	843	mechanism	141	0.017
804	involves	146	0.017	844	observations	141	0.017
805	altered	145	0.017	845	reward	141	0.017
806	benefit	145	0.017	846	young	141	0.017
807	guidelines	145	0.017	847	consider	140	0.017
808	pathogenesis	145	0.017	848	detection	140	0.017
809	release	145	0.017	849	duration	140	0.017
810	resistance	145	0.017	850	expected	140	0.017
811	respiratory	145	0.017	851	glucose	140	0.017
812	responsible	145	0.017	852	history	140	0.017
813	selected	145	0.017	853	nervous	140	0.017
814	taken	145	0.017	854	reproductive	140	0.017
815	therapies	145	0.017	855	disorder	139	0.017
816	translational	145	0.017	856	doses	139	0.017
817	working	145	0.017	857	maintenance	139	0.017
818	below	144	0.017	858	repeated	139	0.017
819	colonies	144	0.017	859	syndrome	139	0.017
820	extensive	144	0.017	860	bias	138	0.016
821	opioid	144	0.017	861	concentrations	138	0.016
822	scientists	144	0.017	862	diagnostic	138	0.016

Rank	HFW	f	%	Rank	HFW	f	%
863	global	138	0.016	903	extent	132	0.016
864	intervention	138	0.016	904	functions	132	0.016
865	maintain	138	0.016	905	past	132	0.016
866	nutrition	138	0.016	906	published	132	0.016
867	problems	138	0.016	907	aged	131	0.016
868	progression	138	0.016	908	avian	131	0.016
869	understand	138	0.016	909	degeneration	131	0.016
870	action	137	0.016	910	dependence	131	0.016
871	binge	137	0.016	911	emergency	131	0.016
872	examined	137	0.016	912	equipment	131	0.016
873	frequently	137	0.016	913	experiment	131	0.016
874	practice	137	0.016	914	institutions	131	0.016
875	primarily	137	0.016	915	length	131	0.016
876	question	137	0.016	916	next	131	0.016
877	remains	137	0.016	917	placed	131	0.016
878	genomes	136	0.016	918	targets	131	0.016
879	spatial	136	0.016	919	energy	130	0.015
880	cross	135	0.016	920	learned	130	0.015
881	finally	135	0.016	921	providing	130	0.015
882	mortality	135	0.016	922	surface	130	0.015
883	able	134	0.016	923	tools	130	0.015
884	advantages	134	0.016	924	allele	129	0.015
885	colony	134	0.016	925	alone	129	0.015
886	far	134	0.016	926	availability	129	0.015
887	marmosets	134	0.016	927	average	129	0.015
888	nutrient	134	0.016	928	feed	129	0.015
889	phenotypes	134	0.016	929	involve	129	0.015
890	project	134	0.016	930	obtained	129	0.015
891	sensitive	134	0.016	931	origin	129	0.015
892	spontaneous	134	0.016	932	proliferation	129	0.015
893	university	134	0.016	933	rare	129	0.015
894	cost	133	0.016	934	address	128	0.015
895	evaluate	133	0.016	935	breed	128	0.015
896	location	133	0.016	936	lab	128	0.015
897	needs	133	0.016	937	left	128	0.015
898	smaller	133	0.016	938	making	128	0.015
899	sources	133	0.016	939	validity	128	0.015
900	table	133	0.016	940	condition	127	0.015
901	tolerance	133	0.016	941	eggs	127	0.015
902	exhibited	132	0.016	942	evaluated	127	0.015

Rank	HFW	f	%	Rank	HFW	f	%
943	investigation	127	0.015	972	distinct	123	0.015
944	rearing	127	0.015	973	experiences	123	0.015
945	respectively	127	0.015	974	family	123	0.015
946	risks	127	0.015	975	laboratories	123	0.015
947	versus	127	0.015	976	applied	122	0.015
948	apes	126	0.015	977	called	122	0.015
949	cancers	126	0.015	978	dietary	122	0.015
950	dish	126	0.015	979	figure	122	0.015
951	diversity	126	0.015	980	inbred	122	0.015
952	just	126	0.015	981	leads	122	0.015
953	linked	126	0.015	982	measure	122	0.015
954	separation	126	0.015	983	protection	122	0.015
955	space	126	0.015	984	sufficient	122	0.015
956	candidate	125	0.015	985	absence	121	0.014
957	five	125	0.015	986	alternative	121	0.014
958	programs	125	0.015	987	bird	121	0.014
959	similarities	125	0.015	988	isolated	121	0.014
960	structures	125	0.015	989	perhaps	121	0.014
961	valuable	125	0.015	990	relapse	121	0.014
962	additionally	124	0.015	991	secondary	121	0.014
963	conduct	124	0.015	992	along	120	0.014
964	infant	124	0.015	993	onset	120	0.014
965	longevity	124	0.015	994	vary	120	0.014
966	periods	124	0.015	995	administered	119	0.014
967	veterinarians	124	0.015	996	adverse	119	0.014
968	via	124	0.015	997	chemical	119	0.014
969	combined	123	0.015	998	density	119	0.014
970	defeat	123	0.015	999	finding	119	0.014
971	disaster	123	0.015	1000	focused	119	0.014
TOTAL						623,583	74.168

APPENDIX B
THE SECOND-1,000 HIGH FREQUENCY WORD LIST IN LARA
CORPUS

Rank	HFW	f	%	Rank	HFW	f	%
1	healthy	119	0.014	32	hypothesis	115	0.014
2	hippocampus	119	0.014	33	imprinted	115	0.014
3	older	119	0.014	34	includes	115	0.014
4	permit	119	0.014	35	practices	115	0.014
5	stimulation	119	0.014	36	respond	115	0.014
6	success	119	0.014	37	thought	115	0.014
7	week	119	0.014	38	breeds	114	0.014
8	enhance	118	0.014	39	largely	114	0.014
9	facilitate	118	0.014	40	policy	114	0.014
10	failure	118	0.014	41	skin	114	0.014
11	last	118	0.014	42	variables	114	0.014
12	removed	118	0.014	43	abnormal	113	0.013
13	right	118	0.014	44	allows	113	0.013
14	surgery	118	0.014	45	avoid	113	0.013
15	account	117	0.014	46	depending	113	0.013
16	always	117	0.014	47	differ	113	0.013
17	diverse	117	0.014	48	here	113	0.013
18	possibility	117	0.014	49	incidence	113	0.013
19	problem	117	0.014	50	main	113	0.013
20	ways	117	0.014	51	organisms	113	0.013
21	cortisol	116	0.014	52	permits	113	0.013
22	fetal	116	0.014	53	physiology	113	0.013
23	full	116	0.014	54	reasons	113	0.013
24	lung	116	0.014	55	received	113	0.013
25	mass	116	0.014	56	represent	113	0.013
26	prenatal	116	0.014	57	spinal	113	0.013
27	promote	116	0.014	58	variable	113	0.013
28	stressors	116	0.014	59	advantage	112	0.013
29	targeted	116	0.014	60	chimpanzees	112	0.013
30	comparison	115	0.014	61	continue	112	0.013
31	differential	115	0.014	62	drinking	112	0.013

Rank	HFW	f	%	Rank	HFW	f	%
63	harm	112	0.013	103	own	109	0.013
64	material	112	0.013	104	readily	109	0.013
65	nutritional	112	0.013	105	see	109	0.013
66	rapidly	112	0.013	106	captivity	108	0.013
67	ratio	112	0.013	107	comparable	108	0.013
68	scale	112	0.013	108	enrichment	108	0.013
69	unknown	112	0.013	109	flow	108	0.013
70	deficits	111	0.013	110	interval	108	0.013
71	discovery	111	0.013	111	introduction	108	0.013
72	elements	111	0.013	112	paradigm	108	0.013
73	indicated	111	0.013	113	pathology	108	0.013
74	interestingly	111	0.013	114	preference	108	0.013
75	latter	111	0.013	115	quail	108	0.013
76	leading	111	0.013	116	science	108	0.013
77	majority	111	0.013	117	third	108	0.013
78	statistical	111	0.013	118	transportation	108	0.013
79	susceptible	111	0.013	119	component	107	0.013
80	adults	110	0.013	120	cortex	107	0.013
81	close	110	0.013	121	hair	107	0.013
82	establish	110	0.013	122	home	107	0.013
83	goal	110	0.013	123	interventions	107	0.013
84	increasingly	110	0.013	124	organism	107	0.013
85	living	110	0.013	125	step	107	0.013
86	maintained	110	0.013	126	task	107	0.013
87	minutes	110	0.013	127	assays	106	0.013
88	open	110	0.013	128	consideration	106	0.013
89	physiologic	110	0.013	129	decades	106	0.013
90	plasticity	110	0.013	130	display	106	0.013
91	poor	110	0.013	131	guide	106	0.013
92	serum	110	0.013	132	investigations	106	0.013
93	situation	110	0.013	133	measurements	106	0.013
94	somatic	110	0.013	134	regarding	106	0.013
95	toxicity	110	0.013	135	serotonin	106	0.013
96	alter	109	0.013	136	strategy	106	0.013
97	conditioned	109	0.013	137	already	105	0.012
98	cycle	109	0.013	138	correlated	105	0.012
99	detect	109	0.013	139	dysfunction	105	0.012
100	diagnosis	109	0.013	140	institutional	105	0.012
101	international	109	0.013	141	markers	105	0.012
102	itself	109	0.013	142	nucleus	105	0.012

Rank	HFW	f	%	Rank	HFW	f	%
143	organization	105	0.012	183	upon	101	0.012
144	containing	104	0.012	184	whose	101	0.012
145	me	104	0.012	185	distal	100	0.012
146	signal	104	0.012	186	generated	100	0.012
147	sleep	104	0.012	187	larval	100	0.012
148	substance	104	0.012	188	nerves	100	0.012
149	alleles	103	0.012	189	patient	100	0.012
150	demonstrate	103	0.012	190	previous	100	0.012
151	eating	103	0.012	191	probably	100	0.012
152	event	103	0.012	192	relevance	100	0.012
153	independent	103	0.012	193	resource	100	0.012
154	induction	103	0.012	194	TRUE	100	0.012
155	instead	103	0.012	195	actions	99	0.012
156	issue	103	0.012	196	assessing	99	0.012
157	mothers	103	0.012	197	benefits	99	0.012
158	progress	103	0.012	198	bladder	99	0.012
159	technique	103	0.012	199	chemotherapy	99	0.012
160	thereby	103	0.012	200	commercial	99	0.012
161	volume	103	0.012	201	invertebrate	99	0.012
162	manner	102	0.012	202	maze	99	0.012
163	modifications	102	0.012	203	processing	99	0.012
164	observation	102	0.012	204	emotional	98	0.012
165	occurred	102	0.012	205	he	98	0.012
166	quite	102	0.012	206	inhibition	98	0.012
167	simple	102	0.012	207	involvement	98	0.012
168	standards	102	0.012	208	mrna	98	0.012
169	abnormalities	101	0.012	209	prior	98	0.012
170	delivery	101	0.012	210	selective	98	0.012
171	documented	101	0.012	211	shows	98	0.012
172	environments	101	0.012	212	substantial	98	0.012
173	euthanasia	101	0.012	213	baseline	97	0.012
174	find	101	0.012	214	committee	97	0.012
175	internal	101	0.012	215	content	97	0.012
176	million	101	0.012	216	crabs	97	0.012
177	morphine	101	0.012	217	describe	97	0.012
178	mother	101	0.012	218	detailed	97	0.012
179	opportunity	101	0.012	219	detected	97	0.012
180	severity	101	0.012	220	dopamine	97	0.012
181	studying	101	0.012	221	easily	97	0.012
182	targeting	101	0.012	222	examine	97	0.012

Rank	HFW	f	%	Rank	HFW	f	%
223	forms	97	0.012	263	hormone	94	0.011
224	gain	97	0.012	264	late	94	0.011
225	genotype	97	0.012	265	litter	94	0.011
226	indicating	97	0.012	266	sexually	94	0.011
227	inflammation	97	0.012	267	typical	94	0.011
228	manipulation	97	0.012	268	affects	93	0.011
229	mechanical	97	0.012	269	apparent	93	0.011
230	pathogen	97	0.012	270	building	93	0.011
231	presented	97	0.012	271	chromatin	93	0.011
232	products	97	0.012	272	contain	93	0.011
233	reared	97	0.012	273	countries	93	0.011
234	turn	97	0.012	274	distribution	93	0.011
235	background	96	0.011	275	earlier	93	0.011
236	communication	96	0.011	276	marks	93	0.011
237	evolution	96	0.011	277	offer	93	0.011
238	institution	96	0.011	278	practical	93	0.011
239	invasive	96	0.011	279	separate	93	0.011
240	marine	96	0.011	280	shock	93	0.011
241	phenotypic	96	0.011	281	zoonotic	93	0.011
242	postnatal	96	0.011	282	birth	92	0.011
243	pressure	96	0.011	283	centers	92	0.011
244	tanks	96	0.011	284	elegans	92	0.011
245	adequate	95	0.011	285	fully	92	0.011
246	almost	95	0.011	286	having	92	0.011
247	axis	95	0.011	287	oxygen	92	0.011
248	bacteria	95	0.011	288	parrots	92	0.011
249	considerable	95	0.011	289	ranging	92	0.011
250	discuss	95	0.011	290	stressful	92	0.011
251	done	95	0.011	291	trait	92	0.011
252	fed	95	0.011	292	ultimately	92	0.011
253	makes	95	0.011	293	understood	92	0.011
254	nearly	95	0.011	294	broad	91	0.011
255	opportunities	95	0.011	295	compounds	91	0.011
256	outside	95	0.011	296	concept	91	0.011
257	pool	95	0.011	297	date	91	0.011
258	purposes	95	0.011	298	effort	91	0.011
259	back	94	0.011	299	embryonic	91	0.011
260	carbon	94	0.011	300	federal	91	0.011
261	conditioning	94	0.011	301	sizes	91	0.011
262	emerging	94	0.011	302	spread	91	0.011

Rank	HFW	f	%	Rank	HFW	f	%
303	stereotypes	91	0.011	343	start	88	0.010
304	testosterone	91	0.011	344	vertebrate	88	0.010
305	beyond	90	0.011	345	antibodies	87	0.010
306	considerations	90	0.011	346	comparisons	87	0.010
307	external	90	0.011	347	house	87	0.010
308	handling	90	0.011	348	mentioned	87	0.010
309	implicated	90	0.011	349	parts	87	0.010
310	normally	90	0.011	350	share	87	0.010
311	off	90	0.011	351	six	87	0.010
312	perform	90	0.011	352	smoke	87	0.010
313	recognized	90	0.011	353	spondylosis	87	0.010
314	reducing	90	0.011	354	vaccine	87	0.010
315	reflect	90	0.011	355	allowing	86	0.010
316	removal	90	0.011	356	express	86	0.010
317	seems	90	0.011	357	hepatic	86	0.010
318	entire	89	0.011	358	isolation	86	0.010
319	evolutionary	89	0.011	359	learn	86	0.010
320	exist	89	0.011	360	movement	86	0.010
321	instance	89	0.011	361	polymorphisms	86	0.010
322	marmoset	89	0.011	362	reduces	86	0.010
323	methyly	89	0.011	363	reporting	86	0.010
324	neoplasia	89	0.011	364	reviewed	86	0.010
325	obtain	89	0.011	365	too	86	0.010
326	pedigree	89	0.011	366	affecting	85	0.010
327	prolonged	89	0.011	367	continued	85	0.010
328	restriction	89	0.011	368	endocrine	85	0.010
329	side	89	0.011	369	excellent	85	0.010
330	though	89	0.011	370	flies	85	0.010
331	vertebrates	89	0.011	371	impaired	85	0.010
332	visual	89	0.011	372	investigated	85	0.010
333	collected	88	0.010	373	limitations	85	0.010
334	confirmed	88	0.010	374	near	85	0.010
335	greatly	88	0.010	375	roles	85	0.010
336	horses	88	0.010	376	sensitization	85	0.010
337	insight	88	0.010	377	uptake	85	0.010
338	located	88	0.010	378	zebra	85	0.010
339	minimal	88	0.010	379	acquisition	84	0.010
340	regulate	88	0.010	380	axons	84	0.010
341	regulations	88	0.010	381	chromosome	84	0.010
342	stable	88	0.010	382	determining	84	0.010

Rank	HFW	f	%	Rank	HFW	f	%
383	directed	84	0.010	423	estimated	81	0.010
384	effectively	84	0.010	424	generate	81	0.010
385	evident	84	0.010	425	identical	81	0.010
386	impacts	84	0.010	426	identifying	81	0.010
387	indicates	84	0.010	427	injections	81	0.010
388	mixed	84	0.010	428	interpretation	81	0.010
389	optimal	84	0.010	429	modified	81	0.010
390	reference	84	0.010	430	navigation	81	0.010
391	replication	84	0.010	431	ongoing	81	0.010
392	respect	84	0.010	432	pharmacological	81	0.010
393	section	84	0.010	433	points	81	0.010
394	shared	84	0.010	434	serious	81	0.010
395	transplantation	84	0.010	435	transcriptional	81	0.010
396	uses	84	0.010	436	transcripts	81	0.010
397	comprehensive	83	0.010	437	treat	81	0.010
398	covariates	83	0.010	438	clinically	80	0.010
399	dorsal	83	0.010	439	cue	80	0.010
400	exposures	83	0.010	440	endogenous	80	0.010
401	immediate	83	0.010	441	genomics	80	0.010
402	importantly	83	0.010	442	hippocampal	80	0.010
403	juvenile	83	0.010	443	membrane	80	0.010
404	peptide	83	0.010	444	minimize	80	0.010
405	regeneration	83	0.010	445	modeling	80	0.010
406	replacement	83	0.010	446	monkey	80	0.010
407	situations	83	0.010	447	neurological	80	0.010
408	sometimes	83	0.010	448	outbreak	80	0.010
409	technology	83	0.010	449	oxidative	80	0.010
410	transfer	83	0.010	450	pets	80	0.010
411	unlike	83	0.010	451	resistant	80	0.010
412	women	83	0.010	452	striatum	80	0.010
413	block	82	0.010	453	stroke	80	0.010
414	completely	82	0.010	454	view	80	0.010
415	costs	82	0.010	455	approved	79	0.009
416	nutrients	82	0.010	456	complexity	79	0.009
417	pairs	82	0.010	457	decline	79	0.009
418	reason	82	0.010	458	except	79	0.009
419	systematic	82	0.010	459	fundamental	79	0.009
420	bacterial	81	0.010	460	head	79	0.009
421	concentration	81	0.010	461	immediately	79	0.009
422	create	81	0.010	462	mapping	79	0.009

Rank	HFW	f	%	Rank	HFW	f	%
463	note	79	0.009	503	percentage	77	0.009
464	organs	79	0.009	504	persistent	77	0.009
465	projects	79	0.009	505	raised	77	0.009
466	strongly	79	0.009	506	repetitive	77	0.009
467	successfully	79	0.009	507	tasks	77	0.009
468	tank	79	0.009	508	transport	77	0.009
469	vocal	79	0.009	509	african	76	0.009
470	advances	78	0.009	510	aggressive	76	0.009
471	discussion	78	0.009	511	apoptosis	76	0.009
472	extensively	78	0.009	512	baboon	76	0.009
473	grade	78	0.009	513	core	76	0.009
474	growing	78	0.009	514	eight	76	0.009
475	his	78	0.009	515	extremely	76	0.009
476	ideal	78	0.009	516	haplotypes	76	0.009
477	mamu	78	0.009	517	humane	76	0.009
478	pregnancy	78	0.009	518	leprosy	76	0.009
479	quantitative	78	0.009	519	locomotor	76	0.009
480	quickly	78	0.009	520	mating	76	0.009
481	reviews	78	0.009	521	migration	76	0.009
482	setting	78	0.009	522	pathological	76	0.009
483	achieve	77	0.009	523	prevention	76	0.009
484	achieved	77	0.009	524	sections	76	0.009
485	activated	77	0.009	525	significance	76	0.009
486	adaptive	77	0.009	526	spectrum	76	0.009
487	again	77	0.009	527	tend	76	0.009
488	aquatic	77	0.009	528	agencies	75	0.009
489	arm	77	0.009	529	aversive	75	0.009
490	bingeing	77	0.009	530	cannabis	75	0.009
491	cephalopods	77	0.009	531	conclusions	75	0.009
492	characterization	77	0.009	532	gestation	75	0.009
493	colonization	77	0.009	533	her	75	0.009
494	decreases	77	0.009	534	interesting	75	0.009
495	designs	77	0.009	535	mirnas	75	0.009
496	ecosystem	77	0.009	536	necrosis	75	0.009
497	etiology	77	0.009	537	pups	75	0.009
498	fewer	77	0.009	538	restricted	75	0.009
499	inhibitors	77	0.009	539	signals	75	0.009
500	mild	77	0.009	540	skeletal	75	0.009
501	organ	77	0.009	541	subsequently	75	0.009
502	pair	77	0.009	542	surveillance	75	0.009

Rank	HFW	f	%	Rank	HFW	f	%
543	base	74	0.009	583	improvement	72	0.009
544	cardiovascular	74	0.009	584	influences	72	0.009
545	challenging	74	0.009	585	larvae	72	0.009
546	depends	74	0.009	586	map	72	0.009
547	disruption	74	0.009	587	month	72	0.009
548	extreme	74	0.009	588	noxious	72	0.009
549	fibers	74	0.009	589	paradigms	72	0.009
550	final	74	0.009	590	repair	72	0.009
551	foods	74	0.009	591	seeking	72	0.009
552	generations	74	0.009	592	special	72	0.009
553	glutamate	74	0.009	593	supply	72	0.009
554	law	74	0.009	594	synaptic	72	0.009
555	possibly	74	0.009	595	tool	72	0.009
556	produces	74	0.009	596	air	71	0.008
557	reliable	74	0.009	597	approval	71	0.008
558	robust	74	0.009	598	characteristic	71	0.008
559	showing	74	0.009	599	collection	71	0.008
560	decision	73	0.009	600	epigenetics	71	0.008
561	embryos	73	0.009	601	excessive	71	0.008
562	enough	73	0.009	602	half	71	0.008
563	evaluating	73	0.009	603	hour	71	0.008
564	existing	73	0.009	604	inhibitor	71	0.008
565	folate	73	0.009	605	locus	71	0.008
566	lasting	73	0.009	606	modification	71	0.008
567	limb	73	0.009	607	profiles	71	0.008
568	neurophilia	73	0.009	608	sheep	71	0.008
569	poultry	73	0.009	609	singing	71	0.008
570	smoking	73	0.009	610	taking	71	0.008
571	tail	73	0.009	611	themselves	71	0.008
572	threat	73	0.009	612	traditional	71	0.008
573	autism	72	0.009	613	brief	70	0.008
574	axonal	72	0.009	614	cattle	70	0.008
575	chemicals	72	0.009	615	depressive	70	0.008
576	consume	72	0.009	616	epithelial	70	0.008
577	define	72	0.009	617	escape	70	0.008
578	down	72	0.009	618	heterogeneity	70	0.008
579	efficient	72	0.009	619	hormones	70	0.008
580	enable	72	0.009	620	induces	70	0.008
581	establishment	72	0.009	621	list	70	0.008
582	genotypes	72	0.009	622	maintaining	70	0.008

Rank	HFW	f	%	Rank	HFW	f	%
623	network	70	0.008	663	avoidance	67	0.008
624	predict	70	0.008	664	diameter	67	0.008
625	progressive	70	0.008	665	easy	67	0.008
626	protective	70	0.008	666	follow	67	0.008
627	reaction	70	0.008	667	initially	67	0.008
628	route	70	0.008	668	leprae	67	0.008
629	sound	70	0.008	669	move	67	0.008
630	structural	70	0.008	670	outbreaks	67	0.008
631	tendon	70	0.008	671	represents	67	0.008
632	utility	70	0.008	672	shorter	67	0.008
633	array	69	0.008	673	solution	67	0.008
634	assessments	69	0.008	674	suitable	67	0.008
635	composition	69	0.008	675	technologies	67	0.008
636	conserved	69	0.008	676	theory	67	0.008
637	experienced	69	0.008	677	widespread	67	0.008
638	ganglia	69	0.008	678	adulthood	66	0.008
639	highest	69	0.008	679	ago	66	0.008
640	materials	69	0.008	680	alpacas	66	0.008
641	mutation	69	0.008	681	barrier	66	0.008
642	powerful	69	0.008	682	capable	66	0.008
643	safe	69	0.008	683	consequence	66	0.008
644	search	69	0.008	684	deletion	66	0.008
645	similarity	69	0.008	685	egg	66	0.008
646	simply	69	0.008	686	give	66	0.008
647	trauma	69	0.008	687	original	66	0.008
648	advanced	68	0.008	688	oversight	66	0.008
649	cats	68	0.008	689	palatable	66	0.008
650	domestic	68	0.008	690	partial	66	0.008
651	hypothalamus	68	0.008	691	pig	66	0.008
652	implications	68	0.008	692	sentinel	66	0.008
653	lived	68	0.008	693	society	66	0.008
654	marked	68	0.008	694	storm	66	0.008
655	methylated	68	0.008	695	american	65	0.008
656	proper	68	0.008	696	anatomy	65	0.008
657	recognition	68	0.008	697	antigen	65	0.008
658	regulated	68	0.008	698	consistently	65	0.008
659	stressor	68	0.008	699	cord	65	0.008
660	teaching	68	0.008	700	correlation	65	0.008
661	trained	68	0.008	701	differentiate	65	0.008
662	applications	67	0.008	702	explain	65	0.008

Rank	HFW	f	%	Rank	HFW	f	%
703	feature	65	0.008	743	estimate	63	0.007
704	fixed	65	0.008	744	examination	63	0.007
705	isoflurane	65	0.008	745	feasible	63	0.007
706	locations	65	0.008	746	frequent	63	0.007
707	lost	65	0.008	747	hypomethylation	63	0.007
708	obvious	65	0.008	748	idea	63	0.007
709	plans	65	0.008	749	nociceptive	63	0.007
710	purpose	65	0.008	750	pen	63	0.007
711	receive	65	0.008	751	position	63	0.007
712	relationships	65	0.008	752	presumably	63	0.007
713	reveal	65	0.008	753	principles	63	0.007
714	team	65	0.008	754	protected	63	0.007
715	trade	65	0.008	755	reverse	63	0.007
716	vascular	65	0.008	756	sense	63	0.007
717	why	65	0.008	757	service	63	0.007
718	attempts	64	0.008	758	shift	63	0.007
719	beginning	64	0.008	759	standardized	63	0.007
720	consumed	64	0.008	760	supported	63	0.007
721	direction	64	0.008	761	accurate	62	0.007
722	dominant	64	0.008	762	added	62	0.007
723	feedback	64	0.008	763	ages	62	0.007
724	investigator	64	0.008	764	aggression	62	0.007
725	lateral	64	0.008	765	began	62	0.007
726	magnitude	64	0.008	766	cardiac	62	0.007
727	matrix	64	0.008	767	circumstances	62	0.007
728	networks	64	0.008	768	come	62	0.007
729	paired	64	0.008	769	confounding	62	0.007
730	planning	64	0.008	770	endpoints	62	0.007
731	protect	64	0.008	771	enzymes	62	0.007
732	recommended	64	0.008	772	equine	62	0.007
733	reproduction	64	0.008	773	hypermethylation	62	0.007
734	screening	64	0.008	774	intervals	62	0.007
735	specimens	64	0.008	775	listed	62	0.007
736	allowed	63	0.007	776	muscles	62	0.007
737	calcium	63	0.007	777	olfactory	62	0.007
738	china	63	0.007	778	regard	62	0.007
739	conduction	63	0.007	779	reinforcement	62	0.007
740	created	63	0.007	780	responding	62	0.007
741	crucial	63	0.007	781	stereotyped	62	0.007
742	embryo	63	0.007	782	toxic	62	0.007

Rank	HFW	f	%	Rank	HFW	f	%
783	addictive	61	0.007	823	random	60	0.007
784	ape	61	0.007	824	systemic	60	0.007
785	completed	61	0.007	825	vitamin	60	0.007
786	examining	61	0.007	826	acquired	59	0.007
787	fiber	61	0.007	827	anxious	59	0.007
788	grooming	61	0.007	828	biochemical	59	0.007
789	immunity	61	0.007	829	capture	59	0.007
790	insights	61	0.007	830	carry	59	0.007
791	integration	61	0.007	831	characterize	59	0.007
792	intended	61	0.007	832	conventional	59	0.007
793	maximum	61	0.007	833	determination	59	0.007
794	molecules	61	0.007	834	employed	59	0.007
795	mycobacteria	61	0.007	835	equivalent	59	0.007
796	nociception	61	0.007	836	estradiol	59	0.007
797	occurrence	61	0.007	837	failed	59	0.007
798	presentation	61	0.007	838	farm	59	0.007
799	reinstatement	61	0.007	839	imprinting	59	0.007
800	researcher	61	0.007	840	imprintome	59	0.007
801	supporting	61	0.007	841	lambs	59	0.007
802	unfortunately	61	0.007	842	oral	59	0.007
803	unit	61	0.007	843	overview	59	0.007
804	vectors	61	0.007	844	parallel	59	0.007
805	white	61	0.007	845	pathogenic	59	0.007
806	yield	61	0.007	846	policies	59	0.007
807	zoos	61	0.007	847	resolution	59	0.007
808	aids	60	0.007	848	rooms	59	0.007
809	antagonists	60	0.007	849	variations	59	0.007
810	bred	60	0.007	850	veterinarian	59	0.007
811	chicken	60	0.007	851	biologists	58	0.007
812	circulating	60	0.007	852	contains	58	0.007
813	creating	60	0.007	853	hosts	58	0.007
814	cynomolgus	60	0.007	854	integrity	58	0.007
815	investigate	60	0.007	855	longitudinal	58	0.007
816	kept	60	0.007	856	matter	58	0.007
817	mutant	60	0.007	857	mediate	58	0.007
818	nuclear	60	0.007	858	molluscs	58	0.007
819	plays	60	0.007	859	monitor	58	0.007
820	postweaning	60	0.007	860	moving	58	0.007
821	probe	60	0.007	861	nucleotide	58	0.007
822	proximal	60	0.007	862	post	58	0.007

Rank	HFW	f	%	Rank	HFW	f	%
863	pregnant	58	0.007	903	became	56	0.007
864	principle	58	0.007	904	causal	56	0.007
865	return	58	0.007	905	contribution	56	0.007
866	sampling	58	0.007	906	copy	56	0.007
867	shell	58	0.007	907	decisions	56	0.007
868	summary	58	0.007	908	driven	56	0.007
869	unclear	58	0.007	909	establishing	56	0.007
870	vulnerability	58	0.007	910	evolved	56	0.007
871	addressed	57	0.007	911	framework	56	0.007
872	aspect	57	0.007	912	goals	56	0.007
873	associations	57	0.007	913	hepatitis	56	0.007
874	autoimmune	57	0.007	914	ischemic	56	0.007
875	chickens	57	0.007	915	limit	56	0.007
876	demonstrating	57	0.007	916	media	56	0.007
877	evacuation	57	0.007	917	mental	56	0.007
878	expanded	57	0.007	918	nuclei	56	0.007
879	experimentally	57	0.007	919	options	56	0.007
880	fields	57	0.007	920	personal	56	0.007
881	fly	57	0.007	921	psychiatric	56	0.007
882	homeostasis	57	0.007	922	reach	56	0.007
883	hypothesized	57	0.007	923	relation	56	0.007
884	interact	57	0.007	924	remaining	56	0.007
885	ischemia	57	0.007	925	supplies	56	0.007
886	kidney	57	0.007	926	yielded	56	0.007
887	managers	57	0.007	927	analgesia	55	0.007
888	moderate	57	0.007	928	aquarium	55	0.007
889	necessarily	57	0.007	929	cerebral	55	0.007
890	neuropathic	57	0.007	930	chimpanzee	55	0.007
891	noise	57	0.007	931	circuits	55	0.007
892	polymorphism	57	0.007	932	continuous	55	0.007
893	predictive	57	0.007	933	cortical	55	0.007
894	quarantine	57	0.007	934	dystrophin	55	0.007
895	rabbits	57	0.007	935	education	55	0.007
896	rank	57	0.007	936	emerged	55	0.007
897	remove	57	0.007	937	finches	55	0.007
898	sequenced	57	0.007	938	introduced	55	0.007
899	tobacco	57	0.007	939	mature	55	0.007
900	translation	57	0.007	940	offers	55	0.007
901	assembly	56	0.007	941	pancreatic	55	0.007
902	attributed	56	0.007	942	peer	55	0.007

Rank	HFW	f	%	Rank	HFW	f	%
943	predator	55	0.007	972	rich	54	0.006
944	proportion	55	0.007	973	settings	54	0.006
945	renewal	55	0.007	974	slow	54	0.006
946	routine	55	0.007	975	soon	54	0.006
947	series	55	0.007	976	specialized	54	0.006
948	statistically	55	0.007	977	transplanted	54	0.006
949	temperatures	55	0.007	978	variance	54	0.006
950	actual	54	0.006	979	accompanied	53	0.006
951	alpha	54	0.006	980	adapted	53	0.006
952	appeared	54	0.006	981	antidepressant	53	0.006
953	armadillos	54	0.006	982	balance	53	0.006
954	away	54	0.006	983	basal	53	0.006
955	carefully	54	0.006	984	causing	53	0.006
956	collagen	54	0.006	985	deprivation	53	0.006
957	compulsive	54	0.006	986	enzyme	53	0.006
958	consensus	54	0.006	987	fast	53	0.006
959	constant	54	0.006	988	fecal	53	0.006
960	depend	54	0.006	989	fluid	53	0.006
961	difficulty	54	0.006	990	habitat	53	0.006
962	endangered	54	0.006	991	inherent	53	0.006
963	forward	54	0.006	992	intact	53	0.006
964	gained	54	0.006	993	legal	53	0.006
965	innate	54	0.006	994	lineage	53	0.006
966	language	54	0.006	995	linkage	53	0.006
967	neuron	54	0.006	996	neurogenesis	53	0.006
968	organizations	54	0.006	997	pre	53	0.006
969	proven	54	0.006	998	profile	53	0.006
970	radiation	54	0.006	999	randomized	53	0.006
971	refinement	54	0.006	1000	reactions	53	0.006
				TOTAL 623,583 74.168			

APPENDIX C
THE HEADWORDS OF ACADEMIC WORD LIST IN LARA
CORPUS

Rank	AWL	f	%	Rank	AWL	f	%
1	regardless	53	0.006	32	influenced	51	0.006
2	surrounding	53	0.006	33	lacking	51	0.006
3	adjacent	52	0.006	34	marker	51	0.006
4	considering	52	0.006	35	mechanistic	51	0.006
5	corresponding	52	0.006	36	morbidity	51	0.006
6	diagnosed	52	0.006	37	office	51	0.006
7	dopaminergic	52	0.006	38	seem	51	0.006
8	drive	52	0.006	39	sodium	51	0.006
9	ethics	52	0.006	40	subset	51	0.006
10	explore	52	0.006	41	temporal	51	0.006
11	eye	52	0.006	42	chain	50	0.006
12	face	52	0.006	43	greatest	50	0.006
13	feces	52	0.006	44	input	50	0.006
14	institute	52	0.006	45	likelihood	50	0.006
15	intravenous	52	0.006	46	link	50	0.006
16	livestock	52	0.006	47	mimic	50	0.006
17	mate	52	0.006	48	minor	50	0.006
18	meet	52	0.006	49	painful	50	0.006
19	newly	52	0.006	50	product	50	0.006
20	predicted	52	0.006	51	profound	50	0.006
21	remodeling	52	0.006	52	publication	50	0.006
22	steps	52	0.006	53	reasonable	50	0.006
23	advance	51	0.006	54	rise	50	0.006
24	begin	51	0.006	55	separated	50	0.006
25	brains	51	0.006	56	threshold	50	0.006
26	certainly	51	0.006	57	timing	50	0.006
27	changed	51	0.006	58	underlie	50	0.006
28	classes	51	0.006	59	wall	50	0.006
29	consists	51	0.006	60	amino	49	0.006
30	domesticated	51	0.006	61	aware	49	0.006
31	exercise	51	0.006	62	beneficial	49	0.006

Rank	AWL	f	%	Rank	AWL	f	%
63	careful	49	0.006	103	competition	47	0.006
64	cognition	49	0.006	104	extend	47	0.006
65	combinations	49	0.006	105	held	47	0.006
66	conducting	49	0.006	106	intense	47	0.006
67	contributes	49	0.006	107	intracellular	47	0.006
68	depth	49	0.006	108	microbial	47	0.006
69	dosing	49	0.006	109	never	47	0.006
70	ear	49	0.006	110	ranges	47	0.006
71	eventually	49	0.006	111	reactive	47	0.006
72	genus	49	0.006	112	referred	47	0.006
73	illness	49	0.006	113	subjected	47	0.006
74	improving	49	0.006	114	supports	47	0.006
75	lives	49	0.006	115	technical	47	0.006
76	opposite	49	0.006	116	users	47	0.006
77	parent	49	0.006	117	vector	47	0.006
78	permanent	49	0.006	118	abilities	46	0.005
79	presents	49	0.006	119	conversely	46	0.005
80	reductions	49	0.006	120	detail	46	0.005
81	storage	49	0.006	121	downstream	46	0.005
82	tract	49	0.006	122	earliest	46	0.005
83	analogous	48	0.006	123	estrogen	46	0.005
84	creation	48	0.006	124	fine	46	0.005
85	department	48	0.006	125	go	46	0.005
86	error	48	0.006	126	heterogeneous	46	0.005
87	industry	48	0.006	127	kinase	46	0.005
88	inhibit	48	0.006	128	postoperative	46	0.005
89	lipid	48	0.006	129	psychological	46	0.005
90	maturation	48	0.006	130	rely	46	0.005
91	modern	48	0.006	131	repeat	46	0.005
92	obese	48	0.006	132	restraint	46	0.005
93	paper	48	0.006	133	stability	46	0.005
94	professional	48	0.006	134	vast	46	0.005
95	prominent	48	0.006	135	bodies	45	0.005
96	thermal	48	0.006	136	circuit	45	0.005
97	threats	48	0.006	137	contexts	45	0.005
98	actually	47	0.006	138	controlling	45	0.005
99	agriculture	47	0.006	139	counterparts	45	0.005
100	amounts	47	0.006	140	dark	45	0.005
101	arms	47	0.006	141	database	45	0.005
102	cluster	47	0.006	142	effectiveness	45	0.005

Rank	AWL	f	%	Rank	AWL	f	%
143	get	45	0.005	183	deep	43	0.005
144	ground	45	0.005	184	economic	43	0.005
145	guidance	45	0.005	185	estimates	43	0.005
146	heat	45	0.005	186	mainly	43	0.005
147	hypertension	45	0.005	187	mostly	43	0.005
148	inform	45	0.005	188	neurobiology	43	0.005
149	insects	45	0.005	189	panel	43	0.005
150	peak	45	0.005	190	principal	43	0.005
151	perspective	45	0.005	191	remarkable	43	0.005
152	private	45	0.005	192	spent	43	0.005
153	prove	45	0.005	193	surprising	43	0.005
154	shape	45	0.005	194	vaccines	43	0.005
155	speed	45	0.005	195	vulnerable	43	0.005
156	students	45	0.005	196	accumulation	42	0.005
157	transporter	45	0.005	197	agency	42	0.005
158	accurately	44	0.005	198	analyzed	42	0.005
159	appearance	44	0.005	199	anterior	42	0.005
160	carried	44	0.005	200	believe	42	0.005
161	confirm	44	0.005	201	century	42	0.005
162	consuming	44	0.005	202	degradation	42	0.005
163	correlate	44	0.005	203	distributed	42	0.005
164	dramatic	44	0.005	204	encoding	42	0.005
165	dynamic	44	0.005	205	exists	42	0.005
166	extracellular	44	0.005	206	families	42	0.005
167	hearing	44	0.005	207	fetus	42	0.005
168	medial	44	0.005	208	middle	42	0.005
169	output	44	0.005	209	negatively	42	0.005
170	phases	44	0.005	210	notably	42	0.005
171	positions	44	0.005	211	path	42	0.005
172	profiling	44	0.005	212	released	42	0.005
173	regimen	44	0.005	213	remained	42	0.005
174	scope	44	0.005	214	stimulate	42	0.005
175	strength	44	0.005	215	today	42	0.005
176	synthesis	44	0.005	216	upper	42	0.005
177	units	44	0.005	217	attractive	41	0.005
178	anesthetized	43	0.005	218	becomes	41	0.005
179	answer	43	0.005	219	chamber	41	0.005
180	arrays	43	0.005	220	closed	41	0.005
181	articles	43	0.005	221	decreasing	41	0.005
182	atrophy	43	0.005	222	desired	41	0.005

Rank	AWL	f	%	Rank	AWL	f	%
223	distance	41	0.005	263	investigating	40	0.005
224	distinction	41	0.005	264	items	40	0.005
225	distinguish	41	0.005	265	minute	40	0.005
226	double	41	0.005	266	option	40	0.005
227	engineered	41	0.005	267	routes	40	0.005
228	entirely	41	0.005	268	seconds	40	0.005
229	functioning	41	0.005	269	survey	40	0.005
230	individually	41	0.005	270	absent	39	0.005
231	instances	41	0.005	271	addressing	39	0.005
232	monitored	41	0.005	272	apply	39	0.005
233	partially	41	0.005	273	arise	39	0.005
234	person	41	0.005	274	assigned	39	0.005
235	popular	41	0.005	275	blocks	39	0.005
236	precise	41	0.005	276	brought	39	0.005
237	preferred	41	0.005	277	commercially	39	0.005
238	preliminary	41	0.005	278	compound	39	0.005
239	prepared	41	0.005	279	continues	39	0.005
240	provision	41	0.005	280	country	39	0.005
241	regular	41	0.005	281	critically	39	0.005
242	school	41	0.005	282	dangerous	39	0.005
243	season	41	0.005	283	ease	39	0.005
244	undergo	41	0.005	284	equal	39	0.005
245	version	41	0.005	285	excess	39	0.005
246	accepted	40	0.005	286	familiar	39	0.005
247	aid	40	0.005	287	fold	39	0.005
248	aim	40	0.005	288	force	39	0.005
249	assay	40	0.005	289	gastrointestinal	39	0.005
250	cold	40	0.005	290	hypotheses	39	0.005
251	contained	40	0.005	291	inducing	39	0.005
252	correct	40	0.005	292	nonetheless	39	0.005
253	eliminate	40	0.005	293	performing	39	0.005
254	exception	40	0.005	294	plant	39	0.005
255	expertise	40	0.005	295	rarely	39	0.005
256	fluorescent	40	0.005	296	real	39	0.005
257	fresh	40	0.005	297	red	39	0.005
258	highlight	40	0.005	298	renal	39	0.005
259	illustrate	40	0.005	299	sets	39	0.005
260	image	40	0.005	300	soft	39	0.005
261	implantation	40	0.005	301	spontaneously	39	0.005
262	index	40	0.005	302	transient	39	0.005

Rank	AWL	f	%	Rank	AWL	f	%
303	translate	39	0.005	343	percent	37	0.004
304	valid	39	0.005	344	promise	37	0.004
305	worldwide	39	0.005	345	proximity	37	0.004
306	clean	38	0.005	346	south	37	0.004
307	committees	38	0.005	347	substrate	37	0.004
308	compare	38	0.005	348	therapeutics	37	0.004
309	constraints	38	0.005	349	transition	37	0.004
310	coupled	38	0.005	350	transmitted	37	0.004
311	develops	38	0.005	351	acts	36	0.004
312	epithelium	38	0.005	352	color	36	0.004
313	expressing	38	0.005	353	construct	36	0.004
314	formed	38	0.005	354	damaged	36	0.004
315	impossible	38	0.005	355	delay	36	0.004
316	injected	38	0.005	356	divided	36	0.004
317	manifestations	38	0.005	357	downregulation	36	0.004
318	modulate	38	0.005	358	experimentation	36	0.004
319	moved	38	0.005	359	indirect	36	0.004
320	neutral	38	0.005	360	infect	36	0.004
321	noninvasive	38	0.005	361	injuries	36	0.004
322	organized	38	0.005	362	matched	36	0.004
323	poorly	38	0.005	363	member	36	0.004
324	portion	38	0.005	364	mind	36	0.004
325	regional	38	0.005	365	overlap	36	0.004
326	shed	38	0.005	366	persist	36	0.004
327	virtually	38	0.005	367	plus	36	0.004
328	adapt	37	0.004	368	reached	36	0.004
329	complement	37	0.004	369	select	36	0.004
330	concluded	37	0.004	370	takes	36	0.004
331	criterion	37	0.004	371	varies	36	0.004
332	cultures	37	0.004	372	adequately	35	0.004
333	decade	37	0.004	373	assist	35	0.004
334	delivered	37	0.004	374	attenuated	35	0.004
335	disadvantages	37	0.004	375	briefly	35	0.004
336	elicit	37	0.004	376	complicated	35	0.004
337	essentially	37	0.004	377	compromise	35	0.004
338	explained	37	0.004	378	concurrent	35	0.004
339	heavy	37	0.004	379	enables	35	0.004
340	know	37	0.004	380	enter	35	0.004
341	labeled	37	0.004	381	foundation	35	0.004
342	minimum	37	0.004	382	harmful	35	0.004

Rank	AWL	f	%	Rank	AWL	f	%
383	ideally	35	0.004	423	priority	34	0.004
384	keep	35	0.004	424	put	34	0.004
385	kingdom	35	0.004	425	quantities	34	0.004
386	locomotion	35	0.004	426	screen	34	0.004
387	mode	35	0.004	427	serves	34	0.004
388	observe	35	0.004	428	sessions	34	0.004
389	prevalent	35	0.004	429	slightly	34	0.004
390	prospective	35	0.004	430	sophisticated	34	0.004
391	recombinant	35	0.004	431	throughput	34	0.004
392	responsive	35	0.004	432	trigger	34	0.004
393	roughly	35	0.004	433	unusual	34	0.004
394	schedule	35	0.004	434	viable	34	0.004
395	severely	35	0.004	435	vital	34	0.004
396	solid	35	0.004	436	acids	33	0.004
397	survive	35	0.004	437	adopted	33	0.004
398	upregulated	35	0.004	438	attempt	33	0.004
399	younger	35	0.004	439	confined	33	0.004
400	affinity	34	0.004	440	council	33	0.004
401	assurance	34	0.004	441	green	33	0.004
402	comes	34	0.004	442	homologous	33	0.004
403	chromosomal	34	0.004	443	leptin	33	0.004
404	deal	34	0.004	444	measuring	33	0.004
405	defects	34	0.004	445	medications	33	0.004
406	deleterious	34	0.004	446	neurodegenerative	33	0.004
407	destruction	34	0.004	447	occasionally	33	0.004
408	dry	34	0.004	448	perceived	33	0.004
409	elucidate	34	0.004	449	predominantly	33	0.004
410	exogenous	34	0.004	450	problematic	33	0.004
411	foot	34	0.004	451	pronounced	33	0.004
412	fruit	34	0.004	452	substances	33	0.004
413	heightened	34	0.004	453	termed	33	0.004
414	hold	34	0.004	454	accordingly	32	0.004
415	inherited	34	0.004	455	asked	32	0.004
416	interfere	34	0.004	456	bind	32	0.004
417	lethal	34	0.004	457	channels	32	0.004
418	load	34	0.004	458	cloning	32	0.004
419	mixture	34	0.004	459	collectively	32	0.004
420	molecule	34	0.004	460	concerned	32	0.004
421	morphology	34	0.004	461	cycles	32	0.004
422	occupational	34	0.004	462	deficient	32	0.004

Rank	AWL	f	%	Rank	AWL	f	%
463	domain	32	0.004	503	neuroscience	31	0.004
464	enriched	32	0.004	504	neurotransmitter	31	0.004
465	ensuring	32	0.004	505	object	31	0.004
466	exact	32	0.004	506	outbred	31	0.004
467	hybrid	32	0.004	507	overcome	31	0.004
468	inclusion	32	0.004	508	prevents	31	0.004
469	incomplete	32	0.004	509	recognize	31	0.004
470	layer	32	0.004	510	represented	31	0.004
471	managed	32	0.004	511	requiring	31	0.004
472	plastic	32	0.004	512	sexes	31	0.004
473	property	32	0.004	513	stated	31	0.004
474	record	32	0.004	514	subtle	31	0.004
475	recruitment	32	0.004	515	sufficiently	31	0.004
476	simultaneously	32	0.004	516	synthetic	31	0.004
477	stop	32	0.004	517	tailed	31	0.004
478	striking	32	0.004	518	truly	31	0.004
479	visible	32	0.004	519	uncommon	31	0.004
480	weights	32	0.004	520	words	31	0.004
481	worth	32	0.004	521	false	31	0.004
482	appropriately	31	0.004	522	accomplished	30	0.004
483	calculated	31	0.004	523	accounted	30	0.004
484	challenged	31	0.004	524	anatomic	30	0.004
485	cited	31	0.004	525	anticipated	30	0.004
486	controversial	31	0.004	526	black	30	0.004
487	cumulative	31	0.004	527	burden	30	0.004
488	deaths	31	0.004	528	characterizing	30	0.004
489	defining	31	0.004	529	choose	30	0.004
490	diminished	31	0.004	530	collaborative	30	0.004
491	dominance	31	0.004	531	ecology	30	0.004
492	emphasis	31	0.004	532	element	30	0.004
493	exhibiting	31	0.004	533	excluded	30	0.004
494	experiencing	31	0.004	534	expand	30	0.004
495	faster	31	0.004	535	fit	30	0.004
496	historical	31	0.004	536	fraction	30	0.004
497	infancy	31	0.004	537	generating	30	0.004
498	integral	31	0.004	538	government	30	0.004
499	justified	31	0.004	539	identity	30	0.004
500	knock	31	0.004	540	immunologic	30	0.004
501	magnetic	31	0.004	541	implemented	30	0.004
502	maximize	31	0.004	542	initiated	30	0.004

Rank	AWL	f	%	Rank	AWL	f	%
543	inspection	30	0.004	583	academic	28	0.003
544	legislation	30	0.004	584	alteration	28	0.003
545	nonspecific	30	0.004	585	breeders	28	0.003
546	notion	30	0.004	586	category	28	0.003
547	overt	30	0.004	587	college	28	0.003
548	pathophysiology	30	0.004	588	composed	28	0.003
549	placental	30	0.004	589	conflict	28	0.003
550	pose	30	0.004	590	connections	28	0.003
551	possess	30	0.004	591	coronary	28	0.003
552	promotes	30	0.004	592	demand	28	0.003
553	reflects	30	0.004	593	detectable	28	0.003
554	returned	30	0.004	594	disturbances	28	0.003
555	run	30	0.004	595	document	28	0.003
556	started	30	0.004	596	drop	28	0.003
557	thoroughly	30	0.004	597	focuses	28	0.003
558	transplant	30	0.004	598	foreign	28	0.003
559	artificial	29	0.003	599	frame	28	0.003
560	born	29	0.003	600	fusion	28	0.003
561	companies	29	0.003	601	healing	28	0.003
562	compelling	29	0.003	602	helpful	28	0.003
563	defense	29	0.003	603	incubation	28	0.003
564	die	29	0.003	604	intermediate	28	0.003
565	enhances	29	0.003	605	introduce	28	0.003
566	euthanized	29	0.003	606	lists	28	0.003
567	fall	29	0.003	607	opposed	28	0.003
568	fashion	29	0.003	608	passed	28	0.003
569	gradual	29	0.003	609	pathologies	28	0.003
570	independently	29	0.003	610	phenomenon	28	0.003
571	infrastructure	29	0.003	611	randomly	28	0.003
572	infusion	29	0.003	612	reagents	28	0.003
573	look	29	0.003	613	replicate	28	0.003
574	missing	29	0.003	614	reversed	28	0.003
575	native	29	0.003	615	seek	28	0.003
576	operating	29	0.003	616	skills	28	0.003
577	receives	29	0.003	617	top	28	0.003
578	replace	29	0.003	618	treating	28	0.003
579	sciences	29	0.003	619	vehicle	28	0.003
580	stock	29	0.003	620	vessels	28	0.003
581	synapses	29	0.003	621	absolute	27	0.003
582	transduction	29	0.003	622	abundance	27	0.003

Rank	AWL	f	%	Rank	AWL	f	%
623	alleviate	27	0.003	663	incorporate	26	0.003
624	antigens	27	0.003	664	indicative	26	0.003
625	attachment	27	0.003	665	leave	26	0.003
626	consecutive	27	0.003	666	limits	26	0.003
627	coordinated	27	0.003	667	liquid	26	0.003
628	cover	27	0.003	668	localized	26	0.003
629	cytokine	27	0.003	669	messenger	26	0.003
630	describes	27	0.003	670	metabolites	26	0.003
631	endpoint	27	0.003	671	offered	26	0.003
632	engage	27	0.003	672	rationale	26	0.003
633	facilitates	27	0.003	673	retention	26	0.003
634	follows	27	0.003	674	unnecessary	26	0.003
635	gross	27	0.003	675	voluntary	26	0.003
636	inner	27	0.003	676	accommodate	25	0.003
637	manipulate	27	0.003	677	candidates	25	0.003
638	modeled	27	0.003	678	captured	25	0.003
639	motivated	27	0.003	679	cavity	25	0.003
640	origins	27	0.003	680	chance	25	0.003
641	participants	27	0.003	681	designing	25	0.003
642	polymerase	27	0.003	682	discrete	25	0.003
643	properly	27	0.003	683	displays	25	0.003
644	raise	27	0.003	684	evaluations	25	0.003
645	sustained	27	0.003	685	flexibility	25	0.003
646	traditionally	27	0.003	686	frequencies	25	0.003
647	trends	27	0.003	687	gap	25	0.003
648	wider	27	0.003	688	grow	25	0.003
649	broader	26	0.003	689	hallmark	25	0.003
650	conjunction	26	0.003	690	hormonal	25	0.003
651	curve	26	0.003	691	interpret	25	0.003
652	device	26	0.003	692	manual	25	0.003
653	differing	26	0.003	693	methodology	25	0.003
654	discomfort	26	0.003	694	narrow	25	0.003
655	discrimination	26	0.003	695	obviously	25	0.003
656	drawn	26	0.003	696	pharmaceutical	25	0.003
657	emerge	26	0.003	697	played	25	0.003
658	expect	26	0.003	698	positively	25	0.003
659	fail	26	0.003	699	probability	25	0.003
660	former	26	0.003	700	pure	25	0.003
661	funded	26	0.003	701	residual	25	0.003
662	hard	26	0.003	702	sharing	25	0.003

Rank	AWL	f	%	Rank	AWL	f	%
703	soluble	25	0.003	736	safely	24	0.003
704	specified	25	0.003	737	sick	24	0.003
705	standardization	25	0.003	738	slowly	24	0.003
706	standing	25	0.003	739	suffer	24	0.003
707	suspected	25	0.003	740	suited	24	0.003
708	syndromes	25	0.003	741	touch	24	0.003
709	topic	25	0.003	742	transferred	24	0.003
710	toxins	25	0.003	743	uniform	24	0.003
711	unrelated	25	0.003	744	accessible	23	0.003
712	adversely	24	0.003	745	accreditation	23	0.003
713	agreement	24	0.003	746	adding	23	0.003
714	barriers	24	0.003	747	annual	23	0.003
715	bilateral	24	0.003	748	assumption	23	0.003
716	call	24	0.003	749	avoiding	23	0.003
717	confer	24	0.003	750	blue	23	0.003
718	consortium	24	0.003	751	dedicated	23	0.003
719	difficulties	24	0.003	752	enormous	23	0.003
720	disrupt	24	0.003	753	instrumental	23	0.003
721	expensive	24	0.003	754	places	23	0.003
722	gender	24	0.003	755	pressures	23	0.003
723	geographic	24	0.003	756	ready	23	0.003
724	gland	24	0.003	757	recommend	23	0.003
725	largest	24	0.003	758	rejection	23	0.003
726	modest	24	0.003	759	resonance	23	0.003
727	neuroendocrine	24	0.003	760	societies	23	0.003
728	obtaining	24	0.003	761	stresses	23	0.003
729	outlined	24	0.003	762	strict	23	0.003
730	paramount	24	0.003	763	summarize	23	0.003
731	phylogenetic	24	0.003	764	suppressed	23	0.003
732	progressively	24	0.003	765	tremendous	23	0.003
733	proof	24	0.003	766	unbiased	23	0.003
734	putative	24	0.003	767	western	23	0.003
735	resident	24	0.003	768	zone	23	0.003
TOTAL						27,656	3.289

APPENDIX D

THE PURPOSE FUNCTION IN LARA CORPUS

PATTERN I: Signal words of objectives and to-infinitive

1. The **objective of this review is to describe** the features of the normal neurogenic niche in the adult mammalian subventricular zone (SVZ1) as well as the factors that promote (1) NSC homing and recruitment to a site of disease and (2) the integration of NSCs into functional tissues in different animal models of human neurological disease.
2. Our **goal is to survey** the results of in vivo experimentation in order to overcome the challenges to successful clinical application of SC-based therapies for neurological disorders.
3. The **goal of this article is to present** lessons learned from the devastating effects of two specific natural disasters in Texas: Tropical Storm Allison, which flooded Houston in June 2001, and Hurricane Ike, which caused severe damage in Galveston in September 2008.
4. Our **purpose is to underscore** the importance of up-to-date crisis management planning based on a knowledgeable, realistic assessment of risks.
5. **An objective was to promote** communication among research groups with the hope that lessons learned in one species might help inform and advance understanding in others.
6. **The aim of this article is to provide** scientific information concerning laboratory confinement of Psittaciformes, with the intention of refining arrangements for their acquisition, housing, and maintenance.
7. Therefore, **the aim of this article is to introduce** the reader to this animal's basic biology and life history, present some preliminary information on characteristics associated with aging in marmosets, and **provide** examples of studies of age-related disease processes in this species.
8. **The aim of this article is to give** potential users an idea of the materials, methods, and effort required to maintain each type of organism in a laboratory or classroom setting.
9. **The goals of this research are to gain** a better understanding of the neurobiology of cocaine addiction and, ultimately, **to use** this knowledge for treatment development.
10. **The goals of the workshop were to** (1) **review** the use of neotropical primates in biomedical research, (2) **identify** areas where neotropical primates are required for research, (3) **identify** areas where neotropical primates facilitate research, and (4) **identify** barriers or challenges to the continued use of these species as biomedical models.
11. **The aim of the study was to design** a protocol to train alpacas successfully to

- remain in a specially designed metabolism pen that would be used for future nutritional studies.
12. **The aim of the present study was** therefore **to modify, optimize, and extend** the in vitro system of bovine respiratory mucosal explants in order to study the BoHV-1 dissemination kinetics in respiratory mucosa.
 13. **The purpose of this article is to review** the existing literature on anesthetic protocols adopted in mice for molecular imaging studies and to report our experience.
 14. **The purpose of this article is to review** the existing literature on molecular imaging studies in mice, **to describe** the effects of different anesthetic protocols on their outcome, and **to report** our own experience with such studies.
 15. **The main aim of our study was to investigate** the stress effect in laboratory rats due to transportation.
 16. **our goal is to identify and summarize** those criteria and methods that should be considered in conducting nutrition research with zebrafish.
 17. **The goal of this review is to provide** researchers, laboratory animal science professionals, architects, and others who may be involved in the planning of a new fish facility with a comprehensive overview of design concerns that should be considered when planning to purchase and install zebrafish housing and life-support equipment in biomedical research settings.
 18. ... **the purpose of this review is to provide** one perspective on the present state and near-term prospects for the field.
 19. The broad **goal of this review is to provide** an overview of various procedures available to study anxiety in the nonhuman primate, with a focus on the behavioral aspects of anxiety.
 20. **The aim of this article is to discuss** the use of systematic review (SR) to address this third moral question.

PATTERN II: Signal words of objectives as the subject and phrases explaining the details of objectives

1. The **focus of this review is** the concerted action of these signals as well as the regulatory and/or stabilizing control circuits exhibited by a class of small RNAs, designated microRNAs.
2. **The focus of this review is** the use of rats as an animal model of cognitive change during aging, and specifically lessons learned from aging rats in behavioral studies of cognitive processes mediated by specialized neural circuitry.
3. **The focus of this article is** the use of anesthesia, analgesia, and euthanasia in invertebrates that are kept for research.
4. **In this article, the focus will be briefly** on the human and canine spinal anatomy, after which this review will particularly focus on spondylosis and DISH in both humans and dogs.

PATTERN III: The personal subject

1. **In this review** we **examine** some of the key issues and challenges in the use of animal models to study human stem cell biology—experimental standardization, body size, immunological barriers, cell survival factors, fusion of host and donor cells, and in vivo imaging and tracking.
2. **In this essay** we **discuss** veterinary disaster planning with information to broaden the reader’s view of disaster response.
3. **In this article** we **describe** how clinical and basic research can substantiate the claims of AAT and we identify specific components for successful AAT.
4. **In this article** the authors (one DVM and one MD) **present** examples of species “cross talk” based on their parallel careers in musculoskeletal oncology surgery.
5. **In this article** we **review** recent progress in the field of pathogen surveillance and discovery with a specific focus on applications in the field of laboratory animal research.
6. **In this article**, we briefly **review** research that contributes to understanding how One Health messages could affect support for wildlife health and conservation.
7. We **identify and discuss** several lessons to be learned from pandemic H1N1 2009 from a One Health perspective, as stronger collaboration among human, animal, and environmental health sectors is necessary to more effectively prevent or detect and respond to influenza pandemics and thus improve human, animal, and environmental health and wellbeing.
8. **In this article** I **discuss** the important role of the avian peripheral ear in the study of acoustic trauma.
9. **In this article** we **review** a model system at the forefront of investigation into the neural bases of information processing, plasticity, and learning: the barn owl auditory localization pathway.
10. **In this review** we **evaluate** methods for the production of transgenic birds, focusing on the advantages and limitations of lentiviral-mediated transgenesis.
11. **In this article** I **describe** an example of progress in a nontraditional species, through studies of the role of sleep in birdsong learning.
12. **In this essay** I **highlight** some of the issues related to the care and maintenance of songbirds in the laboratory.
13. **In this review** we **highlight** some salient contributions of the mouse in aging research: lifespan intervention studies in the Interventions Testing Program of the National Institute on Aging; identification of the genetic underpinnings of the effects of calorie restriction on lifespan; the Aging Phenome Project at the Jackson Laboratory, which has submitted multiple large, freely available phenotyping datasets to the Mouse Phenome Database; insights from spontaneous and engineered mouse mutants; and complex traits analyses identifying quantitative trait loci that affect lifespan. in the laboratory.
14. We **discuss** one other New World monkey group, *Cebus* spp., that might also be an effective NHP model of aging as these species are longer-lived for their body size than any primate except humans.
15. **In this article**, I **explore** a new paradigm—the study of pet dogs—in the field of aging research, describe the prospects and challenges of this alternative approach, and **illustrate** ways in which dog studies can directly complement

work in humans and in conventional model systems.

16. Because of the welfare issues associated with such studies, we **identify** potential concerns that require attention to ensure the appropriate use and humane treatment of the animals involved.
17. **In this article we show** that it is possible to create a reproducible ovine model of postinfarct DCM by inducing an anterolateral infarct of approximately 25% of the LV mass based on the visual selection of the arteries that require ligation.
18. **In this article I review** various criteria that might distinguish nociception from pain.
19. **We review** the physiology of nociceptors and behavioral responses to noxious stimulation in several molluscan taxa, and **discuss** the possibility that nociception may result in painlike states in at least some molluscs that possess more complex nervous systems.
20. **In this review, we discuss** data from human surveys, retrospective and clinical studies, and preclinical research characterizing cannabis dependence.
21. **In this review we provide** background information on Narp and closely related proteins, the neuronal pentraxins, and summarize studies of Narp knockout mice demonstrating that this IEG modulates long-term behavioral responses to drugs of abuse.
22. **In this review we focus on** mechanisms of tolerance and withdrawal after repeated or acute exposure of infants to opiates (mostly morphine), building on the work of others who have addressed similar topics (Noda and Nabeshima 2004; Richardson et al. 2006).
23. **In this essay we review** the importance of research, and of animal models in particular, to shed light on drug addictions.
24. **We briefly review** recent animal models used to study drug addiction and the contribution of data generated by these animal models for the clinical treatment of addictive disorders.
25. **In this review we discuss** neurobiological mechanisms that control food intake, particularly the excessive consumption of foods rich in fat or carbohydrates, and how these systems may be coopted to control the consumption of drugs of abuse.
26. **In this article, we explore** the feasibility of imaging ferret thoracic anatomy by performing CT and PET imaging of uninfected animals.
27. **In the present report we assess** the influence of the animal species origin, with a particular focus on primates, of serum on SOS production and/or deactivation.
28. **In this article, we review** some aspects of *D. rerio* breeding in the laboratory environment.
29. **In this review, we summarize and discuss** the first mechanism, a DNA-binding protein-mediated targeting mechanism, with the recent results that have been derived from genome-wide chromatin immunoprecipitation sequencing (ChIP-Seq1) data, **and we also discuss** the in vivo functions of these histone modification marks with the data derived from mouse and human genetic studies.
30. **In this review, we focus on** the current state of epigenetics as it relates specifically to (murine) secondary palate development and to processes that contribute to CP.
31. **In this review, we propose** that sexual selection provides the foundation of such a framework.

32. **In this review, we highlight** research indicating a link between social and stressful experiences occurring over the lifespan and epigenetic variation and the transgenerational implications of these effects.
33. **In this review, we have examined** the current literature and expert opinions on environmental epigenetics, a term that is narrowly defined as how epigenetics explains the variability in the risk and severity of environmental disease.
34. **In this article, I propose** several underlying factors to explain the general inattentiveness to animal welfare in collaborations between wildlife biologists and veterinarians.
35. **we examine** the ethical and policy-level aspects of research and conservation activities that involve captive wildlife in zoos and aquariums, focusing on some of the implications of accelerating biodiversity decline and rapid environmental change.
36. **In this article, we sketch** a set of constraints to fill that gap.
37. **In this article, we discuss** advantages of feasibility and power gained when using extended pedigrees of rhesus macaques for genetic discovery based on next-generation sequence (NGS) data.
38. **In this review, we summarize** characteristics of pedigreed baboons that have made them valuable models for studies on the genetics of common complex diseases in humans, and **we touch upon** currently available genetic and genomic resources that enhance their value.
39. **In this review, we describe** the steps that we and other members of an international collaborative effort have taken to develop this species into a comprehensive NHP model for systems biology.
40. In particular, **we will review** the use of -omics approaches in studies of simian immunodeficiency virus and respiratory virus pathogenesis and vaccine development, emphasizing the acute and innate responses and the relationship of these to the course of disease and to the evolution of adaptive immunity.
41. **Here, we briefly review** the state of knowledge of human cancer genetics to elaborate on the need for different types of mammalian models, highlighting the strengths of the dog.
42. **In this article we explore** our current understanding of canine cancer genetics.
43. **In this review we will explore** the comparative aspects of human and equine tendon pathophysiology in these specific structures and the practical considerations in obtaining clinically relevant research data for both species.
44. **In this review, we will highlight** the comparative oncology studies that have used molecular imaging techniques, demonstrating the value of spontaneous canine cancers as a research tool in drug and imaging agent development.
45. **In this review, we describe** the social defeat model and **summarize** the current state of knowledge regarding its associated behavioral and physiological effects related to depression.
46. **In this review, we will focus on** the behavioral mechanisms involved in relapse of extinguished fear, taking care to translate work in animal models to humans.
47. **In this article, we focus on** the idiopathic development of SIB in rhesus macaques and assess the utility of the rhesus macaque model for the development of NSSI in the general human population.

48. **In this paper, we will discuss** the main principles and practices for metaanalyses of experimental animal studies.
49. **In this review we examine** some rodent models of bingeing and **determine whether** they meet the criteria for addiction as published in the 4th edition and as proposed for the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM1; APA 2000, 2012) and as established for animal models of drug addiction.
50. **In this review, we will describe** epigenetic changes in the context of carcinogenesis and **offer** examples of models of cancer progression and treatment that allow for the elucidation of the role epigenetics plays in cancer progression and treatment

PATTERN IV: The impersonal subject

1. **This review provides** a description of β cell ablation methods and a discussion of various types of studies of regenerative approaches— β cell proliferation, islet cell transplantation, transdifferentiation, and the use of embryonic and induced pluripotent stem cells—to the treatment of diabetes mellitus.
2. **The study of** novel cancer drugs in dogs with naturally occurring tumors **allows** drug assessment in a cancer that shares many fundamental features with the human cancer condition, and thus **provides** an opportunity to answer questions that inform the cancer drug development path in ways not possible in more conventional models.
3. **This review highlights** the nature of EHEC infection in humans, **provides** a One Health perspective on what is known about EHEC in animal and environmental reservoirs, and **proposes** interventions targeted at pathways of transmission to optimize effective prevention and control measures.
4. **This article provides a review** of the major global trends affecting livestock production and trade, related implications for human, animal, and environmental health, and **reasons** why veterinarians should embrace a One Health approach in dealing with disease problems in this era of globalization.
5. **This essay reviews** aspects of “One Health” research that may be relevant for IACUC consideration in the review and approval of protocols.
6. **This review highlights** the usefulness of investigations using Japanese quail (*Coturnix japonica*) to illustrate general principles about the hormonal control of reproductive and related social behaviors
7. **This review presents** new techniques, experiments, and findings in areas where songbirds have made significant contributions toward understanding of some of the most fundamental questions in neuroscience.
8. **This review concentrates on** these latter quantitative requirements.
9. **This review presents** the design, methods, and main findings of these and other important contributing studies, which have generally revealed beneficial effects of CR on physiological function and the retardation of disease consistent with studies in other species.

10. **This article presents** two case studies illustrating specific concerns in invertebrate research protocols and then **provides** relevant information to address practical IACUC matters related to regulatory and ethical issues, sourcing and record keeping, risk management, assessment of pain and nociception in invertebrates, housing and husbandry, invasive procedures, veterinary care, and humane endpoints.
11. **This review presents** an analysis of the literature on behavioral effects of developmental exposure to nicotine, as assessed in rodent models that mimic the consequences for human offspring of maternal cigarette smoking.
12. **This article presents** an overview of established findings regarding the NMDA receptor and opioid dependence in the context of the immediate withdrawal response and learned aversive behaviors.
13. **This review summarizes** the salient features of Meth-induced neurotoxicity with a focus on the dopamine (DA) neuronal system.
14. **This review presents** a summary of recent research on the uses and effects of ketamine at subanesthetic doses.
15. **This review presents** a discussion of sex differences and the role of gonadal hormones as the biological basis for the sexually dimorphic pattern in behavioral responses to cocaine.
16. **This text presents** a linear regression approach for estimating average daily water consumption rates (g/day) over a water consumption time interval for individual Sprague-Dawley rats.
17. **The present review focuses** specifically **on** the transmission and control of microsporidia in zebrafish facilities.
18. **This article discusses** the variety of neoplasia observed in zebrafish submitted to the ZIRC diagnostic service, neoplasia and related lesions in sentinel fish from selected colonies, results of our prospective tumor studies with WT and mutant lines, our studies of neoplasia in retired broodstock from various colonies, and the diversity of neoplasia documented in carcinogen and genetic research using zebrafish.
19. **This article addresses** the importance of identifying and characterizing the naturally occurring viral infections of zebrafish as the scope of zebrafish models expands into new research areas, and **it briefly discusses** zebrafish susceptibility to experimental viral infection and the characteristics of zebrafish models that enable the study of host factors and viral immunity.
20. **This article presents** a formal review of some published literature and an overview of some unpublished protocols and methods of larval husbandry in Europe and the United States to arrive at a more standardized approach.
21. **This article outlines** some of the current, acceptable methods for providing anesthesia and euthanasia **and provides** some examples of how performance-based approaches can be used to advance the relatively limited number of anesthetic and euthanizing techniques available for zebrafish.
22. **This article explores** aspects that members of institutional animal care and use committees, husbandry care staff, physical plant personnel, veterinarians, principal investigators, and research personnel should consider with regard to the appropriate and necessary care and use of this unique fish model in a teaching, laboratory, or research setting.

23. **This review summarizes** the seminal studies on in utero factors affecting the epigenetic regulation and programming for adult diseases such as obesity, hypertension, type 2 diabetes, and cardiovascular disease.
24. **This essay will discuss** some of the ethical issues raised by African great ape research in the hopes of generating a greater dialogue about best practices.
25. **This review highlights** how recent comparative studies have enhanced the knowledge of human genetics and human evolution.
26. **This text will review** the current understanding of MHC genetics in nonhuman primates, with a focus on Mauritian-origin cynomolgus macaques (*Macaca fascicularis*) and Indian-origin rhesus macaques (*Macaca mulatta*).
27. **This article will focus on** the advancement of transgenic NHPs in the past decade, including transgenic technologies and disease modeling.
28. **This article will detail** some of the issues that must be considered as institutional animal care and use committees (IACUCs) review the use of NHPs in research.
29. **This review will discuss** the most widely used mouse models of DPN, including the main phenotypic characteristics of each model, how to choose the appropriate model, and important considerations when working with selected strains.
30. **this review will focus on** rodent models of CIPN, with special attention to outcome measures used to evaluate the effects of chemotherapy on the peripheral nervous system, and will critically review how faithful these models are to the human disease.
31. **This review article summarizes** earlier seminal observations and **highlights** the recent progress in our understanding of immunopathogenesis of autoimmune neuropathies based on data from animal models.
32. **This review will highlight** the major findings in the SIV-infected pigtailed macaque model of HIV-PN, and **will illustrate** the great value of a reliable large animal model to show the pathogenesis of this complex, HIV-induced disorder of the PNS.
33. **A review of the literature shows** that the toxicity and pathogenesis of chemicals adversely affecting the peripheral nervous system have been studied using animal models.
34. **This review will discuss** the similarities between pediatric and canine OSA with regard to histology, biologic behavior, and molecular genetic alterations that indicate canine OSA is a relevant, spontaneous, large animal model of the pediatric disease and outline how the study of naturally occurring OSA in dogs will offer additional insights into the biology and future treatment of this disease in both children and dogs.
35. **This article provides** an overview of canine TCC, a summary of the similarities and differences between canine and human invasive TCC, and examples of the types of valuable translational research that can be done using dogs with naturally occurring TCC.
36. **This paper reviews** the use of animals in developing pharmacologic therapies in a general sense and for DMD in particular.
37. **This review will discuss** many of the current strategies used in both humans and canines in regards to adoptive T cell therapy.
38. **This review will summarize** historical and recent canine epilepsy research in these five areas.

39. **This article reviews** the current understanding of animal models of HEV infection in both natural and experimental infection settings **and identifies** key research needs and limitations.
40. **This article will discuss** the concept of the model and its background, and present a selection of studies employing and examining the model, alongside the underlying translational rationale of each.
41. **This article will provide** an overview of human depression as well as a developmental model (Diathesis- Stress/Two-Hit) that has been used to explain how underlying vulnerabilities interact with stress to increase the susceptibility to depression.
42. **This overview focuses on** such animal models and tests and how they can be used to assess MDD and BD in rodents.
43. **This review discusses** the methodological features and results of various models of stress in nonhuman primates in the context of their potential relevance for human psychopathology and endocrine dysfunction, particularly mood disorders and dysregulation of the hypothalamic-pituitary-adrenocortical (HPA) system.
44. **This report highlights** the strengths and limitations of the early rearing, genetic, and epigenetic methodologies of macaque neuropsychiatric studies **and ends** by describing the findings of the past 10 to 15 years that have advanced our knowledge of the underlying mechanisms of neuropsychiatric disorders in humans.
45. **This paper will review and discuss** possible sources of biases, **highlight** advantages and limitations of strategies proposed to alleviate them, **and provide** a conceptual framework for improving the reproducibility of laboratory animal research.

PATTERN V: The main verb and to-infinitive

1. **In this article** we **aim to address** the above deficiencies by providing a broad overview of the most salient aspects of passerine biology and their implications for the husbandry and welfare of these species in the laboratory.
2. **We aimed to (1) study** the anatomy of the brain vasculature in our gerbil population and identify heritable characteristics of CoWs and **(2) use** the results of this study to selectively breed an ischemia-sensitive gerbil group.
3. **This review aims to show** that noninvasive brain imaging strategies such as small animal positron emission tomography offer significant potential and promise for modeling motivational disorders such as drug addiction and obesity in humans.
4. **This article aims to provide** insights into the usefulness of the mouse model for epigenetic studies involving nutrition as well as the inherent limitations when compared with epigenetic phenomena in humans.
5. **This review acts as** a succinct resource and provides references for investigators supported by the PHS **to understand** the main expectations and requirements when using animals in research.

PATTERN VI: The main verb as the passive verb and to-infinitive

1. **This brief overview of Internet resources is intended to aid** both novice and experienced individuals in the search for such information.
2. **This overview of Internet resources is intended to aid** both beginning and experienced individuals in the search for such information.
3. **this review is specifically focused on** the phenomenon of genomic imprinting and the imprintome.
4. **This review is meant to briefly describe** the utility of specific preparations that can be used in the fly to explore injury-induced degeneration and functional loss of axons, dendrites, and synapses.
5. **This article is intended to provide** only an **overview** of the permits needed to conduct wildlife research.
6. **A variety of** standard histopathological and immunopathological **procedures** including Epidermal Nerve Fiber Density (ENFD) analysis, Schwann Cell Density, and analysis for other conserved cellular markers **can be used** effectively with armadillos **and will be briefly reviewed in this text.**
7. **In this article,** several allocentric assessment **methods** for rodents **are reviewed and compared with** the MWM. MWM advantages (little training required, no food deprivation, ease of testing, rapid and reliable learning, insensitivity to differences in body weight and appetite, absence of nonperformers, control methods for proximal cue learning, and performance effects) and disadvantages (concern about stress, perhaps not as sensitive for working memory) are discussed.

PATTERN VII: It is + adjective + to-infinitive

Because animal models may not be exact in all features of the disorder being studied, **it is important to investigate** the strengths and weaknesses of using a nonhuman primate model for stereotyped behavior in people with psychological disorders.

PATTERN VIII: The main verb as the passive verb and to-infinitive

This review will focus on the mechanisms through which genetic and environmental factors (specifically dietary) can act alone or in combination to alter DNA methylation patterns, and **we will review** the evidence supporting a critical role for these mechanisms in the etiology of ASDs.

BIOGRAPHY

NAME	Mr. Sirawich Tampanich
DATE OF BIRTH	1 December 1988
PLACE OF BIRTH	Trang, Thailand
INSTITUTION ATTENDED	Kasetsart University, 2007-2010: Bachelor of Arts (English) Mahidol University, 2014-2016: Master of Arts (Applied Linguistics)
HOME ADDRESS	67/11 Plernpitak Road, Tubtaing Subdistrict Muang District, Trang province, 92000 Tel. 0814764688, 0910354115 E-mail: tmeen.sirawich@gmail.com
PRESENTATION	7 th National and International Conference on Humanities and Social Science