

**A CORPUS-BASED STUDY ON FREQUENCY, COLLOCATIONS,
AND PURPOSIVE FUNCTION ON LABORATORY ANIMAL
REVIEW ARTICLES USING WORDS FROM THE OUTSIDE
WORD LIST (OWL)**

VIRATA PANJANON

**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, COLLOCATIONS,
AND PURPOSIVE FUNCTION ON LABORATORY ANIMAL
REVIEW ARTICLES USING WORDS FROM THE OUTSIDE
WORD LIST (OWL)**

Virata Panjanon

.....
Ms. Virata Panjanon
Candidate

Songsri S.

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

Natthapong

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

Y.

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

Patcharee Lertrit

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

W. Pun tai

.....
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, COLLOCATIONS,
AND PURPOSE FUNCTION ON LABORATORY ANIMAL
REVIEW ARTICLES USING WORDS FROM THE OUTSIDE
WORD LIST (OWL)**

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

on
May 30, 2016

Virata Panjanon

.....
Ms. Virata Panjanon
Candidate

Piyatida Changpueng

.....
Assoc. Prof. Piyatida Changpueng
Ph.D. (English as an International
Language)
Chair

Songsri S.

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

Y. T.

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

Natthapong

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

Patcharee L.

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

Aphilak K.

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

ACKNOWLEDGEMENTS

This thesis owes its existence to the help, support, guidance, patience, and expertise of the following persons.

First and foremost, I would like to convey my grateful thanks to my thesis advisor Assoc. Prof. Dr. Songsri Soranastaporn for rendering me the invaluable advice, academic guidance, wholehearted support as well as encouragement along the way from the beginning to the final destination in MA program. My study would not be successfully done without my advisor being there for me as always.

I would also like to express my sincere gratitude to my co-advisors Lect. Natthapong Chanyoo and Lect. Yuwadee Tirataradol, for their substantial assistance in terms of insightful suggestions and practical recommendation. My special thanks also go out to Lect. William Martin, the lecturer of the Applied Linguistics Program and Dr. Vinod Dumblekar for providing me with editing assistance and kind suggestions. Also, I would also like to express my appreciation to all of the lecturers of Applied Linguistics Program at the Faculty of Liberal Arts, Mahidol University, for exchanges of knowledge which enrich my academic experiences and skills to become a more efficient researcher.

This work is supported by the 60th Year Supreme Reign of his Majesty King Bhumibol Adulyadej Scholarship, granted by the Faculty of Graduate Studies Academic Year 2014, Mahidol University. Therefore, I would like to thank for all supporting that provided me.

A special acknowledgement also goes to all of my friends in MA program for their generous support and help as well as motivation and inspiration.

Finally, I would also like to offer my profoundest gratitude to my family including parents, grandmother, and sister for their love and support for my whole life.

Virata Panjanon

A CORPUS-BASED STUDY ON FREQUENCY, COLLOCATIONS, AND
PURPOSIVE FUNCTION ON LABORATORY ANIMAL REVIEW ARTICLES
USING WORDS FROM THE OUTSIDE WORD LIST (OWL)

VIRATA PANJANON 5736185 LAAL / M

M.A. (APPLIED LINGUISTICS)

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

ABSTRACT

The purposes of this corpus-based study were: 1) to produce a collection of the content words of General Service List (GSL), Academic Word List (AWL), and Outside Word List (OWL) used in laboratory animal research articles, 2) to identify the high-frequency content words of OWL used in laboratory animal research articles, 3) to analyze the types of grammatical and lexical collocations of the highest-frequency content words in OWL used in laboratory animal research articles, and 4) to investigate the writing patterns used to state the purposes of the study in the laboratory animal research articles.

The source of corpus included 555,526 running words which were collected from 100 review articles of Institute for Laboratory Animal Research (ILAR) between 2010 and 2014. The 100 review articles were selected by stratified random sampling and sample random sampling techniques. Then only content words in the corpus were categorized into three groups: General Service List (GSL), Academic Word List (AWL), and Outside Word List (OWL) based on the new revised version of GSL and AWL (Browne, Culligan & Phillips, 2013). In the present study, two research instruments were employed to collect data. First, the VocabProfile program was used to categorize words into GSL, AWL, and OWL. In addition, AntConc (3.4.4) was used to create a laboratory animal word list. Then the concordance lines of the highest-frequency words were analyzed to find grammatical and lexical collocations and purposive function. The data were analyzed and presented by percentage, frequency, and quartile.

The results of this study are as follows. 1) OWL has the highest coverage (65.87%), followed by GSL (27.4%) and AWL (6.73%). 2) The high-frequency words in OWL ranked from 1st to 3055th (1st quartile) such as *zebrafish*, *methylation*, *epigenetic*, *cocaine*, and *genome*. 3) Lexical collocations were mostly found such as noun + noun (55.81%), noun + verb (17.83%), adjective + noun (14.15%), and verb + noun (7%). 4) The results showed that purposive sentences were written by an employment of the main verb expressing the objective of the study and a to-infinitive to express the objectives of the study.

KEY WORDS: ACADEMIC WORD LIST (AWL) / COLLCOATION / CORPUS-
BASED STUDY / GENERAL SERVICE LIST (GSL) / LABORATORY
ANIMAL / OUTSIDE WORD LIST (OWL) / PURPOSIVE FUNCTION

205 pages

การศึกษาความถี่ การเกิดร่วมจำเพาะ และหน้าที่ในการแสดงวัตถุประสงค์ของคำเนื้อหาใน OUTSIDE WORD LIST (OWL) โดยใช้คลังข้อมูลภาษาจากบทความปริทัศน์ที่เกี่ยวข้องกับสัตว์ทดลอง

A CORPUS-BASED STUDY ON FREQUENCY, COLLOCATIONS, AND PURPOSIVE FUNCTION ON LABORATORY ANIMAL REVIEW ARTICLES USING WORDS FROM THE OUTSIDE WORD LIST (OWL)

วิรดา ปัญจานนท์ 5736185 LAAL / M

ศศ.ม. (ภาษาศาสตร์ประยุกต์)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: ทรงศรี สรณสถาพร, Ph.D., ณัฐพงษ์ จันทร์อยู่, Ph.D., ยุวดี ภิรชราดล, Ph.D.

บทคัดย่อ

งานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อ 1) รวบรวมคลังข้อมูลคำศัพท์ทางด้านสัตว์ทดลอง ซึ่งประกอบด้วยคำเนื้อหาของรายการคำศัพท์ General Service List (GSL) Academic Word List (AWL) และ Outside Word List (OWL) 2) จำแนกคำศัพท์เฉพาะทางด้านสัตว์ทดลอง ในรายการคำศัพท์ OWL ที่มีความถี่สูงในบทความปริทัศน์ 3) วิเคราะห์การปรากฏร่วมจำเพาะของคำที่มีความถี่สูงสุดในรายการคำศัพท์ OWL และ 4) ศึกษารูปแบบการเขียนวัตถุประสงค์ของงานวิจัยสัตว์ทดลอง

แหล่งของคลังข้อมูลภาษาประกอบด้วยคำศัพท์จำนวน 555,526 คำ จากบทความปริทัศน์ใน วารสาร Institute for Laboratory Animal Research (ILAR) จำนวน 100 ฉบับ ที่ตีพิมพ์ตั้งแต่ ค.ศ. 2012-2014 โดยใช้วิธีการสุ่มตัวอย่างแบบแบ่งชั้นและการสุ่มตัวอย่างแบบง่าย จากนั้นจึงรวบรวมคำศัพท์เฉพาะคำเนื้อหาและแบ่งเป็น GSL AWL และ OWL โดยอ้างอิงรายการคำศัพท์ฉบับปรับปรุงใหม่ของ Browne Culligan และ Phillips (2013)

เครื่องมือที่ใช้ในการวิจัยประกอบด้วยโปรแกรมจำแนกคำศัพท์ VocabProfile และโปรแกรมวิเคราะห์ภาษา AntConc (3.4.4) โปรแกรม VocabProfile จะใช้ในการแบ่งคำศัพท์ออกเป็น GSL AWL และ OWL จากนั้นโปรแกรม AntConc (3.4.4) จะใช้ในการรวบรวมรายการคำศัพท์และแสดงครรชนีคำในบริบท เพื่อวิเคราะห์การปรากฏร่วมจำเพาะของคำและรูปแบบการเขียนวัตถุประสงค์งานวิจัย จากนั้นวิเคราะห์ข้อมูลจากคลังภาษาสัตว์ทดลองโดยการหาคำร้อยละ ค่าความถี่ และการแบ่งควอไทล์

ผลการวิจัยพบว่า 1) คลังข้อมูลสัตว์ทดลองประกอบด้วยรายการคำศัพท์ OWL (65.87%) มากที่สุด รองลงมาคือรายการคำศัพท์ GSL (27.4%) และ รายการคำศัพท์ AWL (6.73%) ตามลำดับ 2) คำศัพท์ที่มีความถี่สูงคือรายการคำศัพท์ที่ปรากฏใน OWL จากลำดับที่ 1 ถึง 3,055 (ควอไทล์ที่ 1) ยกตัวอย่างเช่นคำว่า zebrafish, methylation, epigenetic, cocaine และ genome 3) การปรากฏร่วมจำเพาะแบบ lexical collocation พบมากที่สุดในคลังข้อมูลภาษานี้ โดยเป็นการปรากฏร่วมแบบ noun + noun (55.81%), noun + verb (17.83%), adjective + noun (14.15%) และ verb + noun (7%) และ 4) จากการศึกษาพบว่า วัตถุประสงค์งานวิจัยสามารถเขียนได้โดยการใช้กริยาหลักที่แสดงวัตถุประสงค์หรือ to-infinitive

CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
ABSTRACT (ENGLISH)	iv
ABSTRACT (THAI)	v
LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER I INTRODUCTION	1
1.1 Background of the study	1
1.2 Rationale of the study	4
1.3 Importance of the study	6
1.4 The Problems of the study	8
1.5 Statement of problems	12
1.6 Conceptual framework	13
1.7 Purpose of the study	14
1.8 Research questions	14
1.9 Scope and limitation of the study	15
1.10 Definition of terms	16
CHAPTER II LITERATURE REVIEW	19
2.1 English for Specific Purpose (ESP)	21
2.1.1 English for Academic Purposes (EAP)	21
2.1.2 English for Science and Technology (EST)	23
2.1.3 Field of laboratory animal	29
2.2 Corpus Studies	30
2.2.1 Definitions of corpus	31
2.2.2 Development and use of corpus-based study	32
2.2.3 Concordance software program	33
2.3 Word Lists in a Corpus	36

CONTENTS (cont.)

	Page
2.3.1 General Service List (GSL)	37
2.3.2 Academic Word List (AWL)	40
2.3.3 New General Word List (NGSL) and New Academic Word List (NAWL)	44
2.3.4 Outside Word List (OWL)	45
2.3.5 Technical words	45
2.4 Collocations	49
2.4.1 The importance of collocations	49
2.4.2 Perspectives of collocations	51
2.4.3 Grammatical features of collocations	55
2.4.4 Previous study of collocation research	58
CHAPTER III RESEARCH METHODOLOGY	62
3.1 Source of Corpus	64
3.1.1 Description of corpus	64
3.1.2 Source of corpus	64
3.1.3 Size of corpus	67
3.2 Research Instruments	69
3.2.1 AntConc	69
3.2.2 VocabProfile	70
3.3 Steps of the Study	71
3.4 Data Analysis	77
3.4.1 Rhetorical frameworks	77
3.4.2 Statistical devices	79
CHAPTER IV RESULTS	82
4.1 Content words	82

CONTENTS (cont.)

	Page
4.2 High-frequency content words in OWL	93
4.3 Collocations	95
4.3.1 Grammatical collocation	96
4.3.2 Lexical collocation	98
4.4 Purposive functions	108
CHAPTER V DISCUSSION	115
5.1 Content words	116
5.1.1 The content words of GSL, AWL, and OWL used in the laboratory animal review articles	116
5.1.2 The high-frequency content words in OWL used in the laboratory animal review articles	119
5.2 Collocations	122
5.2.1 Grammatical collocations	123
5.2.2 Lexical collocations	124
5.3 Communicative function	125
5.4 Implications of the Study	127
CHAPTER VI CONCLUSION	128
6.1 Conclusions	128
6.2 Recommendations for Further Studies	131
REFERENCES	132
APPENDICES	146
Appendix A A List of GSL 1 in the Laboratory Animal Corpus	147
Appendix B A List of GSL 2 in the Laboratory Animal Corpus	157
Appendix C A List of GSL 3 in the Laboratory Animal Corpus	162
Appendix D A List of AWL in the Laboratory Animal Corpus	165
Appendix E A List of OWL in the Laboratory Animal Corpus	169

CONTENTS (cont.)

	Page
Appendix F The Concordance Lines of Grammatical Collocations (noun + preposition)	175
Appendix G The Concordance Lines of Lexical Collocations (noun + noun)	178
Appendix H The Concordance Lines of Lexical Collocations (noun + verb)	193
Appendix I The Concordance Lines of Lexical Collocations (adjective + noun)	197
Appendix J The Concordance Lines of Lexical Collocations (verb + noun)	201
Appendix K The Concordance Lines of Purposive Function	204
BIOGRAPHY	205

LIST OF TABLES

Table	Page
1.1 The English Average Scores of The Ordinary National Educational Test (O-NET) Between 2010 and 2014	3
2.1 IMRD Sections and Purpose	25
2.2 Components of Review Articles	29
2.3 The Overview Functions and Features of AntConc	34
2.4 The Comparison of Word Lists in the Corpus	37
2.5 Top-Forty Most Frequent Words in GSL (West, 1953)	38
2.6 The Examples of GSL	39
2.7 The Most Frequent Members of the Word Families in Each Sublist	41
2.8 The Previous Researches of Corpus-Based Studies	43
2.9 Examples of Collocations' Patterns	53
2.10 The Semantic Functions of Prepositions 'In', 'With', and 'From'	55
2.11 The Examples of Primary Compound Nouns	56
2.12 The Examples of Secondary Compound Nouns	56
2.13 The Examples of Active Voice	57
2.14 The Examples of Passive Voice	57
2.15 Previous studies of collocation research	59
3.1 Size of Corpus	68
3.2 Examples of Grammatical Collocations	78
3.3 Examples of Lexical Collocations	79
4.1 The Coverage of Word Lists	84
4.2 The 100 Most Frequent Words in GSL 1	86
4.3 The 100 Most Frequent Words in GSL 2	87
4.4 The 100 Most Frequent Words in GSL 3	88
4.5 The 100 Most Frequent Words in AWL	90

LIST OF TABLES (cont.)

Table		Page
4.6	The 100 Most Frequent Words in OWL	92
4.7	The Examples of Lexical Collocations (noun + noun)	100
4.8	The Examples of Lexical Collocations (noun + verb)	103
4.9	The Examples of Lexical Collocations (adjective + noun)	104
4.10	The Examples of Lexical Collocation (verb + noun)	107
5.1	The Coverage of the Top Five High-frequency Words in OWL	120

LIST OF FIGURES

Figure	Page
1.1 Setting keywords and field for searching	10
1.2 The searching results of corpus study in laboratory animal field from ScienceDirect	10
1.3 Setting keywords for searching	11
1.4 The searching results of corpus study in laboratory animal field from Scopus	11
1.5 Conceptual framework	13
2.1 The overview of literature reviews	20
2.2 The overview of ESP	21
2.3 EST Rhetorical Process Chart by Trimble (1985)	26
2.4 The comparison of software programs used to analyze corpora (cited in Anthony, 2013)	35
2.5 Word lists in a corpus	36
2.6 Meanings of the word ‘operate’	48
2.7 Perspectives of collocations	51
2.8 Grammatical and lexical collocations (Benson, 1985)	54
2.9 Types of compound noun	56
3.1 The overview of methodology	63
3.2 The searching result of the ‘Institute for Laboratory Animal Research’ journal from Beall’s list	66
3.3 The ILAR journal ranking	67
3.4 Comparison of software programs used to analyze corpora (cited in Anthony, 2013)	70

LIST OF FIGURES (cont.)

Figure	Page
3.5	73
3.6	75
3.7	76
3.8	77
3.9	78
3.10	80
4.1	82
4.2	83
4.3	84
4.4	93
4.5	94
4.6	95
4.7	96
4.8	99
4.9	109
4.10	113
5.1	115
5.2	116
5.3	117
5.4	119
5.5	121
5.6	123

CHAPTER I

INTRODUCTION

1.1 Background of the Study

English in present day has been influential to the world. The notion of global language (Kirkpatrick, 2007; Sharifian, 2009) which desires the identical language for the world drives the demand of English for communication especially in academic disciplines (Paquot, 2010). Consequently, most countries try to overcome the barrier in communication by promoting English to their citizens. For example, ASEAN countries, including Thailand (Ministry of Education, 2001), increasingly provide English education to the learners. It is essential to realize the importance of English, the roles of English in ASEAN including Thailand, and English problems in Thailand.

English is important because it is a world language used in everyday life and the Internet. English is widely used in daily life because it is considered to be an international or global language (Kirkpatrick, 2007; Sharifian, 2009). For example, English is used for professional and occupational needs (EOP). Most of the English countries are influential to the world economy (Flower & Peacock, 2001). Those countries drive the industrialization and commercialization throughout the world (Broughton, 1999). Therefore, English is mainly used for business and commerce. English also serve for academic purposes (EAP). Most of academic information is presented in English. For examples, scientific journal, research paper, articles, reports and abstracts were published in English (Cosby, 1987; Ulrich, 1996; Wood, 2001).

Moreover, English is mostly used in the Internet. Hasman (2000) stated that more than 90% of electronic data are in English. Internet World Stats (2015) showed that the most language used on websites was English. There is also the analytical report on User Language Preferences Online (2011) which indicated that in almost all countries, English was the most frequently mentioned language for reading or watching content on the Internet, showing that the most widely used language on the Internet is English. All in all, the age of information and the growth of science, technology, business and commerce has driven the demand of English both in daily life and the Internet world (Hasman, 2000; Montgomery & Crystal, 2013).

English plays significant roles as a second or foreign language and official language among *ASEAN countries including Thailand*. Since Thailand is a member of ASEAN Economic Community (AEC), English as a foreign language becomes more influential to Thais. It is not only in Thailand where English has been predominant. Cheng (2012) mentioned that English was used widely as a second and foreign language among Asia. English is also an official and semi-official language among the Asia, and is used in instruction as the requirement course or core subjects in schools. Thus, English has become a common language or lingua franca (Cheng, 2012). This paves the way to a new dominant role for English in Asian countries (Kirkpatrick, 2012). In addition, English is the only official language of many countries in ASEAN (Kirkpatrick, 2007). It works as an intranational and international communications (Honna, 2012; Low & Hashim, 2012). There are different official languages among ASEAN countries. For example, the official language of Thailand is Thai, whereas the official language of Myanmar is Burmese, and the official language of Vietnam is Vietnamese. Thus, one official language is needed among ASEAN countries. According to the ASEAN Charter, Article 34, it mentioned; “The working language of ASEAN shall be English” (Association of Southeast Asian Nations, 2008). Consequently, English becomes an official communicative tool in the ASEAN Community. As English is a working language in ASEAN Community, there is a commitment for all ASEAN members to support and improve the use of English within their countries. Rooth and Saarela (2007) mentioned that English is one of the most important working skills of workers’ human capital. As the demand for English has risen, the ASEAN country members including Thailand should improve people’s

English language skills. Thus, English is a crucial language in terms of second or foreign language and official language for Thai and other ASEAN members.

However, there are *English problems in Thailand* such as insufficient English language education and low English proficiency in Thai students. As the access of English in Thailand is inadequate, English language education becomes a problem. Baker (2012) stated that there is a lack of resources for English language teaching in Thailand. Hence, Thais do not get enough access to English education. Moreover, proficiency in English of many Thai students is quite low. It could be seen that many government English tests score are not successful, becoming a problem in Thailand (Hayes, 2010). The Ordinary National Educational Test (O-NET) between 2010 and 2014 showed that the English of Thai high school students were below the average scores.

Table 1.1 The English Average Scores of The Ordinary National Educational Test (O-NET) Between 2010 and 2014

Year	Number of Examinees	Total scores	\bar{x}	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>Median</i>	<i>Mode</i>
2014	430,877	100	23.44	11.59	99.00	1.00	20.00	18.00
2013	414,688	100	25.35	12.55	98.00	0.00	22.00	19.00
2012	392,468	100	22.13	11.63	98.00	0.00	19.00	17.00
2011	370,561	100	21.80	11.09	98.00	0.00	19.00	17.00
2010	354,531	100	19.22	12.01	100	0.00	16.00	13.00

(National Institute of Educational Testing Service, 2010-2014)

These scores were considered as “very low” (Office for National Education Standards and Quality Assessment, 2010-2014). In addition, English skills of Thai university graduates are regarded as weakness (Marukatat, 2012). Major problem among graduate students are writing (Pawapatcharandom, 2007). When writing academic paper, students face problems such as applying grammar and forming structure in sentences. To solve these critical problems, more and better education in English language should be provided to improve Thais’ use of English.

Thus, English language education and students' proficiency of English are problems in Thailand.

In conclusion, the importance of English, the roles of English in Thailand, and the English problems in Thailand demand attention. English is regarded, worldwide, as the tool for regular communications in many countries including Thai and ASEAN countries. Although English skill is recommended to promote and improve, Thais still have problems with English.

1.2 Rationales of the Study

English has dominated the most part of the world because it is an applicable language worked for communicative purposes such as studying and working (Crystal, 2003). Hence, effective communication needs bigger vocabulary knowledge. In other words, to achieve specific communicative purposes, people seek for vocabulary knowledge for improving their academic knowledge and communication in professional careers. These reasons are the rationales for the requirement for English for Specific Purpose (ESP), the demand of vocabulary knowledge, and the way to corpus study.

The notions of English for Specific purposes (ESP) (Paltridge & Starfield, 2014) and English for Science and Technology (EST) (Trimble, 1985; Montgomery, 2013) have played a vital role in the second language and EFL teaching and learning. As aforementioned, the age of information and the growth of science, technology, business, and commerce have driven the demand for learning English for specific purpose. For professional achievement in careers, adult learners tend to learn English for particular disciplines, such as English for Business Communication that focuses on common speaking practice in business areas (Bhatia & Bremner, 2012). An important branch in English for Specific purposes (ESP) is English for Science and Technology (EST) (Trimble, 1985; Montgomery, 2013). With the development of science and technology, English is accepted internationally as the most effective tool for communication of science and technology in a majority of the research journals (Belcher, Johns & Paltridge, 2011). Scientific journal, research paper, articles, reports

and abstracts have been published in English, aiding the growth of science (Cosby, 1987; Ulrich, 1996; Wood, 2001). Therefore, EST is needed for adult learners who want to read scientific and technical literature, or keep touch with development of science knowledge and professional areas. Thus, there is a greater demand for English for Specific purpose (ESP) and English for Science and Technology (EST). In order to climb up the ladder of English in a specific discipline, technical terms or vocabulary knowledge is an enforced facilitator to comprehend (Bromley, 2007) and produce (Qian, 2005) language.

Limited vocabulary knowledge could affect the difficulty in reading and writing of EFL and ESL learners; therefore, vocabulary knowledge is essential for reading (Qian, 2005) and writing (Martin-Chang & Gould, 2008). Increasing vocabulary knowledge could help reading comprehension, and give teachers new approaches of teaching. It is important to expand vocabulary because:

“Vocabulary and word knowledge can contribute to improved comprehension, and it provides a sound rationale for increased emphasis on vocabulary instruction”

(Jitendra, Edwards, Sacks, & Jacobson, 2004, p. 299)

“Vocabulary is a principle contributor to comprehension, fluency, and achievement”

(Bromley, 2007, p. 528)

Moreover, researchers agree that vocabulary knowledge influences the quality of writing (Qian, 2005; Weigle, 2002). Research results indicated that the correlation of vocabulary knowledge and EFL writing proficiency is higher (.63) than L1's (.47) (Schoonen, Gelderen, Stoel, Hulstijn & Glopper, 2011). Therefore, vocabulary is necessary for EFL and ESL's reading and writing (Alderson, 2005). The demand of vocabulary knowledge for specific purpose leads to the corpus study because the corpus can provide specific words and actual uses in a particular field (Chung & Nation, 2004; Gotti & Giannoni, 2014).

The investigation of *specific words* in a particular discipline could be gathered in a corpus. Words in a corpus are generally classified into three types: the General Service List (GSL) (West, 1953), the Academic Word List (AWL) (Coxhead, 2000), and Outside Word List (OWL). These three lists are beneficial for vocabulary

learning (Nation, 2001). Technical words or the OWL is investigated by linguistic researchers (Chung & Nation, 2004; Coxhead & Hirsh, 2007) because technical words are likely to occur in a particular text. Hewings (2002) pointed out that the improvement of analytical tools namely genre and corpus analysis is beneficial for teaching pedagogy in a specific discipline. Chung and Nation (2004) found that technical words investigated by corpus-based study could provide a consistency rate of 91.9%. Paltridge and Starfield (2011), the editors of *English for Specific Purposes Journal (ESPj)*, pointed out the outstanding research interests in ESP that discourse or text analysis investigated by genre or corpus base studies has been predominant mostly in *ESPj*, but the studies did not make an empirical study of these and other methods that had been used over time. These are the reasons that a corpus-based study is required.

Moreover, a corpus can show the *actual uses* of the words within specific contexts. For example, Walker (2011) employed a corpus study to examine collocations for EAP students. Knowing words of co-occurrence or collocation can make naturalized language become more fluent (Hill, 2000; Nation, 2001; Durrant & Schmitt, 2010). Consequently, the corpus also provides reliable results and practicality. For these reasons, researchers are interested in the corpus-based approach (Römer & Wulff, 2010; Charles, 2012). Overall, the corpus study is required as a tool to gather specific words and usages.

Thus, ESP is essential for work and study in a particular discipline. Especially in the study, learners need vocabulary to develop their listening, speaking, reading, and writing (Basturkmen, 2010; Montgomery & Crystal, 2013). Therefore, linguists have tried to collect texts and vocabulary to compile a linguistic corpus for the purpose of teaching and learning in a specific discipline.

1.3 Importance of the Study

The present study is aimed to collect words used in laboratory animal review articles. Therefore, the corpus-based study was properly employed to collect the data because a corpus is a collection of linguistic data, and it is useful for teachers and students for teaching and learning.

1.3.1 The Benefits of Corpus-based Study for Teachers

The implications of this study are advantageous for teachers in terms of ESP course design and teaching pedagogy. To design the courses, students' needs are essential, and must be evaluated. Then, the course will be formed for a specific group or purpose of learners (Hutchinson & Waters, 1987; Jordan, 2002; Belcher, Johns, & Paltridge, 2011). The outcomes of the corpus study will be expected to help teachers design an ESP course (Flowerdew, 1993) for learners who are in the field of laboratory animal suitably. Moreover, teaching pedagogy will be enhanced for an ESP course. The results from this study will be useful and worthy in facilitating teachers to develop more effective teaching pedagogy. For example, teachers can select vocabulary and exemplify the use of words in the corpus since the corpus study is the collection of actual used words (Flowerdew, 1993; Aston, Bernardini, & Stewart, 2009; Boulton, Thomas & Jolivet, 2012). In addition to communicative function, teachers can use the patterns found in the corpus to teach students in a writing course. For instance, teachers can use the pattern of purposive function to teach students to write the objectives of the study. Thus, the corpus study could help teachers to design ESP course and teaching pedagogy.

1.3.2 The Benefits of Corpus-based Study for Students

The findings of the study also enhance students' English skills in terms of vocabulary and collocations knowledge. Since the vocabulary in the present study consists of GSL, AWL, and OWL in a specific scientific field, these word lists could increase vocabulary knowledge of scientific students. The GSL and AWL would be regarded as high-frequency words, and will be advantageous for reading comprehension (Nation, 2001). The reading and writing skills also facilitate learners to deal with research papers (Wallace & Wray, 2011). On the other hand, the OWL is dominant for specialized words in the particular investigated field. Therefore, the OWL is similar to technical words that should be emphasized for ESP contexts. As the vocabulary in the study corpus is related to students' interests, they can learn vocabulary from those groups of word lists. This increases students' vocabulary knowledge for reading contexts in their particular field (Qian, 2005). In addition, the findings of collocations will assist students' use of English, fluently. With collocation

knowledge, students can apply words more fluently, proficiently, and like a native speaker (Hill, 2000; Nation, 2001; Durrant & Schmitt, 2010). Thus, English learners can gain benefits from the discovery of specific words and collocations.

Therefore, the importance of the study is expected to contribute benefits to teachers and learners. The high-frequency content words can be used for ESP course design. Also, collocations of words can facilitate students' use of English language, naturally. Importantly, teachers can use the pattern of communicative function as a guideline and a model to teach students how to write a paper.

1.4 The Problems of the Study

“While without grammar very little can be conveyed, without vocabulary nothing can be conveyed” (Wilkins, 1972). The emphasis on vocabulary knowledge is a beginning stage for learning a language. Although vocabulary knowledge is emphasized in language learning, the problem that seems to obstruct learners is lexical or vocabulary (Laufer, 2014). The problems are: 1) insufficiency of vocabulary knowledge, 2) inconsistency of vocabulary learned in the interested field, and 3) the inadequate role model of vocabulary slots and usages.

Without sufficient vocabulary, it could affect students' reading and writing. As vocabulary contains the core meaning of the message, readers may not comprehend the message conveyed due to their inadequacy of vocabulary knowledge. As the ESP students may meet semi-technical and academic terms in particular contexts within their own discipline, the difficulty of reading specialized texts is an obstacle for them (Christie, 2002). Some studies (Laufer, 1989; Sutarsyah, Nation, & Kennedy, 1994; Worthington & Nation, 1996; Waring & Nation, 1997) revealed that a certain number of high frequency words are needed for second language learners to understand the written text. It is similarly to writing. A meaningful message cannot be expressed if enough words are not present. In short, reading and writing cannot be completed without adequate vocabulary.

Moreover, the vocabulary taught does not match with students' learning field. The differences of field of study require inputs from different vocabulary; therefore, it leads to an English for Specific Purpose (ESP) course. According to

Hutchinson and Waters (1987), an ESP course is designed so as to respond the students' need in specific groups or purposes. However, organizing an ESP course may not be practical for every field of study. Students tend to learn the general vocabulary in fundamental course. This can cause problems or difficulties for them when they need to apply the vocabulary knowledge for an academic context in their particular fields because different disciplines contain different and specific list of vocabulary. The study of general vocabulary knowledge of pre-medical science at Mahidol University (Prawanpath, 1984) showed that learners could apply general English vocabulary in daily life, but they lacked experience of scientific English vocabulary used in specific field in a university level. Thus, ESP students in a professional career require particular linguistic and lexical needs for their specific uses such as community practice and discourses and types of documentation (Northcott & Brown, 2006). These needs are fulfilled by providing access to lexical ESP training to the workplace (Dovey, 2006). Thus, the general vocabulary could not respond to learners' needs because it is not matched with learners' field of study.

According to the two databases, 'ScienceDirect' and 'Scopus', there are inadequate corpus studies of vocabulary slots and usages in the field of study of laboratory animals. These two databases were accessed through e-database of Mahidol University on 1 June, 2015. The three main keywords of 'corpus', 'laboratory', and 'animal' was searched in advance by restricting the searching area in Arts and Humanities in order to find the related empirical study of corpus in the field of laboratory animals. The results showed that there were corpus studies in different fields such as medical (Grabowski, 2015; Staples, 2015), agriculture (Martinez, Beck, & Panza, 2009; Muñoz, 2015), chemistry (Valipouri, & Nassaji, 2013), environment (Liu & Han, 2015), business (Walker, 2011), and applied linguistic (Wang, 2014). However, there is no corpus study related to the field of laboratory animals (see Figure 1.1, 1.2, 1.3, and 1.4).

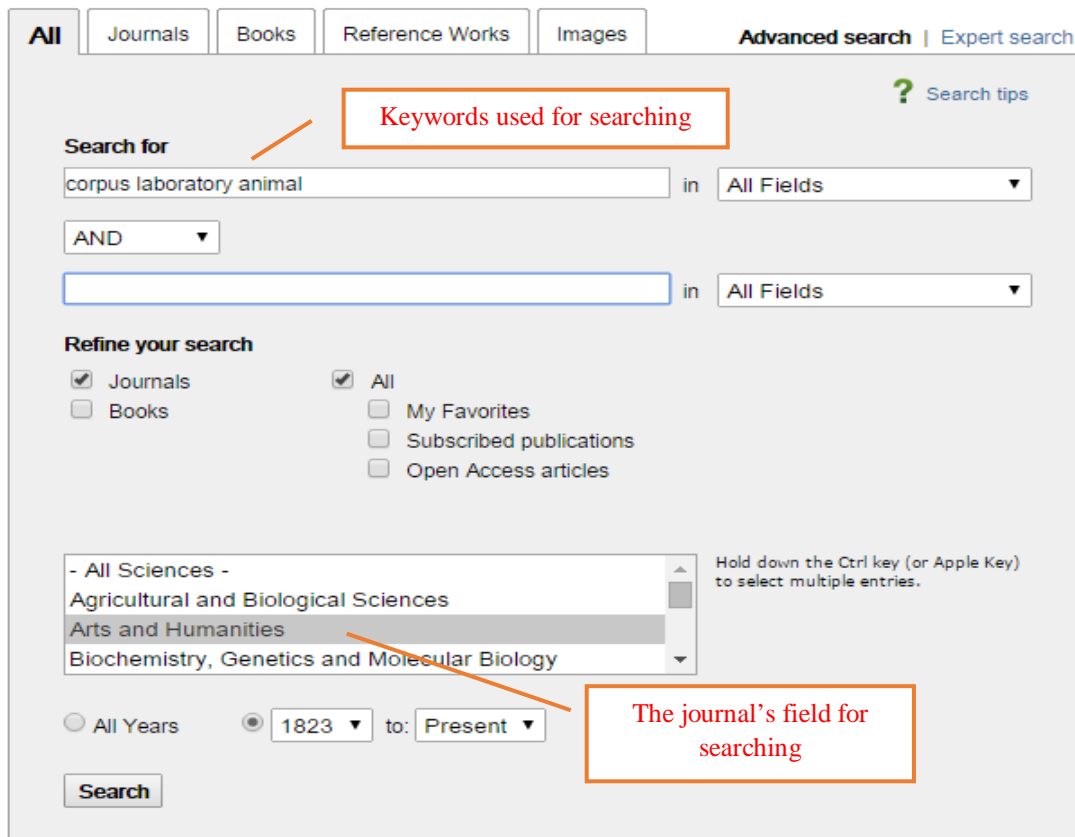


Figure 1.1 Setting keywords and field for searching

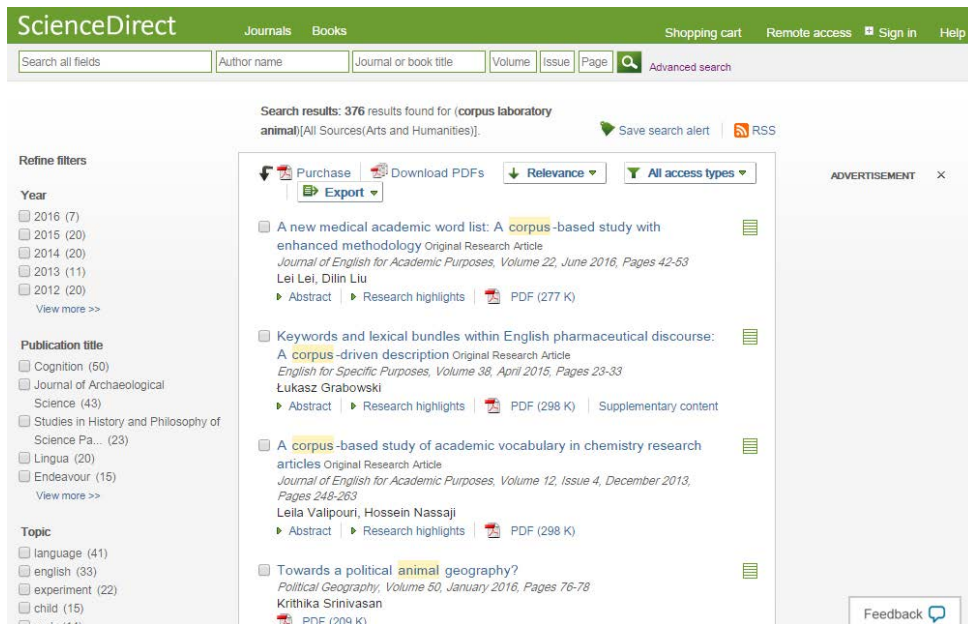


Figure 1.2 The searching results of corpus study in the field of laboratory animals from ScienceDirect

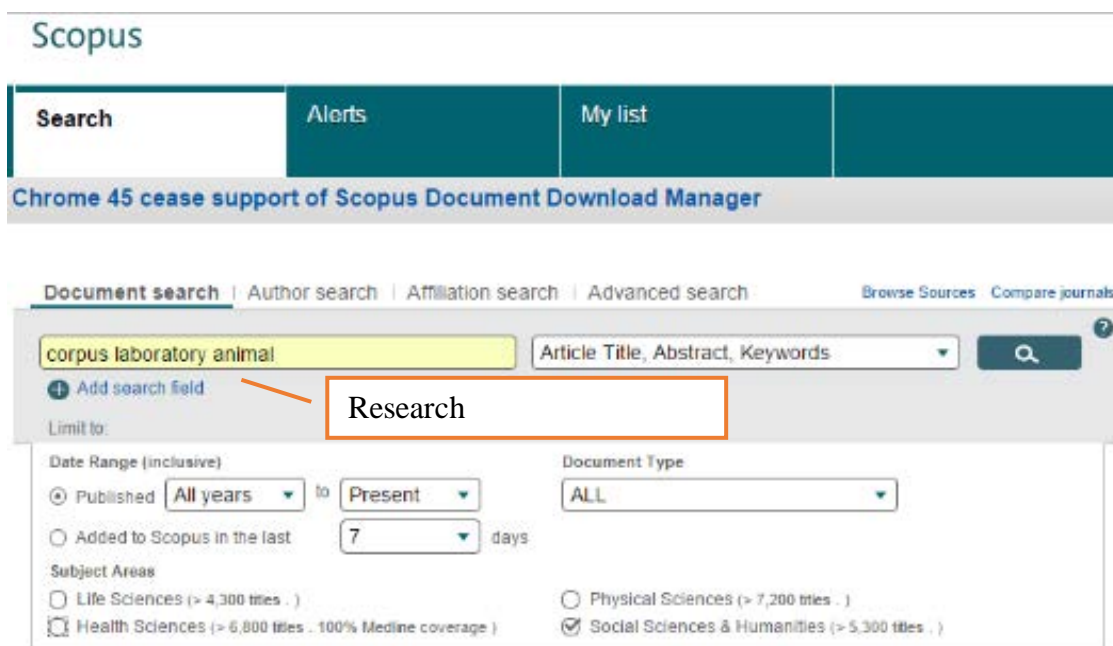


Figure 1.3 Setting keywords for searching

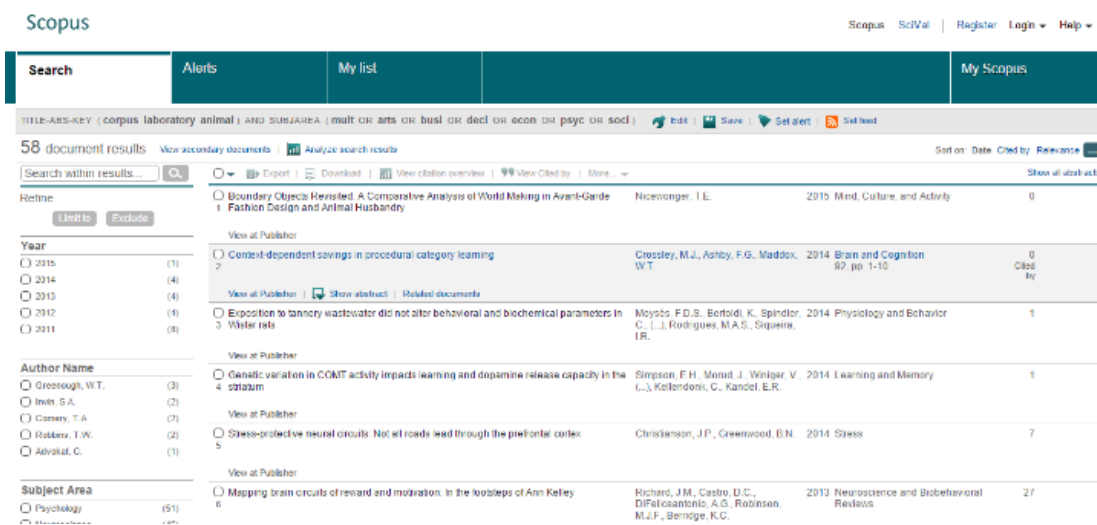


Figure 1.4 The searching results of corpus study in the field of laboratory animals from Scopus

Thus, no corpus study in the field of study of laboratory animals was ever conducted. That is to say, the role model of vocabulary slots and usages in the field of laboratory animals are insufficient. This can be the critical obstacle to design an ESP course for students who study in the field of laboratory animals. All in all, due to the

absence of the role model of vocabulary slots and usages in the field of laboratory animals, the research needs to be conducted.

Thus, students lacked vocabulary knowledge that caused them problems in listening, speaking, reading, and writing. Moreover, students needed to know vocabulary relating to their academic discipline. Due to these the problems, the corpus-based study was conducted and expected to create benefits in teaching and learning.

1.5 Statement of the Problems

Though English for Specific Purpose (ESP), especially English for Science and Technology (EST) is required, and vocabulary knowledge and the use of the words in writing are demanded, no corpus-based study in the field of laboratory animals has been conducted. The director of National Laboratory Animal Center at Mahidol University (1 June 2015) was concerned that the research results in the field of laboratory animals produced by Thai researchers cannot be published worldwide because English language, especially vocabulary knowledge in the laboratory animal science, was their big obstacle. Moreover, there was no ESP course in the field of laboratory animals taught in any university. Above all, the researcher used the keywords of “corpus”, “laboratory animal”, “collocation”, and “purposive function” to search for corpus-based studies in the field of laboratory animals from ScienceDirect and Scopus (1 June 2015). The results showed that there were corpus studies in different fields such as medicine (Staples, 2015), agriculture (Martinez, Beck, & Panza, 2009), chemistry (Valipouri, & Nassaji, 2013), and environmental science (Liu & Han, 2015). However, there were no corpus studies related to the field of laboratory animal research. In addition, the principle word lists of GSL (West, 1953) and AWL (Coxhead, 2000) were used for long time; while, a new version of GSL and AWL was recently revised by Browne, Phillip, and Colligan (2013). Therefore, it was appropriate to employ the new versions of GSL and AWL to the study. To fill the gaps, this present study aimed to conduct a corpus-based study in the field of laboratory animals.

1.6 Conceptual Framework

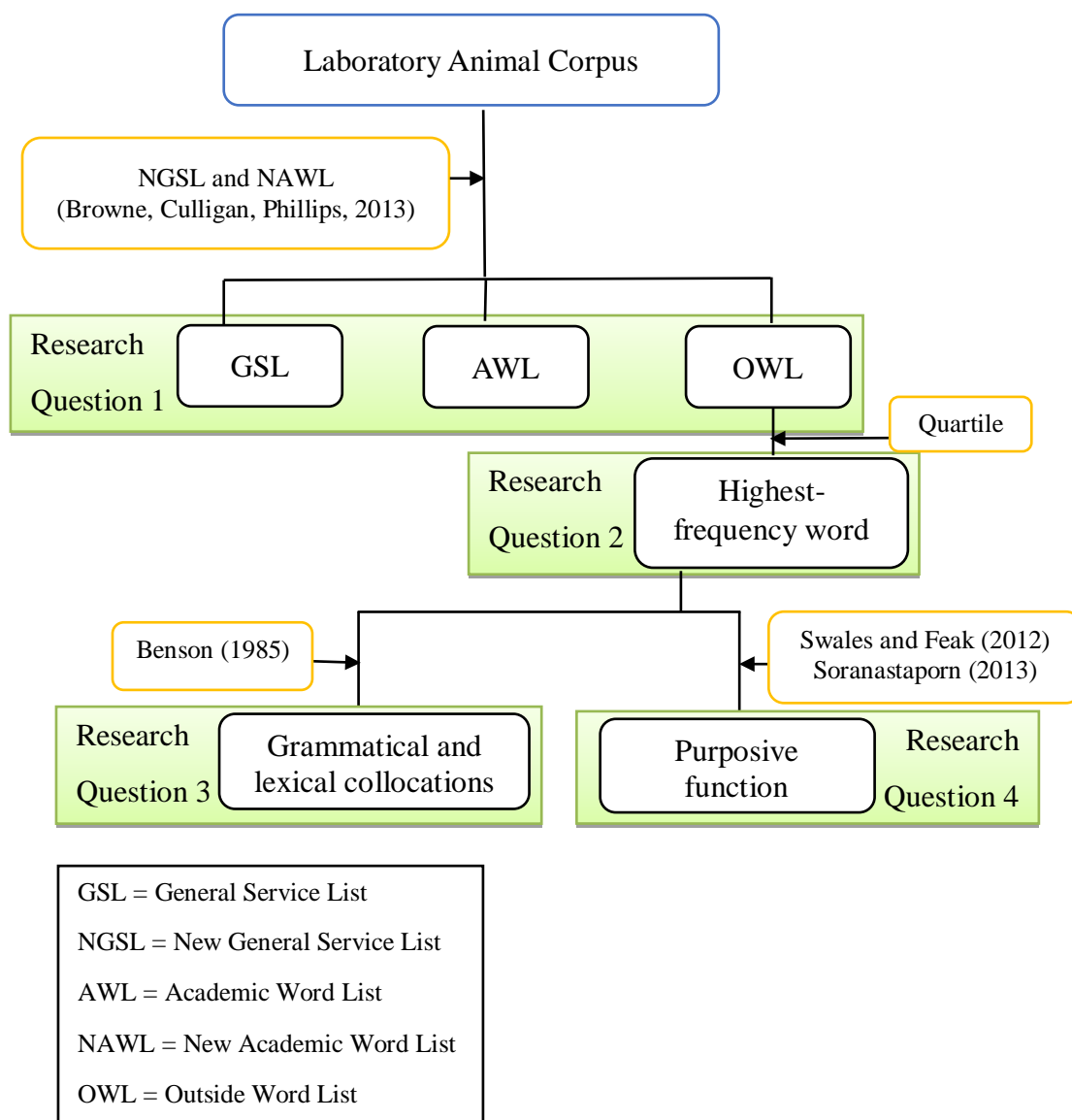


Figure 1.5 Conceptual framework

The researcher adopted the frameworks of word lists ((Browne, Culligan, Phillips, 2013), collocations (Benson, 1985), and purposive function (Swales & Feak, 2012; Soranastaporn, 2013) in order to compile a laboratory animal corpus. The new versions of GSL and AWL were used as reference word lists to find GSL, AWL, and OWL occurring in laboratory animal review articles. The researcher applied the framework of GSL and AWL (Browne, Culligan, Phillips, 2013) because it was

revised and provided an additional 5% of text coverage. Then the words in OWL were identified for high frequency words by using quartile. The emphasis of the present study was based on OWL because it had the specialized or technical words that particularly occurred in the specific discipline. Moreover, the highest-frequency word in OWL was examined for collocation and purposive function. Collocations in the present study were categorized into grammatical and lexical collocations (Benson, 1985). For the purposive function, the researcher applied the framework of writing purposive statement from Swales and Feak (2012) and Soranastaporn (2013).

1.7 Purposes of the Study

1. To produce a collection of the content words of the General Service List (GSL), the Academic Word List (AWL), and the Outside Word List (OWL) used in laboratory animal review articles
2. To identify the high-frequency content words of OWL used in laboratory animal review articles.
3. To analyze the types of grammatical and lexical collocations of the highest-frequency content words in OWL used in laboratory animal review articles.
4. To investigate the writing patterns used to state the purposes in the laboratory animal review articles.

1.8 Research Questions

1. What were the content words of GSL, AWL, and OWL used in the laboratory animal review articles?
2. What were the high-frequency content words in OWL used in the laboratory animal review articles?
3. What collocations of the highest-frequency content word in OWL were used in laboratory animal review articles?
4. What were the patterns used to state the purposes in the laboratory animal review articles?

1.9 Scope and Limitation of the Study

1. The journal investigated in the study

The journal of laboratory animal used in the present study was the Institute for Laboratory Animal Research (ILAR). The selected articles were published between 2010 and 2014. The total numbers of selected articles were 100.

2. The data input for the corpus

The selected articles were investigated for the words in text body including abstract, introduction, method, result, and discussion sections. The labels, titles, pictures, graphs, tables, formulae, and bibliography were excluded.

3. The field of the study

The investigation was limited to the field of laboratory animals which a branch of science. Therefore, the findings of the study may not be generalized for other fields.

4. The classification of word lists

Only content words in GSL, AWL, and OWL were investigated. Words that can be both function and content words were removed from the corpus. That is to say, 'verb to be', 'verb to do', and 'verb to have' were excluded because they can perform as a main verb and auxiliary verb. OWL had words outside of GSL and AWL which were identified for high-frequency words, and the highest-frequency word in OWL was analyzed for collocations and communicative function.

5. The classification of collocations

The collocations in the study were classified into two types, grammatical collocations and lexical collocations. Other collocations were excluded such as discourse marker, multi-word preposition phrase, binominal, trinominal, fixed phrase, fixed expression, semi-fixed expression, part of a proverb, and parts of a quotation.

6. Communicative functions

The investigated function of the present study was purposive function. In other words, the present study investigated the patterns of the statements used to state the purpose of the study.

1.10 Definitions of Terms

Academic Word List refers to the content words found in the Academic Word List (AWL) created by Browne, Culligan, and Phillips (2013).

Cluster is a group of words which were put together in a text (Stubbs, 2001).

Collocation refers to the sequence of words or terms that occurs with two words (Benson, 1985).

Concordance is a list of a given word which often shows in a phrase in a corpus (McEnery & Wilson, 2010).

Content words refer to nouns, verbs, adjectives, and some adverbs (Nation, 2001).

Corpus refers to the collection of linguistic data in form of spoken records and written texts (Sinclair, 1991).

Function words refer to adverbial particles, auxiliary verbs, prepositions, conjunctions, determiners (Nation, 2001).

General Word List refers to the content words found in the General Word List (GSL) created by Browne, Culligan, and Phillips (2013).

Grammatical collocations are “the type in which a dominant word fit together with a grammatical word, typically a noun, verb, or adjective followed by a preposition.” (Benson, 1985)

High-frequency words are the words occurring more than most others in the present corpus.

Lemma refers to the headword.

Lexical collocations “consist of combinations of two basically equal words such as noun + verb, verb + noun, and adjective + noun, in which both words contribute to the meaning.” (Benson, 1985)

Node word is the word used to find the collocation. In other words, node word is the central word used to collocate for their left and right words.

Outside Word List refers to the content words which do not belong to GSL or AWL created by Browne, Culligan, and Phillips (2013).

Token or running words are every form of the words in a word list.

Chapter Summary

The importance of English nowadays has been influential in most parts of the world because English was considered as a global language that has been used worldwide in communication and especially in academic discipline. As the international language in science is English, well-known publishers require to publish papers in English. This becomes a requirement for scientific researchers to read and write English paper. Importantly, they are obliged to establish a clear purposive sentence of the study since the purposive sentence represents the central ideas of the article, and is stated in articles of all standard journals. Therefore, to overcome the barriers in communication, they need to know specialized vocabulary or technical terms and its functions or usage in their particular disciplines. To rectify the problem,

this present study aimed to compile a corpus-based study because corpus can represent words and usages in the specific discipline.

CHAPTER II

LITERATURE REVIEW

This section is provided in order to give an overview of the study. As the purposes of this corpus-based study are to examine the high-frequency words including General Service List (GSL), Academic Word List (AWL), Outside Word List (OWL), and collocations in laboratory animal research articles, the relevant theoretical frameworks and related previous studies have been collected here for review. This literature review is comprised of four main sections: 1) English for Specific Purposes (ESP), 2) corpus studies, 3) word lists in a corpus, and 4) collocations.

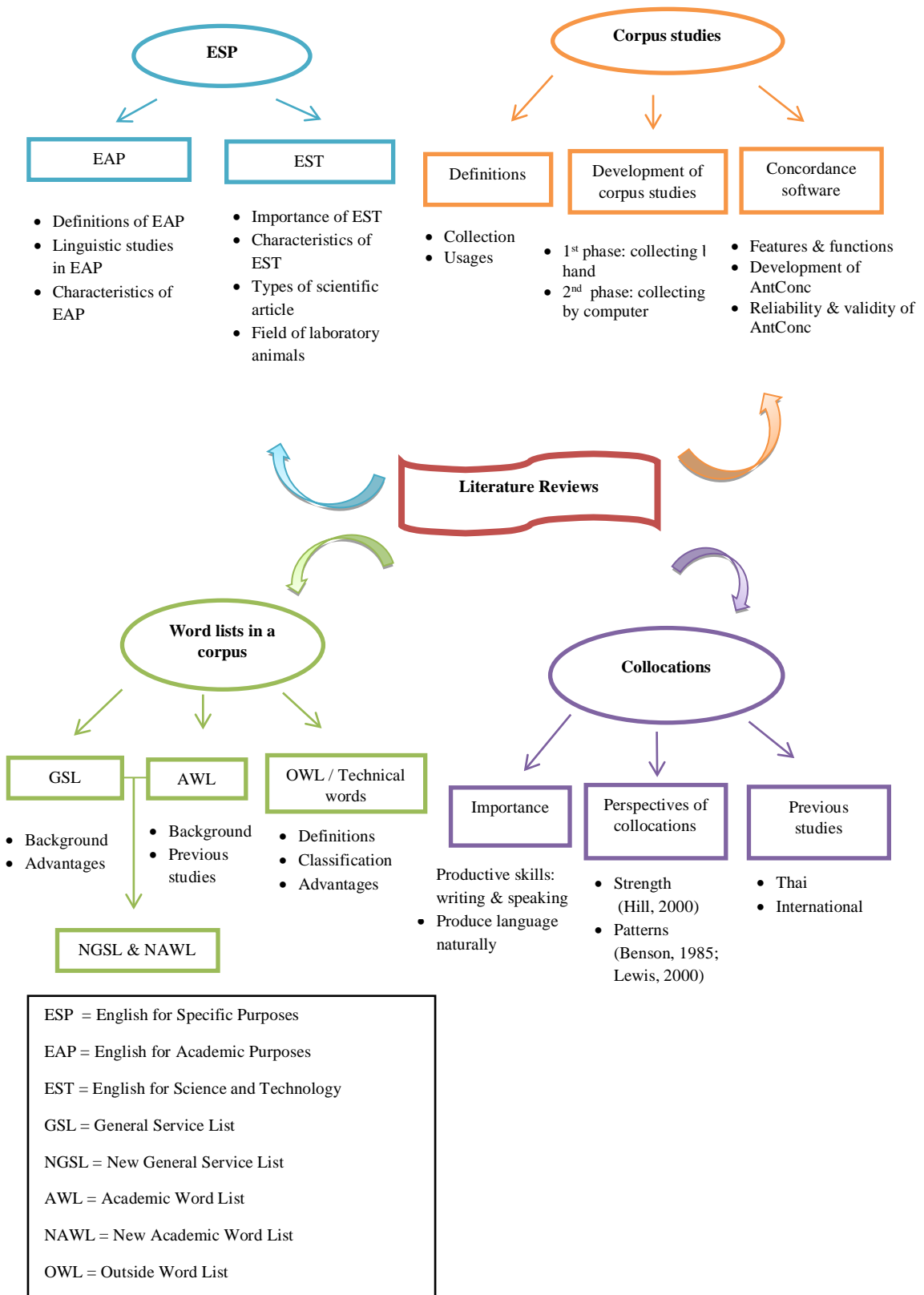


Figure 2.1 The overview of literature reviews

2.1 English for Specific Purposes (ESP)

English for Specific Purposes (ESP) has been utilized for long times. ESP is comprehensible for English as a second language learning (Basturkmen, 2010). Hutchinson and Waters (1987) initially suggested that students' need was so important that ESP course need to be designed for a specific group of students. This notion has been interested and employed by researchers such as Flowerdew and Peacock (2001), Jordan (2002), Gotti and Giannoni, (2014) and Belcher, Johns, and Paltridge (2011). Research has played a significant role in ESP; therefore, the growth of the quality and quantity research in the field has increased, continuously (Belcher, 2006; Hewings & Hewings, 2002). Generally, ESP is separated into various branches according to particular fields and specific purposes. One of the prominent branches of ESP is responsible for English for Academic Purposes (EAP) especially in English for Science and Technology (EST).

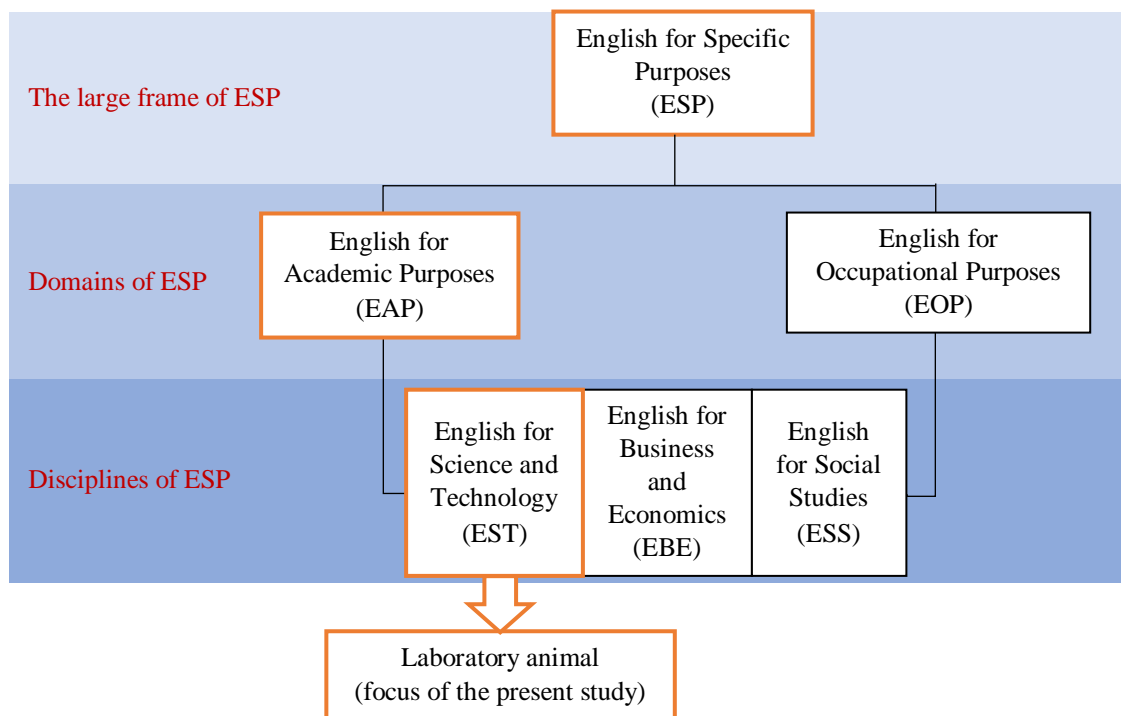


Figure 2.2 The overview of ESP

2.1.1 English for Academic Purposes (EAP)

English for Academic Purposes (EAP) includes the wide range of academic disciplines such as medical, science, business, and arts. EAP is essential to

students because the wide uses of English includes academic disciplines such as research articles (Hasman, 2000). In term of academic teaching and learning, EAP involves the study on specialized words, patterns, and organization of its discipline.

2.1.1.1 Definitions of EAP

Flowerdew and Peacock (2001) defined EAP as the teaching of English with the specific purpose to facilitate learners in academic disciplines. EAP focuses on the academic domain so as to facilitate learners in their studies and to prepare them for future professional occupations. For example, EAP emphasizes on study, conducting research, and teaching language relating to students' study field (Stevens, 1988). Dudley-Evans and John (1998) mentioned that "For most of its history, ESP has been dominated by English for Academic Purposes (EAP)...[and it] continues to dominate internationally".

2.1.1.2 The linguistic studies in EAP

The linguistic studies in EAP were explained through applied linguistics perspectives in terms of *register*, *contextual*, and *discourse analysis*. *Register analysis* was used to explain the description of language in a particular context. For example, Halliday, McIntosh and Stevens (1965) wrote the book named *The Linguistic Sciences and Language Teaching*. They employed the concept of register analysis to analyze the frequency of linguistic forms by using lexis and syntax to explain the language in a particular discipline. The concept of register analysis presented the frequency of linguistic forms, but did not explain how those forms were used with the register or why those forms appeared with high or less frequency. Hence, *Contextual Analysis* (Celce-Murcia, 1980; Tarone, 1981; Swales, 1990) were proposed to explain the use of language in a particular context. For example, Tarone (1981) analyzed the use of passive voice in astrophysics articles. The results of this study explain generalization and communicative purpose of the use of passive voice in the selected genre. Although register and contextual analysis could explain grammar use in sentences, discourse analysis can make a better understanding of how sentences were combined to produce meaning (Trimble, 1985). *Discourse analysis* was the study of content that was beyond the sentence boundary in order to find the purpose and natural occurring of language use. For instance, Hu and Li (2015) investigated the use

of discourse connectives in the argumentative essays written by L1 and L2 university students. The study revealed that students frequently used resultive and summative discourse connectives, and too much use of discourse connectives was likely to make the essays redundant. Therefore, the writers need to be mindful of the pragmatic functions of discourse connectives because they could affect the interpretation of the message (Hu & Li, 2015). To conclude, it is necessary to analyze lexis, syntax and discourse in a particular context of a specific discipline in order to know the use of language, its meanings, and communicative purposes.

2.1.1.3 The characteristics of EAP

According to Strevens (1988) and Basturkmen (2010), there are four characteristics of EAP. First is *design*. The teaching course needs to design to meet specific needs of the learners for examples, English for Medical Studies, English for Economics, and English for Psychology. Second is *content*. The contents such as themes and topics need to relate to particular disciplines, occupations, and activities. For instance, nursing students need to practice conversation like greeting patients or giving medicine instruction. Third is *language*. The language should be appropriate to the activities and learners. Course designers need to consider language teaching such as lexis, syntax, and discourse. Fourth is *uniqueness*. EAP is English language teaching that is designed to teach in a particular discipline relating to students' learning area. In short, EAP characteristics comprise of design, contents, language, and uniqueness to meet the specific objectives of teaching and learning.

Overall, English for Academic Purposes (EAP) is an approach for teaching and learning for specific purposes in particular areas (Hyland, 2006). In addition, English is mainly used in academic areas because most academic information is presented in English such as research articles in scientific field and electronic data on the Internet (Hasman, 2000). Thus, linguistic studies in EAP have explored the lexis, syntax, and discourse in a particular discipline in order to collect data for teaching and learning.

2.1.2 English for Science and Technology (EST)

English for Science and Technology (EST) is a branch of ESP which focuses on academic purpose (Hutchinson & Waters, 1987; Jordan, 2002; Basturkmen, 2010). EST emphasizes on teaching English language skills that students need for

scientific communication. According to Trimble (1985) and Samraj (2005), EST is only one of the ESP and EAP contexts, it is important to scientific teaching and learning because EST has its own linguistic characteristics depending research types.

2.1.2.1 The importance of EST

EST is important to scientific publication and teaching. EST is essential in scientific publication because most well-known scientific publications are required to write in English (Hasman, 2000; Wood, 2001; Swales, 2002). As English is used increasingly in science, English is called the language of international science. In order to achieve scientific publications, EST is focused on teaching English relating to scientific and technological disciplines such as physics, chemistry, biology, and mathematics. Scientific texts, such as research articles, were used to teach as authentic materials, and were analyzed for the linguistic discourse and communicative functions. The results of linguistic findings such as organization in scientific articles can be used to form the syllabus of the EST and a model for teaching writing.

2.1.2.2 The linguistic characteristics of EST in scientific research

Linguistic researchers have worked with a variety of different research and found that genre demonstrates how a particular text distinguishes from others and causes differences in terms of organizations, rhetorical structures, and word choices (Dudley-Evans & John, 1998; Orr, 1999; Mavor & Trayner, 2001).

Scientific texts have a definite *organization* that scientists use to form their pieces of writing to convey a message in content. For example, Swales (2005) demonstrated that the component of experimental research in writing generally combined with four sections: introduction (I), methodology (M), results (R), and discussion (D). He further explained the purposes of each section as shown in Table 1.

Table 2.1 IMRD Sections and Purposes

Sections	Purposes
Introduction:	to present the problems, significance, and purposes of the study.
Methodology:	to explain instrument and procedures used to conduct the study.
Results:	to present what was discovered.
Discussion:	to analyze the importance of results and implications.

Moreover, Trimble (1985) suggested the patterns of EST discourse which were used to organize the scientific research articles as follows:

The Rhetorical Process Chart of EST: A Discourse Approach by Louis Trimble (1985)

- A. The objectives of the total discourse:
 1. Detailing an experiment
 2. Making a recommendation
 3. Presenting new hypotheses or theory
 4. Presenting other types of EST information
- B. The general rhetorical functions that develop the objectives of level (A):
 1. Stating purpose
 2. Reporting past research
 3. Stating the problem
 4. Presenting information on apparatus used in an experiment
 - a) Description
 - b) Operation
 5. Presenting information on experimental procedures
- C. The specific rhetorical functions that develop the general rhetorical functions of level B:
 1. Description: physical, function, and process
 2. Definition
 3. Classification
 4. Instruction
 6. Visual-verbal relation
- D. The rhetorical techniques that provide relationships within and between the rhetorical units of level C:
 - I. Orders
 1. Time orders
 2. Space Orders
 3. Causality

II. Patterns

1. Causality and result
2. Order of importance
3. Comparison and contrast
4. Analogy
5. Exemplification
6. Illustration

Figure 2.3 EST Rhetorical Process Chart by Trimble (1985, p.11)

From Figure 2.3, Trimble (1985) explained that information in Level A presents the objectives of the total discourse, usually found in the introductory section of the discourse. Level B is major pieces of text which will be combined to make up the complete discourse and usually appears in section headings or sub-headings. Level 3 shows obvious specific rhetorical functions that are commonly found in written EST discourse: description, definition, classification, instructions, and visual-verbal relationships between a visual aid and its accompanying text. Finally, Level D shows the relationship of each section such as cause and effect or compare and contrast.

In a scientific writing genre, unique linguistic features such as *rhetorical structures* and *word choices* are used to make a form for communication. Regarding to the writing standard, particularly, the purpose sentence is stated in articles of all standard journals (Swales & Feak, 2012) which people in a specific discipline accept (Paltridge & Starfield, 2011). For example, a purposive function in a scientific research articles is generally stated at the end of introductory section (Pyrzszak & Bruce, 2000; Bailey, 2011; McMillan, 2012; Swales & Feak, 2012). According to Swales and Feak (2012), purposive statements are obligatory to all research articles. It has two variants: Purposive (P) and Descriptive (D).

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, p. 355)

In addition, the patterns of purposive statements are written with words carrying purposive meaning as following (Soranastaporn, 2013).

Pattern I: Presenting the purpose by using signal words of purpose and to-infinitive

Subject (signal words of purposes)			Verb to be	To-infinitive (to + V.inf)	Noun phrase
The purpose (s)	of	paper	is (are) was (were)	to	investigate explore discover find determine develop
The objective (s)		book			
The goal (s)		research			
The aim (s)		study			
The reason (s)					
Noun details of objective(s).					

(Soranastaporn, 2013, p. 122)

Pattern II: Presenting the purposes with main verb in an active form

Subject		Main verb (V2)	Phrase / Clause
This	study	investigated explored discovered analyzed developed	Noun phrase / Noun clause

(Soranastaporn, 2013, p. 125)

Pattern III: Presenting the purposes with main verb in passive form and to-infinitive

Subject		Main Verb (Passive voice)		To-infinitive (to + V.inf)		Phrase / Clause
		be	Past participle (V3)			
This	paper	is was	designed concluded	to in order to	study	Noun phrase Noun clause
	book				compare	
	research				analyze	
	study				ascertain examine evaluate	

(Soranastaporn, 2013, p. 125)

Overall, the purpose of the research can be written by using signal words of purposes and to-infinitive. The examples of signal words that tell the purpose are *purpose*, *objective*, *goal*, and *aim*. These words can be used as a subject of the purposive sentence, and followed by a main verb and to-infinitive (Swales & Feak, 2012; Soranastaporn, 2013).

2.1.2.3 Types of scientific research

Academic literature, especially research papers, is crucial to scientific disciplines. Research articles record and represent the methods and solution of the experiment that generate the knowledge for scientific readers. Review articles are one of the article types published in the scientific disciplines (Noguchi, 2006). This following section discusses detailed descriptions and organization of review articles which relate to the target article of the current study.

Review articles are an analysis of the literature in a particular discipline by summarizing, classifying, analyzing, and comparing (Noguchi, 2006). A scientific review articles rely on previously researched studies or data. The aims of review articles are: to organize a collection of similar research, to assess data, to recognize patterns and movements in the field, to analyze literature, to find research gaps, and to introduce new research areas. Review articles are usually written by

senior scholars at the invitation of journal editors (Swales & Feak, 2012). According to Ridley (2008), review articles consist of title, informative abstract, introduction, body, conclusion, and references (see Table 2.2).

Table 2.2 Components of Review Articles

Sections	Purposes
Title	to inform the title of a review and help readers to decide whether the paper meet the reader' s objective or not.
Abstract	to inform about the main purposes and findings of the review article (informative abstract) or describe the text structure (descriptive abstract).
Introduction	to provide details about the text, identify the requirement for the review, defines the main point, the research question.
Body	to include material, methods, results, and discussion.
Conclusion	to summarize a listing of findings analyzing from previous research.
References	to acknowledges the work mentioned in the text.

In summary, English for Science and Technology (EST) is a major subdivision of English for Academic Purposes (EAP) and English for Specific Purposes (ESP). EST courses are different from others since they pay attention to scientific English and to the selection of the appropriate rhetorical discourse and communicative functions that are generally used in science and technology (Trimble, 1985; Mavor & Trayner, 2001; Swales, 2002). Review articles are one of article types that present an overview of a particular field of professional research. So as to achieve knowledge and communication requirements in scientific disciplines, students need to improve their communicative skills such as writing and reading to increase their EST competence.

2.1.3 Field of Laboratory Animals

Laboratory animal are the focus of the present corpus-based study. It was a branch of English for Science and Technology (EST). Journal Citation and Report (JCR) categorized laboratory animal in the subfields of animal science and zoology, biochemistry, and genetics and molecular biology. Animal laboratory is also known

as animal testing, animal experimentation, animal research, and in vivo testing. Nowadays, the use of laboratory animals is increasingly due to higher demand in the use of animals as a subject or scientific experiments on medical purposes (The Royal Society, 2004).

The use of animals plays significant roles in experimental research. According to The Royal Society (2004), most medical achievement in the 20th century relied on the use of animals. The focus of laboratory animals is based on the research that tries to find a cure for a disease. Examples of laboratory animal research include testing, breeding, toxicology, and cosmetics testing. Laboratory animals used in the experiments are: mice, rats, fish, amphibians, and reptiles. Before an experiment, these laboratory animals are euthanized. In educational field, animal testing can be a brunch of biology courses. National Academy of Sciences (2004) stated that research in the field of animal are not able to be replaced by even intelligent computer models because they cannot cope with the complex interactions between molecules, cells, tissues, organs, organisms, and the environment. Thus, rising numbers of laboratory animal papers published during these years are witnessed (Cohn, 2010).

Overall, laboratory animal research relates to the subfield of English for Science and Technology (EST). The researchers use animals to experiment in order to test medicine and find the solution for treatment. Mice, rats, fish, amphibians, and reptiles are the laboratory animals that are commonly used in the experimental research.

2.2 Corpus Studies

Corpus studies have been investigated widely (Fukushima et al., 2012; Liu, 2012). Large corpora contained a variety of words subdivided by disciplines; for example, British National Corpus (BNC) and Corpus of Contemporary American English (COCA). Specific corpora are smaller because they aim to investigate only selected disciplines; for example, a corpus of chemistry words (Valipouri & Nassaji, 2013) and a corpus of medical words (Wang, Liang & Ge, 2008). This kind of corpora contains specific words used in particular disciplines. To give more details of corpus studies, definitions of corpus, development, and use of corpus-based study,

concordance software program, and the implications of corpus study were explained below.

2.2.1 Definitions of Corpus

The terms 'corpus' refers to a gathering source of data and an area of linguistic study. Therefore, corpus is a tool or method that linguistic researchers use to analyze language. According to the definitions of corpus, the corpus can be described in two aspects: collections and usages.

First, corpus was defined by its role which is the *collection* of written and spoken texts. Sinclair (1991) defined a corpus as a collection of linguistic data in the form of recorded written texts. However, Boswood (1997) added that a corpus is a collection of data which can be consisted of written texts and a transcription of recorded speech. Therefore, a corpus also investigated spoken language. Moreover, corpus is a collection of linguistic data that can be collected in both paper-based and computer-based forms. McEnergy and Wilson (2010) explained that "the term corpus is now almost synonymous with the term machine-readable corpus" Thus, a corpus is normally in computer-based form, nowadays. To sum up, the corpus refers to the collection of written and spoken texts which can be collected in paper base and computer base forms.

In addition, the term of 'corpus' was focused on its *usage* as a tool for linguistic study. Biber, Conrad, and Reppen (2000) indicated that corpus is a study of a large collection of natural texts. It is a source of an authentic language that presents actual (Huston, 2002) and natural uses (Richards & Schmidt, 2010). Thus, corpus is a tool for linguistic study such as specialized words occurred in a real specific area. The term 'corpus' also used in literature area. Mahlberg (2007) referred that "corpus is the practice of linguistic analysis of literary texts, making use of a collection of electronic texts, sampled to be maximally representative of a writer's works or a particular literary genre". Briefly, the definitions of corpus emphasized on its use for linguistic analysis.

In short, 'corpus' is defined by its role which is a collection of spoken or written texts and its usage for linguistic analysis in a particular discipline. The data in

the corpus come from the authentic materials which are likely collected and presented in electronic form, nowadays.

2.2.2 Development and Use of Corpus-Based Study

Corpus studies have been investigated for decades (Hunston, 2011; Richards & Schmidt, 2010). There were two critical phases of corpus developments, i.e, initially, corpus was gathered by hands, and later, it was gathered by computer.

2.2.2.1 The first phase of corpus development

The early corpus was created manually around 19th century. The purposes of corpus were to facilitate English teaching and learning by providing actual use of English samples (Biber, Conrad, & Reppen, 2000). At that time, corpus was done manually because the computer program for concordancing was not available. Therefore, it consumed lots of time. The corpora of frequency word lists such as Thorndike and Lorge's List (1944) and West's General Service List (1953) were collected manually. The General Service List (GSL) (West, 1953) contained a list of 2,000 words that represent the most frequent English words. The list was collected manually from a corpus of written English. The frequencies of various senses of the words in the list were counted and arranged from high to low.

2.2.2.2 The second phase of corpus development

After the computer and technology had been developed, corpus was collected by computer, more frequently. The word 'frequency lists' became more popular among English teachers and learners. Fernández and Faber (2011) stated that the utilities and facilities of computer techniques on lexicon approach or corpus can help ESP students to find and learn the real use of words relating to their study field. Due to the advance of computer, a large electronic data was stored systematically. For example, the use of computer was beneficial for the collection of dictionary. In 1987, Collins Cobuild English Language Dictionary, the first English dictionary gathered by computer-based corpus was published. Since then many dictionaries have been created by text corpora; for instance, Oxford Advanced Learner's Dictionary of Current English, Longman Dictionary of Contemporary English, and Collins COBUILD Advanced Learner's Dictionary. Corpus could be further created as online dictionaries and lexical resources (Fernández & Faber, 2011).

Overall, the two phases of corpus development are first manually collecting by hands and later, automatically, collecting by computers. The development of corpus was influential to linguistic exploration because the corpus is used as a tool to collect linguistic data such as words or vocabulary for teaching and learning.

2.2.3 Concordance Software Program

The corpus linguist uses concordance programs as a tool for investigating and analyzing linguistic data. Since most corpora contain a large quantity of data, it is difficult to search for the detail in a corpus without the help of a computer. Concordance programs are able to search for the electronic data automatically such as word, phrase, or even word in a sentence. There are various concordance software programs such as Wordsmith (Scott, 1996), MonoConc Pro (Barlow, 2002), and AntConc (Anthony, 2014) which have similar features.

2.2.3.1 The features and functions of concordance software

The features and functions of concordance software namely Wordsmith (Scott, 1996), MonoConc Pro (Barlow, 2003), and AntConc (Anthony, 2014) were similar. These programs can install on Windows, Macintosh, and Linux operation. In addition, the software can process texts in several languages such as Chinese, Japanese, and Korean. Moreover, the concordance software is practical and useful for multi-purpose as a word collection. They have functions for creating word lists and a keyword tool which can identify words frequency within a corpus. They are also an analysis device which is essential for text analysis such as cluster and collocate tools. Most concordance software applications were commercial programs, and there is no diskette version available for software outlet. However, there is an exception for AntConc.

2.2.3.2 Development of the AntConc program

AntConc was created by Laurence Anthony of Waseda University in Tokyo, Japan. The first version of AntConc was released in 2002. The program has been continuously improved and developed, and the latest version is AntConc 3.4.3 (Anthony, 2014). The recent version is advantageous and described as "A user-friendly concordancer with an intuitive interface" (Maher, 2008). The

program can count the running words and make a word list. It can show and arrange the word by either frequencies or alphabets, and is free and available for download. Thus, AntConc was convenient to installation since it was ready to use. See Table 2.3.

Table 2.3 The Overview Functions and Features of AntConc

<ul style="list-style-type: none"> • Freeware License • Small memory requirement (2 B of disk space) • Multiplatform <ul style="list-style-type: none"> - Windows 98 or newer - Unix / linux • Extensive set of text analysis tools <ul style="list-style-type: none"> - KWLC Concordance - Original File View - Word Clusters / Lexical Bundles - Word lists <ul style="list-style-type: none"> - Keyword lists 	<ul style="list-style-type: none"> • Powerful Search Features <ul style="list-style-type: none"> - Regular Expressions (REGEX) - Extensive Wildcards • Multiple-Level Sorting • HTML / XML Tag Handling • Unicode Support • Easy-to-use, intuitive GUI
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

2.2.3.3 Reliability and Validity of AntConc

The AntConc program claimed that it was one of the popular concordance programs that most linguists apply to analyze linguistic data such as words and sentences because the program is available for online download and can be used easily (Diniz, 2005). Froehlich (2015) indicated that Antconc is able to process the data such as frequency, structure, and collocation of the target word or sentence which the user needs to analyze. In addition, Tribble (2012) compared software programs used to analyze corpora, and demonstrated that AntConc was one of the most popular concordance software programs (Figure 2.4).

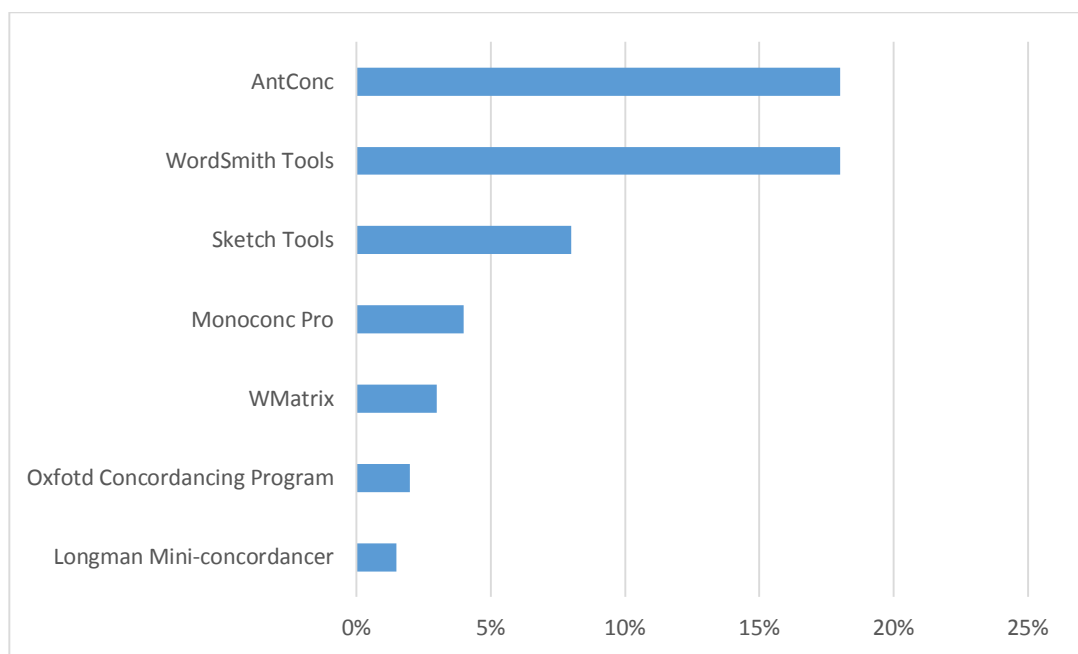


Figure 2.4 The comparison of software programs used to analyze corpora (cited in Anthony, 2013)

Figure 2.4 indicated that WordSmith Tools (Scott, 2012) and AntConc (Anthony, 2014) were the software programs mostly used to analyze corpora. These two programs are available for the researchers who want to process raw data to create own corpus. However, WordSmith Tools come with a restrictive commercial license in a diskette that requires user to pay for the program. The commercial license prevents the user from using the program on other laptops. In contrast, the AntConc program is a freeware tool that is available to download on their Internet website. It can process raw corpus data of various kinds and can be used for diverse purposes such as counting word frequencies, finding concordance lines of the target words, drawing a concordance plot, and collocating the word. As a result, it can be useful in many contexts.

To sum up, concordance programs are generally used for word queries and searching for word combinations within a context. The program also makes a word list of collocates or frequency lists. However, AntConc is different from other concordance programs because it is freeware and an off-line program which is available for online download. Consequently, it is easy and practical to use.

2.3 Word Lists in a Corpus

Corpus is a collection of words that can provide specific words and usage in a particular field (Nation, 2001). Nation (2001) stated that we can distinguish four kinds of vocabulary in the text: high-frequency words, academic words, technical words, and low-frequency words. According to West (1985), the high frequency words refer to General Service List (GSL) which contains 2000 most frequent words. Academic word List (AWL) are the words occurs in academic disciplines. Technical words are the words used in a particular discipline, or the words that do not occur in GSL or AWL. That is to say, technical words refer to the outside words list (OWL). In addition, the low frequency words occur with low frequency or are rarely found in the corpus. Thus, the present corpus-based study did not include low frequency words in the corpus. To conclude, words in the present corpus are classified into the General Service List (GSL), the Academic Word List (AWL), and Outside Word List (OWL).

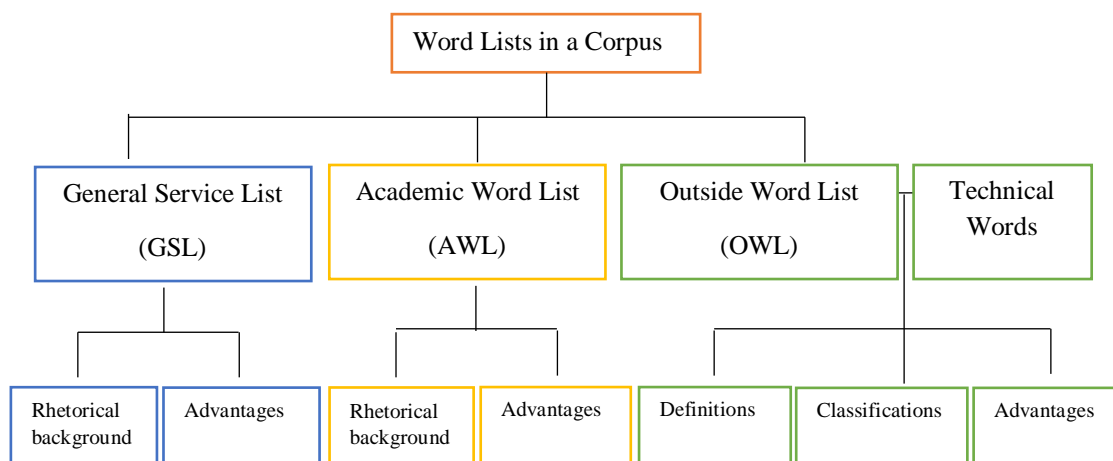


Figure 2.5 Word lists in a corpus

GSL and AWL were first explored by West (1953) and Coxhead (2000). However, the new versions of New General Service List (NGSL) and New Academic Word List (NAWL) have been recently revised by Brown, Culligan, and Phillips (2013). See Table 2.4 below.

Table 2.4 The Comparison of Word Lists in the Corpus

Word Lists	Author	Field	Number of word families	Coverage
GSL	West (1953)	General texts	2000	78-80%
AWL	Coxhead, (2000)	Academic texts	570	8-10%
NGSL and NAWL	Brown, Culligan, and Phillips (2013)	General texts	3,331	92%
OWL	-	Specialize texts	-	-

GSL = General Service List
 AWL = Academic Word List
 NGSL = New General Service List
 NAWL = New Academic Word List
 OWL = Outside Words List

Table 2.4 showed the comparison of word lists in the corpus. From the table, the new revised versions of GSL and AWL cover 92% of text coverage which have around 5% more than the old version of GSL and AWL. GSL refers to the words appear in the general texts, and AWL refers to the words appear in the academic texts. Words, which appear in neither GSL nor AWL, belong to OWL. In other words, OWL refers to specialized words or technical terms used in the particular discipline.

2.3.1 General Service List (GSL)

In 1953, Michael West published his English words called the General Service List (GSL). The GSL is the primary principal word list in teaching and learning (Nation, 2001) and plays an important role for English language learners and ESL teachers as a fundamental word list.

2.3.1.1 The background of General Service List (GSL)

The GSL was created by West (1953). The GSL consisted of general or non-specialized words which appeared frequently in general texts (Coxhead & Nation, 2001; Nation, 2001; Brezina & Gablasova, 2013). There were 2000 word families which were collected from a corpus of written English. The GSL includes both content and function words which covers around 80% of running words in

context (Nation, 2001). The content words are noun, verb, adjective, and adverb. Fires (1952) claimed that learning content words increase students' lexical knowledge because the content words contain the core meanings of message. On the other hand, the function words referred to adverbial particles, modal verbs, prepositions, conjunctions, determiners, pronoun, and numbers (Nation, 2001; Murphy, 2010). The high-frequency words in GSL were mostly composed of function words. See Table 2.5.

Table 2.5 Top-Forty Most Frequent Words in GSL (West, 1953)

Rank	Word	<i>f</i>	Rank	Word	<i>f</i>	Rank	Word	<i>f</i>	Rank	Word	<i>F</i>
1	the	69975	11	that	10790	21	by	5307	31	would	3062
2	be	39175	12	for	9495	22	this	5146	32	all	3002
3	of	36432	13	they	8555	23	we	4931	33	will	2903
4	and	28872	14	I	8516	24	you	4695	34	there	2851
5	a	26800	15	with	7289	25	do	4389	35	say	2793
6	to	26190	16	as	7254	26	but	4383	36	who	2678
7	in	21338	17	not	6976	27	from	4371	37	make	2378
8	he	20033	18	on	6742	28	or	4204	38	when	2333
9	have	12458	19	she	6164	29	which	3560	39	can	2218
10	it	11247	20	at	5377	30	one	3480	40	more	2203

Table 2.5 showed that most of top-forty highest frequent words in GSL (West, 1953) were function words. The article 'the' was used with the highest frequency in the corpus, and it appeared around more twice times than 'be' which was the second highest frequency.

In addition, Schmitt (2000) suggested that not only just frequency count, GSL was more helpful in terms of the different parts of speech and the different meanings of word were listed. See examples below.

Table 2.6 The Examples of GSL

Head words	Part of speech	<i>f</i>	Meaning	%
able	adj.	940	(1) <i>Able</i> to take care of	89%
			(2) An <i>able</i> man [Able-bodied, 0.1%]	11%
ability	n.	340	(1) (cleverness) A man of great <i>ability</i> Musical <i>ability</i>	58%
			(2) (<i>power to</i>) <i>Ability</i> to learn [legal sense, 2%]	40%

adj. = adjective

n. = noun

% = frequency of use

For example, the use of the family word *able* can be seen in two forms: *able* (adj.) and *ability* (n.). However, each form of the family word *able* contains more than one meaning. For example, *able* is to take care of, and *able* is able-bodied. Between these two meanings, *able* as ‘to take care of’ is used for 89% of these two meanings.

2.3.1.2 The advantages of GSL

The GSL has the advantage to become the fundamental English vocabulary used in teaching and learning materials for foreign and second language learning (Coxhead & Nation, 2001; Nation, 2001; Brezina & Gablasova, 2013). The list was used to produce materials, especially reading books and referenced books such as ESL course books and dictionaries for graded readers. Although the GSL was created for decades, it is applicable for learning because of its availability. The list provided text coverage and frequency of meanings which indicate the actual uses of word.

Briefly, the General Service List (GSL) is a list of words which were selected from the corpus of written English because they are the 'general' words used for learning English. The frequency in English was the factor used to consider when

choosing the words. Moreover, the GSL shows the frequency of words including part of speech and various meanings of word sense. The GSL is the basic word list suitable for the initial vocabulary inputs, and therefore, was adopted for material teaching.

2.3.2 Academic Word List (AWL)

Academic Word List (AWL) is the word list that does not belong to the most frequent 2000 words of GSL. The AWL and its sublists were created by Averil Coxhead in order to prepare learners for tertiary level study in a university. The rhetorical background of Coxhead's AWL has been published twice, and has been the prototype of the research in AWL.

2.3.2.1 The background of AWL

The Academic Word List (AWL) was first published in 1998, and was later revised in 2000. It consisted of academic words which did not generally occur except for a wide range of academic texts from different disciplines (Nation, 2001). In other words, academic words were the words that have specialized meanings in a scientific context (Childs & O'Farrell, 2003; Strevens et al., 2000). Coxhead's AWL contained 570 word families which were based on a corpus of 3.5 million tokens. It provided approximately 8 to 10% of text coverage in academic reading, 3.9 percent of newspapers, and 1.7 percent of fiction. Moreover, there were 28 subjects which were categorized into four disciplines namely: commerce, law, arts, and science. There were 10 sub-lists in the AWL. Sublist 1 contained the most common used words that appeared with the highest frequency in the corpus. On the other hand, Sublist 10 contained the least used words with less frequency. Each sublist contained 60 word families except Sublist 10. It contained only 30 word families. See examples in Table 2.7

Table 2.7 The Most Frequent Members of the Word Families in Each Sublist

Sublist 1	Sublist 2	Sublist 3	Sublist 4	Sublist 5
analysis	achieve	alternative	access	academic
approach	acquisition	circumstances	adequate	adjustment
area	administration	comments	annual	alter
assessment	affect	compensation	apparent	amendment
assume	appropriate	components	approximated	aware
authority	aspects	consent	attitudes	capacity
available	assistance	considerable	attributed	challenge
benefit	categories	constant	civil	clause
concept	chapter	constraints	code	compounds
consistent	commission	contribution	commitment	conflict
constitutional	community	convention	communication	consultation
context	complex	coordination	concentration	contact
contract	computer	core	conference	decline
create	conclusion	corporate	contrast	discretion
data	conduct	corresponding	cycle	draft
definition	consequences	criteria	debate	enable
derived	construction	deduction	despite	energy
distribution	consumer	demonstrate	dimensions	enforcement
economic	credit	document	domestic	entities
environment	cultural	dominant	emerged	equivalent
established	design	emphasis	error	evolution
estimate	distinction	ensure	ethnic	expansion
evidence	elements	excluded	goals	exposure
export	equation	framework	granted	external
factors	evaluation	funds	hence	facilitate
financial	features	illustrated	hypothesis	fundamental
formula	final	immigration	implementation	generated
function	focus	implies	implications	generation
identified	impact	initial	imposed	image
income	injury	instance	integration	liberal
indicate	institute	interaction	internal	licence
individual	investment	justification	investigation	logic
interpretation	items	layer	job	marginal
involved	journal	link	label	medical
issues	maintenance	location	mechanism	mental
labour	normal	maximum	obvious	modified
legal	obtained	minorities	occupational	monitoring
legislation	participation	negative	option	network
major	perceived	outcomes	output	notion
method	positive	partnership	overall	objective
occur	potential	philosophy	parallel	orientation
percent	previous	physical	parameters	perspective
period	primary	proportion	phase	precise
policy	purchase	published	predicted	prime
principle	range	reaction	principal	psychology
procedure	region	registered	prior	pursue
process	regulations	reliance	professional	ratio
required	relevant	removed	project	rejected
research	resident	scheme	promote	revenue
response	resources	sequence	regime	stability
role	restricted	sex	resolution	styles
section	security	shift	retained	substitution
sector	sought	specified	series	sustainable
significant	select	sufficient	statistics	symbolic
similar	site	task	status	target
source	strategies	technical	stress	transition
specific	survey	techniques	subsequent	trend
structure	text	technology	sum	version
theory	traditional	validity	summary	welfare
variables	transfer	volume	undertaken	whereas

Table 2.7 The Most Frequent Members of the Word Families in Each Sublist (cont.)

Sublist 6	Sublist 7	Sublist 8	Sublist 9	Sublist 10
abstract	adaptation	abandon	accommodation	adjacent
accurate	adults	accompanied	analogous	albeit
acknowledged	advocate	accumulation	anticipated	assembly
aggregate	aid	ambiguous	assurance	collapse
allocation	channel	appendix	attained	colleagues
assigned	chemical	appreciation	behalf	compiled
attached	classical	arbitrary	bulk	conceived
author	comprehensive	automatically	ceases	convinced
bond	comprise	bias	coherence	depression
brief	confirmed	chart	coincide	encountered
capable	contrary	clarity	commenced	enormous
cited	converted	conformity	incompatible	forthcoming
cooperative	couple	commodity	concurrent	inclination
discrimination	decades	complement	confined	integrity
display	definite	contemporary	controversy	intrinsic
diversity	deny	contradiction	conversely	invoked
domain	differentiation	crucial	device	levy
edition	disposal	currency	devoted	likewise
enhanced	dynamic	denote	diminished	nonetheless
estate	eliminate	detected	distorted/distortion -	notwithstanding
exceed	empirical	deviation	equal figures	odd
expert	equipment	displacement	duration	ongoing
explicit	extract	dramatic	erosion	panel
federal	file	eventually	ethical	persistent
fees	finite	exhibit	format	posed
flexibility	foundation	exploitation	founded	reluctant
furthermore	global	fluctuations	inherent	so-called
gender	grade	guidelines	insights	straightforward
ignored	guarantee	highlighted	integral	undergo
incentive	hierarchical	implicit	intermediate	whereby
incidence	identical	induced	manual	
incorporated	ideology	inevitably	mature	
index	inferred	infrastructure	mediation	
inhibition	innovation	inspection	medium	
initiatives	insert	intensity	military	
input	intervention	manipulation	minimal	
instructions	isolated	minimised	mutual	
intelligence	media	nuclear	norms	
interval	mode	offset	overlap	
lecture	paradigm	paragraph	passive	
migration	phenomenon	plus	portion	
minimum	priority	practitioners	preliminary	
ministry	prohibited	predominantly	protocol	
motivation	publication	prospect	qualitative	
neutral	quotation	radical	refine	
nevertheless	release	random	relaxed	
overseas	reverse	reinforced	restraints	
preceding	simulation	restore	revolution	
presumption	solely	revision	rigid	
rational	somewhat	schedule	route	
recovery	submitted	tension	scenario	
revealed	successive	termination	sphere	
scope	survive	theme	subordinate	
subsidiary	thesis	thereby	supplementary	
tapes	topic	uniform	suspended	
trace	transmission	vehicle	team	
transformation	ultimately	via	temporary	
transport	unique	virtually	trigger	
underlying	visible	widespread	unified	
utility	voluntary	visual	violation	

(Coxhead, 2000, p.232)

2.3.2.2 The previous studies of AWL

Research has been widely conducted in academic word lists in specific disciplines and investigated for linguistic use. The discipline-based lexical studies focused on the academic vocabulary relating to disciplines such as medical science, chemistry, agricultures, nursing, and environmental science. Most studies examined academic words used in research articles from different disciplines. For examples, Chen and Ge (2007) and Wang, Liang, and Ge (2008) created a Medical Academic Word List which was noteworthy as the most common medical academic words used in medical research papers. In a chemistry corpus, Valipouri and Nassaji (2013) found that AWL in chemistry covered 9.06 %. Martínez, Beck, and Panza (2009) also compiled an Agricultural Academic Word List including 92 word families. In addition, the recent corpora of nursing (Yang, 2015) and environmental science (Liu & Han, 2015) found that AWL in the nursing' corpus covered 13.64%; while AWL in the environmental science corpus covered 12.82%. See Table 2.8.

Table 2.8 The Previous Research of Corpus-Based Studies

Authors	The study field of corpus	Number of running words	Findings of text coverage (%)
Kaewphanngam (2002)	psychology	236,086	AWL (7.78%)
Chen and Ge (2007)	medical	190,425	AWL (10.07%)
Wang, Liang, and Ge (2008)	medical	1,093,011	AWL (12.24%)
Martínez, Beck, and Panza (2009)	agricultural sciences	826,416	AWL (9.06%)
Vongpumivitch, Huang and Chang (2009)	applied linguistics	1,500,000	AWL (11.17%)
Valipouri and Nassaji (2013)	chemistry	4,000,000	AWL (9.60%)
Liu and Han (2015)	environmental science	862,242	AWL (12.82%)
Yang (2015)	nursing	1,006,934	AWL (13.64%)

Other studies investigated the use of words through linguistic approach such as discourse analysis. For examples, academic and non-academic words

were investigated in applied linguistics research articles (Vongpumivitch, Huang, & Chang, 2009). Their Applied Linguistics Academic Word List consisted of 603 word families. Others have explored linguistic features namely discourse analysis through academic words. For example, one aspect of linguistic exploration was in scientific discourse (Cabr , 1999; Bowker & Pearson, 2002; Halliday & Webster, 2002). They verified that different disciplines of research articles contained different types of vocabulary used. In short, the previous research showed that the AWL was investigated in two aspects: 1) collecting the words in a specific discipline and 2) examining the use of words in a specific discipline.

In summary, the AWL created by Coxhead (2000) was different from the most frequency 2000 word in GSL. The 570 words in AWL were gathered because they occurred with high frequency in a broad range of academic texts, and thus it was used by teachers to teach at colleges and universities. In addition, the studies of AWL were investigated in various disciplines because the diversity of academic words in each discipline was different.

2.3.3 New General Word List (NGSL) and New Academic Word List (NAWL)

GSL and AWL were created by West (1953) and Coxhead (2000). Although AWL was revised in 2000, the new version of GSL and AWL were recently created by Browne, Culligan, and Phillips (2013). In order to see the development of GSL and AWL, the old GSL (West, 1953) and AWL (Coxhead, 2000) were compared to the new GSL and AWL (Browne, Culligan, & Phillips, 2013).

It is noted that the combinations of GSL and AWL cover approximately 87% of the running words in a text (Coxhead, 2000; Coxhead & Nation, 2001; Nation, 2001). The GSL coverage was 84%; while, the AWL covered 8% to 10% of the text. Browne, Culligan, and Phillips (2013) have created the New General Word List (NGSL) and New Academic Word List (NAWL) (Table 2.3).

The GSL and AWL were recently revised by Browne, Culligan, and Phillips in 2013 and named as the New General Service List (NGSL) and the New Academic Word List (NAWL). These word lists are considered as the newest versions and provide more coverage. There were 2,368 headwords for NGSL and 963

headwords for NAWL. The NGSL coverage was 90% which was higher than the GSL (84%). Therefore, the total of NGAL and NAWL provided an additional 5% of text coverage than GSL (West, 1953) and AWL (Coxhead, 2000) (Table 2.8).

In conclusion, the revised versions of GSL and AWL comprised by Browne, Culligan, and Phillips (2013) were more reliable due to their updated versions and increased coverage. Due to the additional coverage of NGSL and NAWL, their lists of words can increase learners' vocabulary knowledge (Browne, Culligan, & Phillips, 2013).

2.3.4 Outside Word List (OWL)

Although the General Service List (GSL) and the Academic Word List (AWL) are the frequently used words, the Outside Word List (OWL) is also necessary for specific groups of learners due to following reasons. First, the outside word list is similar to technical words that particularly occur and are used in particular disciplines (Nation, 2001). In other words, the OWL is the word list that does not belong to the GSL and AWL. Second, several research had already examined high frequency words namely the GSL and AWL, but words for particular group of learners and discipline especially the field of laboratory animals were rarely analyzed in depth. Therefore, the OWL or technical words were mainly emphasized. Technical words are further explained below.

2.3.5 Technical Words

Technical words are the common used words which are used in a particular area of knowledge. Within a certain discipline, technical words may frequently occur, but they are unlikely to occur in general or outside disciplines (Nation, 2001; Nation & Chung 2004). Nation (2001) mentioned that students who learn in ESP course need to know and comprehend words used in their field of study because each specific field has their own characteristics of specialized words used.

2.3.5.1 Definitions of technical words

Technical words are defined from two perspectives. The first is *applied linguistics* (Chung & Nation, 2004; Coxhead & Nation, 2001; Nation, 2001). Nation (2001) observed that there was little research providing “a consistently applied

operational definition of what words are technical words” (p. 18). For applied linguistics field, Chung and Nation (2004) defined technical words as the words that occurred frequently in a particular context and in a specialized discipline but appeared infrequently in other fields. Although it is quite difficult to calculate the exact total number of technical words in a discipline, Coxhead and Nation (2001) claimed that the technical words in a discipline accounts for approximately 1,000 words or less. This was the typical number of vocabulary in technical dictionaries. The examples of technical words were acronyms, abbreviations, chemical formulas, symbols, and words related to a specific discipline (Nation, 2001). The second perspective is *terminology* (Pearson, 1998; Cabré, 1999; Bowker & Pearson, 2002). Terminology perspective examines the difference between terms and words. Terms or technical words are defined as a distinction of meaningful units in specialized languages. Cabré (1999) defined technical words as “used to designate concepts pertaining to special disciplines and activities”. In short, the definitions of technical words refer to the words that occur frequently in its own discipline rather than in others, and they contain specialized meaning used exclusively in that discipline.

2.3.5.2 The classifications of technical words

The classifications of technical words can be categorized by three methods: 1) using statistical devices, 2) using GSL and AWL as referenced lists, and 3) using a specialized dictionary.

Statistical devices

Technical words can be identified by statistical devices such as frequency and range. Cabré (1999), Bowker and Pearson (2002), and Chung and Nation (2004) suggested that statistical formulas and quantitative data, and using term definitions were the methods used to separate technical words in the corpus. The study could compare the frequency and range of words in corpora between different disciplines for collecting quantitative data. By counting word frequency, the researcher was able to know how the word was frequently used in the investigated area. In addition, range can be used to identify the various uses of the word in different areas or text coverage by certain word lists.

GSL and AWL as referenced lists

GSL and AWL can be used as referenced lists to classify technical words. For example, Coxhead (2011) selected technical vocabulary using the GSL and the AWL as referenced lists, as same as Coxhead and Hirsh (2007). This method was based on the assumption on the boundary of GSL and AWL; therefore, the specialized vocabulary which occurred outside the GSL and the AWL is assumed as technical words. Moreover, the development of lexical programs was developed to categorize words according word's levels. For instance, VocabProfile created by Tom Cobb in 2003 was used to categorize the words into GSL, AWL, and OWL. The database of the program was developed and updated from the new findings of GSL and AWL (Browne, Culligan & Phillips, 2013).

A specialized dictionary

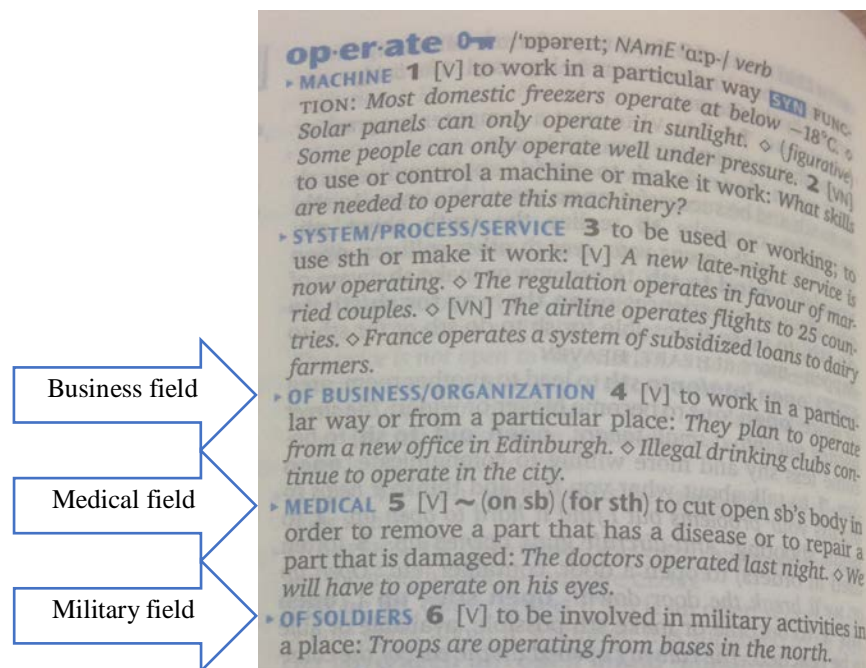
According to Nation (20001), words were checked against a specialized dictionary in order to verify technical meanings. For example, Kaewphanngam (2002) and Pitukwong (2012) compiled a business corpus and a psychology corpus by using specialized dictionaries relevant to the field of study, as a criterion to select technical words. Dictionaries were not only used for looking up spellings and meanings of a broad selection of terms, but also covered a relatively restricted definition used in a particular area.

Briefly, technical word classification can be identified by formulas or quantitative data, the GSL and AWL boundary, specialized dictionary. The criteria of word classifications were flexibly employed to separate words in a corpus according to the appropriateness of learners' levels and the purpose of the study.

2.3.5.3 The advantages of technical words

The advantages of technical words were distributed to ESP learners, ESP material design, and EAP course (Chung & Nation, 2004). The restriction of meaning that was used differently from the usual can appear in a particular discipline. Thus, technical words play a vital role for ESP and EAP course design especially for learning material. The advantages of technical words are as following.

First, technical words are essential for students who are in the area of English for Specific Purposes (ESP) (Chung & Nation, 2004; Chujo & Utiyama, 2006; Kwary, 2011). As each discipline of the study has own characteristics including technical lexical items, ESP learners are required to understand these lexical in their field (Nation, 2001). That is to say, specialized vocabulary contains different specific meanings when it is used in the particular discipline. See the example of the word ‘operate’.



(Hornby, 2005, p. 1064)

Figure 2.6 Meanings of the word ‘operate’

Second, in order to designing ESP teaching materials, identifying technical terms is an important stage because there are no well-established approaches for determining which words are technical terms or not (Chung & Nation, 2004, p. 251). To rectify this problem, a corpus study needs to undertake to find a systematic, reliable, and practical approach for identifying technical words in specific context.

Third, technical words are essential for EAP course. Donley and Reppen (2001, p.7) stated that specific words are useful for learners in a particular discipline. The words can facilitate and prepare learners when they learn in EAP

course. For example, technical words in a scientific English context require not only English knowledge but also scientific knowledge. Therefore, technical words are useful for learners if they are taught within the content area of relevant discipline. All in all, technical words are beneficial for ESP learners, ESP material design, and EAP course.

In summary, the OWL can be referred to the technical words used in a particular discipline. The OWL can be classified by using GSL and AWL as a referenced list; therefore, the words that do not appear in GSL or AWL are assumed as OWL or technical words. The uses of OWL are significant to ESP teaching and learning since the words are technically used in a specific meaning.

2.4 Collocations

Collocation refers to the sequence of words or terms that two words usually occurs together (Benson, 1985). The problems of learning English language can normally occur such as grammatical errors, insufficient lexical knowledge, or pronunciation. These problems can be resolved by learning from grammar books or consulting dictionaries. When the problems are collocations, it is not easy to find out the answers because collocation errors are difficult to identify especially by a non-native speaker (Lubis, 2013). In a corpus, word lists provide only word frequency, but it is necessary to know how words are used. Therefore, collocations of words should be exemplified for the natural use of words.

2.4.1 The importance of Collocations

The other keystone of vocabulary learning is collocations. Collocation is important since it can assist learners for producing productive skills such as writing and speaking (Wray, 2002) and naturalize learners' language (Nation, 2001).

2.4.1.1 Collocation knowledge assisting learners' procuring productive skills

High-frequency words and meanings may be adequate for receptive skills (listening and reading); however, knowing their language feature uses such as their collocations can help in productive skills (writing and speaking)

(Nations, 2001; Wray, 2002). Collocations associate with writing and speaking as a level of word, phrase, and sentence production. Moreover, knowing word collocations, learners can produce sentences naturally (Nations, 2001). Collocations or formulaic sequence is emphasized on communicative competence (Barfield & Gyllstad, 2009; Nation, 2001; Schmitt, 2004; Wood, 2010; Wray, 2002). Word collocations can facilitate speakers to use language fluently (Pawley & Syder, 1983) and to achieve basic communicative needs (Wray, 2002) because collocation is a string of words that is able to use automatically.

2.4.1.2 Collocation knowledge assisting learners to generate language naturally

Nation (2001) stated that collocation knowledge can help to produce language spontaneously. For native speakers, “knowledge of acceptable and unacceptable collocations is largely instinctive” (Conzett, 2000). Unlike L2 learners, collocation in L1 interferes the production of collocation in L2. Due to their L1 interference with L2, they seem to have problems in collocations especially the beginning students (Yuan, 2014). Furthermore, Celce-Murcia and Olshtain (2000) suggested that L2 learners usually used incorrect collocations and combined the unsuitable words together. Lubis (2013) supported Celce-Murcia and Olshtain (2000) by showing the common errors of collocations among Indonesian students as following.

Common mistakes

take decision
drive a motorcycle
find money
work homework

Correct collocation

make decision
ride a motorcycle
make money
do homework

(Lubis, 2013, p.24)

Lubis (2013) suggested that the collocation errors are attributed to the students’ ignorance about collocation. They were not aware that wrong collocations could cause unnatural language. The ignorance about collocation caused them to do literal or direct translations which produced wrong collocations. The examples above showed that the students directly translated verbs into English. The students did not know that in English the verb ‘take’ does not collocate with decision but with ‘make’. Similarly, ‘money’ does not collocate with ‘find’ because

‘find’ means ‘to discover’. Briefly, students made wrong collocation due to the differences between L1 and English language and because they were ignorant about English collocation.

In summary, collocations can help learners to develop skills to language, naturally. In other words, students can apply vocabulary in their productive skills, naturally.

2.4.2 Perspectives of Collocations

J.R. Firth defined collocation as ‘the company words keep’ or ‘the relationships with other words’ (cited in Hill, 2000). Another definition of collocation referred to ‘the way words combine in predictable ways’ (Hill, 2000). Due to the large number of words in English, millions of word combinations are possible. Therefore, it is significant to become aware of the nature of collocations that can be usually put together in English. Collocations can be viewed from two perspectives: 1) strength (Hill, 2000) and 2) patterns (Benson, 1985; Lewis, 2000)

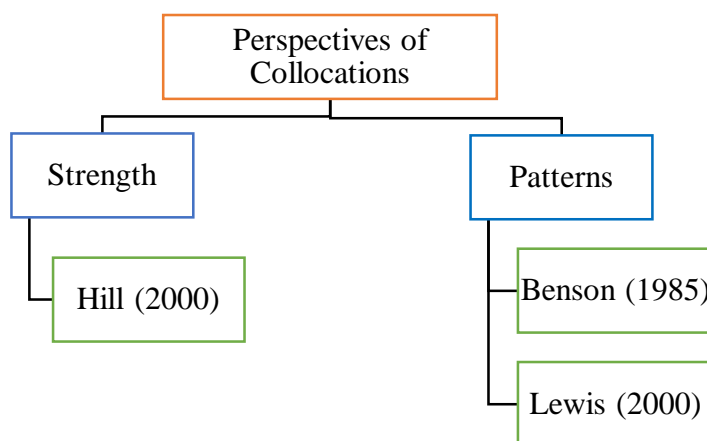


Figure 2.7 Perspectives of collocations

Collocations viewed by its strength

Hill (2000) viewed collocation by its strength. In other words, collocations are considered by how fixed and flexible words are put together. Consequently, Hill (2000) categorized collocations into two types: fixed and flexible collocations. ‘Fixed-meaning collocations’ or ‘Unique collocations’ refer to a group of words stuck together. It cannot be changeable.

Examples of fixed-meaning collocations

Kill two birds with one stone
Time and tide wait for no man
whisk an egg
curly hair
winding road
blissfully ignorant

On the other hand, 'Flexible-meaning collocation' or 'Weak collocation' refers to a group of words joined together, but changeable or replaceable by other words (Hill, 2000).

Examples of flexible-meaning collocations

extremely/ particularly/ deeply/ statistically + significant
big /enormous/ large + house/ lorry/ cup
fast/ shiny/ expensive + car/motorbike/ airplane
very/ really/ extremely + interesting/ hot/ generous
brown/ straight/long + fence/ hair/ line

It is noted that, in fixed-meaning collocations, very few words can collocate with the head noun. Flexible-meaning collocations can collocate with several words.

Collocations viewed by its patterns

Benson (1958) and Lewis (2000) provided patterns of collocations considered by parts of speech. Benson categorized collocation into six patterns called grammatical collocation and lexical collocations. Lewis added more patterns of collocations which included phrases and idioms (see Table 2.9).

Table 2.9 Examples of Collocations' Patterns

Patterns of collocations	Benson (1985)	Lewis (2000)
	Grammatical collocations	
1. noun + preposition	abstinence from, access to, accusation against, acquiescence in (to), admiration for	-
2. adjective + preposition	absent from, accountable to, acquainted with, adept at (in), adequate for (to)	aware of
3. verb + preposition	abide by, abstain from, account for, approve of	turn in
	Lexical collocations	
4. verb + noun	set an alarm, roll a hoop, fly a kite, launch a missile	submit a report
5. noun + verb	adjectives modify, bells ring	the fog closed in
6. adjective + noun	a confirmed bachelor, a pitched battle, pure chance, a responsive chord	a difficult decision
	Others	
7. noun + noun	-	radio station
8. verb + adverb	-	examine thoroughly
9. adverb + adjective	-	extremely inconvenient
10. verb + adjective + noun	-	revise the original plan
11. compound noun	-	fire escape
12. binominal	-	backwards and forwards
13. discourse marker	-	To put it another way
14. multi-word prepositional phrase	-	a few years ago
15. trinominal	-	hook, line and sinker
16. fixed phrase	-	On the other hand
17. incomplete fixed phrase	-	A sort of ...
18. fixed expression	-	Not half!
19. semi-fixed expression	-	See you later / tomorrow / on Monday
20. part of a proverb	-	Too many cooks ...
21. part of a quotation	-	To be or not to be ...

Benson (1985) suggested that collocations were divided into two principle types: 1) grammatical or syntactic collocations and 2) lexical or semantic collocations.

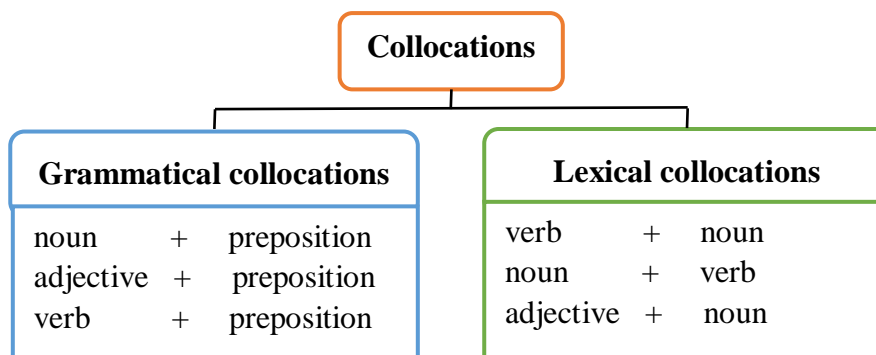


Figure 2.8 Grammatical and lexical collocations (Benson, 1985)

Grammatical collocations refer to a noun, verb, or adjective followed by a particle such as a preposition. The lexical collocations are the combination of two equal words such as noun + verb, verb + noun, and adjective + noun.

However, Lewis (2000) suggested more patterns of collocations. From the Table 2.8, it shows that the definitions of collocation are very wide, and cover many different patterns including a string of words such as expression, proverb, idiom, and quotation. Overall, collocations refer to the group of words that are regularly found together.

To summarize, collocations generally refer to a group of words usually occurring together (Benson, 1985; Hill, 2000; Lewis, 2000). Collocations can be viewed by the strength and pattern of words joined together. For example, fixed and flexible collocations were based on how words were put and stayed together, strongly (Hill, 2000), while, grammatical and lexical collocations were considered by patterns of words co-occurring (Benson, 1985; Lewis, 2000). This current study adopted the framework of grammatical and lexical collocations (Benson, 1985) because it aimed to find the patterns of collocations used in laboratory review articles. As the source of investigated corpus was compiled from review articles which were written in formal language, some patterns of collocations such as expressions and proverbs rarely occur, and some patterns proposed by Lewis (2000) were not adopted in this study. In addition, by analyzing patterns of collocations, the researcher found some grammatical features used in the review articles.

2.4.3 Grammatical features of collocations

From the analyzing this laboratory animal corpus by using AntConc program, ‘zebrafish’ was the highest-frequency word in the corpus. Thus, ‘zebrafish’ was the target word of the study. Moreover, five patterns of collocations that collocated with the target word (zebrafish) were preposition, noun, verb and adjective. To understand how word collocations were used to form the sentence, the grammatical features which related to the target word were reviewed. The grammatical features of the target word are semantic functions of prepositions, compound nouns, active and passive voices, and types of adjectives.

Semantic functions of prepositions

The common semantic functions of prepositions relate to the concept of space, time, and degree. In order to illustrate this functions, Table 2.11 demonstrates the use of prepositions ‘in’, ‘with’, and ‘from’ which exhibit space, time, degree, and other usages.

Table 2.10 The Semantic Functions of Prepositions ‘In’, ‘With’, and ‘From’

Prep.	Space	Time	Degree	Other usages
in	<i>enclosure</i> : The man is in the room	<i>in a period</i> : WW II ended in 1945. <i>future appointment</i> : Come in 10 minutes.	-	<i>currency</i> : Pay me in dollars. <i>language</i> : Write, say it in English.
with	<i>alongside, near</i> : even with the wall	<i>together</i> : He grew wiser with the years. He rises with the chickens. work from 9 ton 5	<i>equal standing or ability</i> : rank with the best; run with the fastest. from 60 to 80 degrees from 5 to 7 dollars	<i>in regard to</i> : pleased with the gift <i>manner</i> : spoke with ease.
from	<i>a starting point</i> : We traveled from N.Y. to L.A. <i>origin</i> : man from New York	-		<i>source</i> : paper is made from wood <i>cause</i> : wet from the rain

prep. = preposition

(Celce-Murcia & Larsen-Freeman, 1999, pp.257-258)

Compound nouns

A compound noun is a noun that is formed by two or more words. Most compound nouns in English are made up by nouns modified by other nouns. The types of compound noun were categorized in two types: 1) primary compound noun and 2) secondary compound noun (McCarthy, 2002).

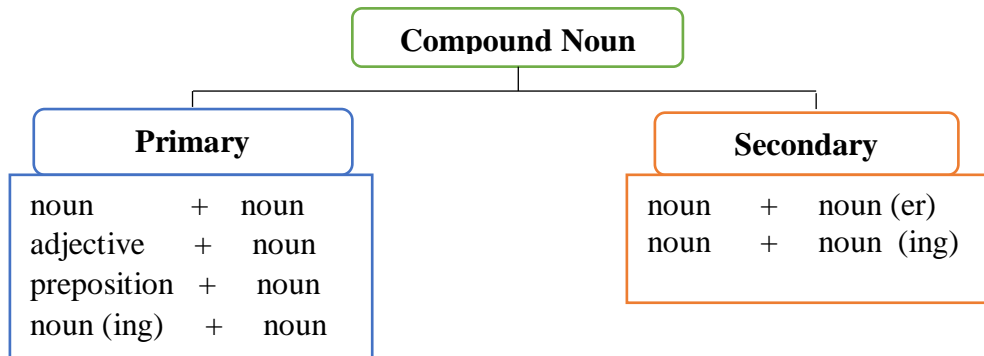


Figure 2.9 Types of compound noun

1. *Primary compound noun* refers to the head noun that is modified by noun, adjective, preposition, and noun with a morpheme (ing). See Table 2.11.

Table 2.11 The Examples of Primary Compound Nouns

noun + noun	adjective + noun	preposition + noun	noun (ing) + noun
Football	full moon	underworld	drawing room
bedroom	blackboard	underground	washing machine
water tank			swimming pool

2. *Secondary compound noun* refers to the head noun that is formed by verb with a morpheme –er or –ing as a noun. See Table 2.12.

Table 2.12 The Examples of Secondary Compound Nouns

noun + noun (er)	noun + noun (ing)
shoemaker	trainspotting
speechwriter	

Active and passive voices

English sentences can be formed by at least a subject and a verb. Also, verb in English can be written in ‘active voice’ and ‘passive voice’.

1. *Active voice* refers to the subject who has the agent role. In other words, a subject does the action to the passive or object (Vanderweide, Rees-Miller, Aronoff & O'Grady, 2010).

Table 2.13 The Examples of Active Voice

Subject [Thing doing action]	verb [verb]	Object [thing receiving action]
We	did	a lot of sightseeing in London last year.
I	have left	a message on your answer-phone.
Jim	will buy	a new car.

2. *Passive voice* refers to the grammatical subject expressing the theme or patient of the main verb – that is, the person or thing that undergoes the action or has its state changed (Vanderweide, Rees-Miller, Aronoff & O'Grady, 2010). The passive form is used to emphasize the importance of the nonhuman thing that receives the action. Moreover, the passive form is used when the writer does not know who is doing the action or does not want to mention who is doing the action.

Table 2.14 The Examples of Passive Voice

Subject [Thing receiving action]	verb [be + past participle of verb]	Object [by + thing doing action]
The tree	was pulled down	by someone.
The heart transplant	was carried out	by a surgeon.
Scotland's railway network	has been paralyzed	by the one-day strike.

Types of adjectives

Bailey (1975), in an empirical investigation of attributive adjectives in English, pointed out that descriptive adjectives have four subcategories:

1. *subjective/evaluative adjectives*: poor, sweet, pretty, old, nice (may also be used to measure or describe with other more literal meanings)

2. *shape adjectives*: round, square, oblong, irregular, triangular

3. *measurement adjectives*: big, little, tiny, large, short

4. *coloration adjectives*: red, blue, pale, dark

(cited in Celce-Murcia & Larsen-Freeman, 1999, pp.396-397)

In conclusion, the collocations of the target word in the present study were analyzed for grammatical features such as semantic functions of prepositions, compound nouns, active and passive voices, and types of adjectives.

2.4.4 Previous study of collocation research

The linguistic researchers conducted a corpus-base study which investigated word collocations. The linguists used one or more corpora to investigate words and analyze for the collocations in academic disciplines (Ackermann & Chen, 2013). In addition, the findings of words from more corpora were compared with each other in order to generalize the results (Nesselhauf, 2003). Consequently, it can be concluded from the previous research that the corpus-based studies were aimed to investigate and compare word collocations.

The researchers used one or more corpora to investigate aspects of lexis associated with the specific areas such as English novels, academic discipline, and food and beverage advertisements (Table 2.15).

Table 2.15 Previous studies of collocation research

Author	Field of investigation	Results of collocation
Intaraprasit (2013)	English novels	adjective + noun (33.74%) adverb + verb (27.34%) noun + noun (16.11)
Buakaew (2015)	Food and beverage advertisements	adjective + noun (45.90%) verb + preposition (44.06%) noun + noun (15.85%) preposition + noun (15.25%) noun + preposition (13.56%) verb + noun (6.01%) verb + adverb (5.46%) adjective + preposition (5.08%) verb + noun / pronoun (4.37%) adjective + to-infinitive (1.69%) adverb + adjective (1.09%)
Ackermann and Chen (2013)	Pearson International Corpus of Academic English (PICAЕ)	adjective + noun (71.8%) verb + noun (12.6%) adverb + past participle (5.0%) adverb + adjective (5.0%) noun + noun (2.5%) verb + adjective (1.2 %) verb + adverb (1.2%) adverb + verb (0.6%)

Table 2.15 showed the investigation of collocation in the three particular areas. It can be seen that these previous studies determined collocation by its patterns and parts of speech. For example, Intaraprasit (2013) compiled English novels to investigate collocation. He found that the adjective collocated with the noun was used mostly in the English novel. Moreover, Buakaew (2015) employed corpus-based study

to organize the type of collocations and their characteristics in food and beverage advertisements. The results indicated that collocations of verb combinations such as phrasal verbs were used mostly in the corpus. In addition, Ackermann and Chen (2013) gathered the academic words in order to find academic collocations used in research articles, and the results showed that the 2,468 most frequent collocations found were relevant to EAP teaching pedagogy.

In addition, corpus-base studies were used to investigate the plausibility of applying a common list of frequent collocations compared to different academic disciplines. For example, Durrant (2014) and Hyland (2008) examined the use of lexical bundles across a field of academic disciplines, and found that there were significant differences in their use. In addition, corpora were used to compare English academic writing of non-native speakers (NNS) to native speaker students. For instance, Nesselhauf (2003, 2005) found that English essays written by German undergraduates and native speakers were differently significant in the use of collocations. The results showed that learners' first language influenced the collocation used when they translated their first language into English. Granger (1998) also investigated the use of collocations and prefabricated expressions in NNS writing. The results showed that the NNS writers usually "cling onto certain fixed phrases and expressions".

In summary, previous studies on collocation were conducted to investigate and compare the use of collocations in various fields. Words and collocations were examined through one or more corpora. Moreover, corpus-based studies were employed to measure the use of English between non-native speakers (NNS) and native speaker students. As the purposes of the present study were to investigate words and usages in a specialized field of laboratory animals, the specialized word in that corpus was investigated for the usage of collocation.

Chapter Summary

In summary, EST courses have its own unique rhetorical discourse and communicative functions that are used in the discipline, so learners and researchers need to improve their English in order to achieve knowledge and communication requirements in scientific disciplines. Since the present study focuses on the field of laboratory animals which was in the English for Science and Technology (EST) area, a corpus-based study was conducted to collect the words and usages in the field. Words in the corpus comprised of General Service List (GSL), Academic Word List (AWL), and Outside Word List (OWL). The emphasis was on OWL because it referred to the technical words that appeared in a particular discipline. Moreover, knowledge of collocations should be provided to assist learners to form sentences, effectively. As a researcher, it is also important to know communicative functions, especially purposive function, because purposive sentences are required to be used in all standard journals. Therefore, the researcher needs to state the purposes of the study when writing article. Briefly, the knowledge of vocabulary, collocation, and communicative functions can facilitate learners and researchers to read and write research articles.

CHAPTER III

RESEARCH METHODOLOGY

This section is provided to overview the research methodology of the present study. The purposes of the present corpus-based study are to investigate content words, collocations, and communicative functions. To achieve the research purposes, the methodology section is organized systematically as follows (Figure 3.1):

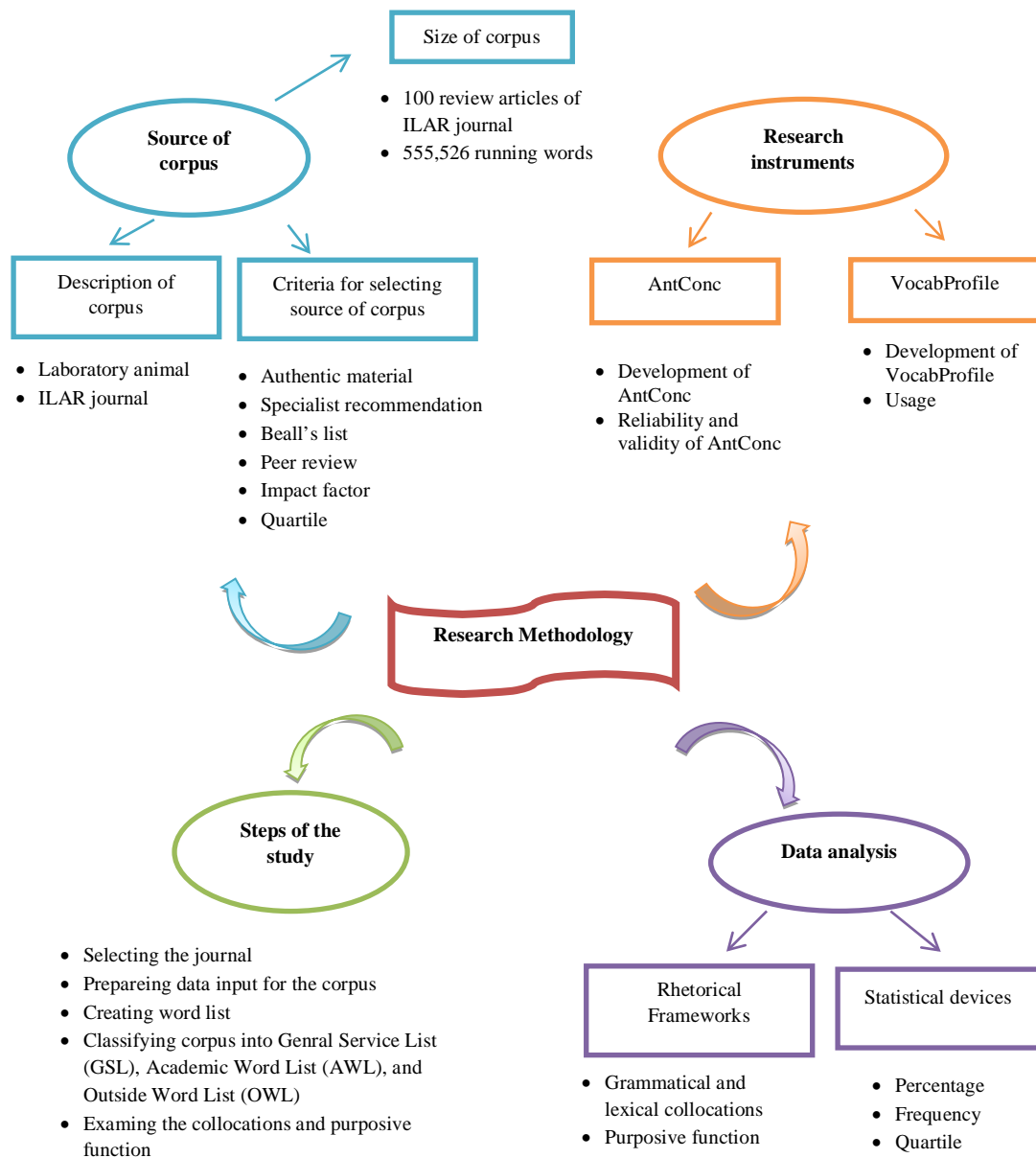


Figure 3.1 The overview of methodology

3.1 Source of Corpus

In order to investigate words and usage, data was collected from the quality journals for the present study. The description of corpora was introduced, and then, the source of corpus was validated for conducting the research. Moreover, size of corpus included words in GSL, AWL, and OWL from laboratory animal review articles. Description of corpus, source of corpus, and size of corpus were explained in details.

3.1.1 Description of Corpus

The input data of this present corpus came from the Institute for Laboratory Animal Research journal (ILAR) which was described as the laboratory animal corpus. The laboratory animal corpus comprised of the content words which were divided by VocabProfile program into three word lists: 1) the General Service List (GSL), 2) the Academic Word List (AWL), and 3) the Outside Word List (OWL). The corpus was collected from 100 articles of ILAR journal between 2012 and 2014. The ILAR journal was described as a peer reviewed journal published by Oxford University Press. It publishes four issues per year, and provides articles in the form of both paper and electronic data. The journal focused on the research in the field of laboratory animals and original articles which promoted the use of animals and human care. Furthermore, the ILAR journal had review articles providing a summary of previous laboratory animal research. Briefly, the principle corpus of this present study contained the linguistic data which was relevant to the field of laboratory animals.

3.1.2 Source of Corpus

The source of this corpus-based study was the Institute of Laboratory Animal Research Articles (ILAR), and it was selected based on six criteria as following.

Authentic material

First, source of corpus was collected from review articles because they were authentic materials. According to Underwood (1989), authentic materials gave students a proper representation of real and natural language used in real life. As the

language was changed and adapted to the needs of users all the time, authentic materials can introduce new vocabulary and new patterns or structures of language which students were likely to encounter in the real communications (Breen, 1995; Otte, 2006). Thus, authentic material can enrich students' language learning experiences and their use of English, and it can sensitize students to the target language use in real world situations (Harmer, 1994; Wong, 1995). For these reasons, the review articles, which were authentic material, were selected as the source of corpus in order to collect the linguistic data for this study.

Specialist recommendation

Second, the source of corpus should be well-known and acceptable in the specialized field. Therefore, the source of the present corpus was suggested by an expert in the related field. To select the renowned journal in the field of laboratory animals, the researcher interviewed the director of Laboratory Animal Center at Mahidol University, Thailand, who suggested that the ILAR journal was internationally well-known and reliable.

Beall's list

A third criterion was Beall's list. To ensure the reliability of the journal, the journal must not appear or belong to Beall's list which is the list of suspicious journals or predatory journals which do not meet the journal quality. Therefore, ILAR journal was searched in Beall's list which did not show the journal in the list (Figure 3.2).

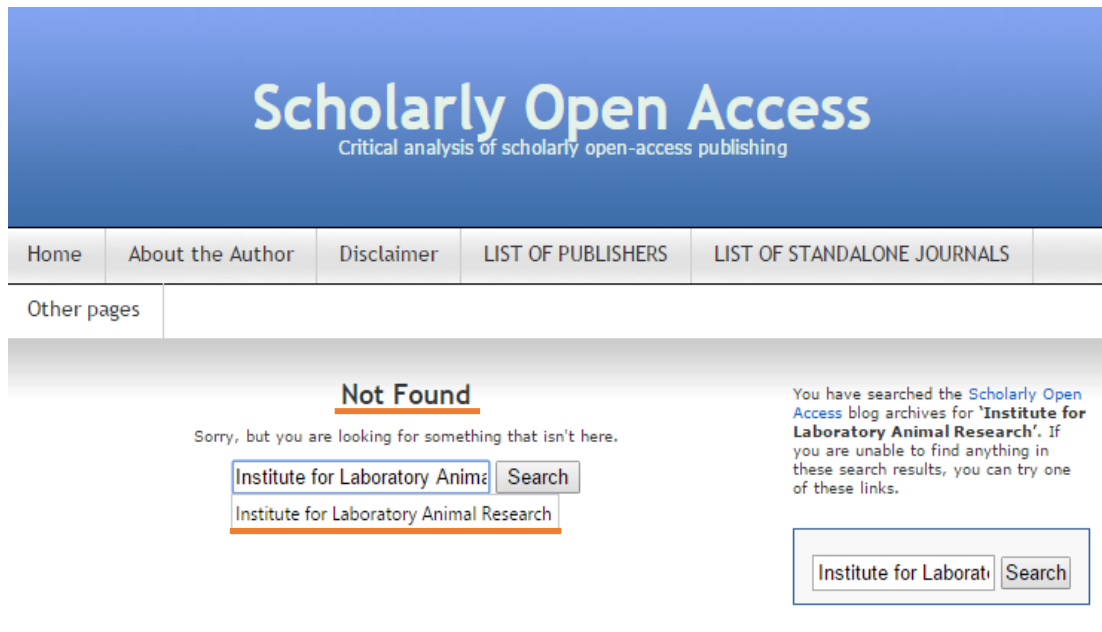


Figure 3.2 The searching result of the 'Institute for Laboratory Animal Research' journal from Beall's list

Peer review

The journal needs to be validated by a peer-review process. In academic discipline, peer review process is often employed to scrutinize an academic paper for correctness of methodology, content, and language before publication. As described in the official ILAR website, this journal was peer-reviewed by professional reviewers in the field of laboratory animals.

Impact factor

The impact factor of ILAR journal was validated by Journal Citation Reports (JCR), which provided an evaluation of the world's leading journals by presenting statistical information based on citation data. Articles' citations or references can help to measure research impact in respective category. ILAR's impact factor over a 5-year period in the fields of Animal Science and Zoology, Biochemistry, and Genetics and Molecular Biology was acceptable at 2.393. Consequently, the journal of ILAR was sufficiently credible and acceptable for use.

Quartile

The quartile of journal rank was based on SCImago which has four key values (Q1, Q2, Q3, and Q4). Q1 is the highest rank, while Q4 represents the lowest rank. According to Journal Citation Reports between 2010 and 2014, which is the target publication of the present study, *ILAR J* was evaluated with high values (Q1 and Q2) (Figure 3.3).

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 ILAR journal ranking

Thus, the source of the present corpus was selected based on six criteria including authentic material, the specialist recommendation, Beall’s list, peer review, impact factor, and quartile. The ILAR journal met all the mentioned criteria; therefore, this journal was used as the input data of the current linguistic corpus.

3.1.3 Size of Corpus

The data collection investigated in the particular discipline can be controlled by its size and its publication years (Teubert, 1996; Stubbs, 2001). The size of the present corpus-based study contained 555,526 running words from 100 review articles of ILAR journal. The 100 articles were published between 2010 and 2014 because a range of five years can show varying trends, and it was not too outdated for use.

Compared to previous corpus-based studies, the size of corpus in this study was not too large. For example, Swales (1981) collected 48 articles in the fields of natural and social sciences in order to analyze the introduction section. Samraj (2005)

used 24 abstracts of Conservation Biology and Wildlife Behavior in her exploration of genre on research article introductions and abstracts. Pho (2008) studied the macro-organization of 30 abstracts in applied linguistics and educational technology. Tribble and Jones (1990) suggested that size of corpus should be adequate to generalize the results; therefore, the words investigated in the corpus should be at least 50,000 of running words. The previous corpus-based study of psychology technical words (Kaewphanngam, 2002) investigated 180,000 running words from 60 psychology articles in three psychology textbooks. However, in order to make the study more reliable, the present study increased the number of samples by collecting 100 review articles in the field of laboratory animals which covered 555,526 running words.

To prevent bias from randomizing the articles, the stratified random sampling technique was used to determine the number of selected articles in each year because the number of articles in each publication year was not equal (Table 3.1). All random articles were review articles that analyzed and discussed about previous research in laboratory animals.

Table 3.1 Size of Corpus

Publication years	Number of population (N)	Number of samples (n)	Percentage (%)
2014	38	24	15.00
2013	19	12	7.50
2012	37	23	14.38
2011	32	20	12.50
2010	34	21	13.12
Total	160	100	62.50

In summary, the corpus was explained by its description, source, , and size. This laboratory animal corpus was compiled from 100 review articles between 2010 and 2014 in the field of laboratory animals. The articles were collected from the ILAR journal which was selected based on six criteria: authentic material, the specialist recommendation, Beall's list, peer review, impact factor, and quartile. Thereafter, the 100 articles were chosen by stratified random sampling.

3.2 Research Instruments

The research instruments used to collect and analyze the data in the present study were AntConc and VocabProfile. AntConc and VocabProfile were software programs that were used to process only the electronic data. As the data gathered in this corpus-based study was downloaded from the website of ILAR journal, the data was available in electronic form.

3.2.1 AntConc

The study used AntConc version 3.4.3 to create laboratory animal corpus. AntConc program was applied to collect the electronic texts from ILAR journal in order to make a word list and present the linguistic data of the laboratory animals.

3.2.1.1 Development of AntConc

The AntConc program (3.4.3) was the concordance program developed by Laurence Anthony of Waseda University in Tokyo, Japan. The program was used to count the running words and make a word list. AntConc can show and arrange the word frequencies in order. Moreover, it was used to collocate right and left of the node words for analyzing collocations and communicative function.

3.2.1.2 Reliability and Validity of AntConc

The linguists applied AntConc program to analyze linguistic data such as words and sentences because the program was available and easy to use (Diniz, 2005). Froehlich (2015) mentioned that Antconce can show the frequency, structure, and collocation of the target word or sentence which needed to be analyzed. Furthermore, Tribble (2012) compared the software programs used to analyze corpora (Figure 3.3).

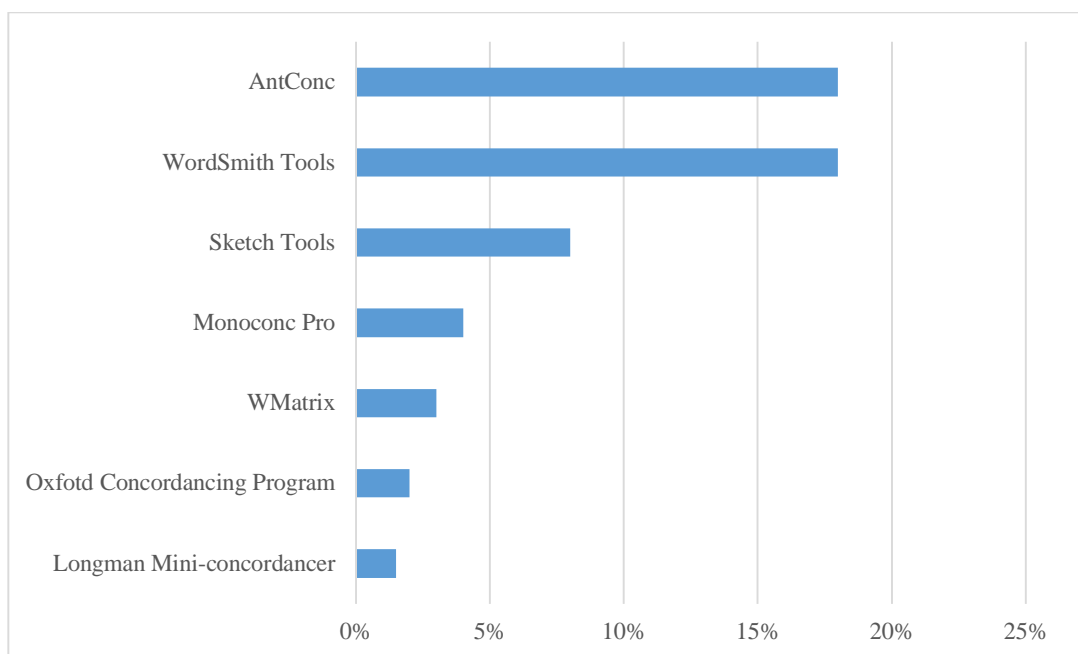


Figure 3.4 Comparison of software programs used to analyze corpora (cited in Anthony, 2013)

Figure 3.4 showed that AntConc and WordSmith Tools were the most popular programs. WordSmith Tools is a licensed software that available for installation only by diskette, after the user buys the program. Moreover, WordSmith Tools restricts the number of users by a series number while installing the program. Once installed on one computer, WordSmith Tools cannot be used on other computers. However, AntConc is a freeware that is available by online download. Importantly, AntConc allows the user to use own data in order to create own corpus. As the purpose of the present study aim to create the word list of laboratory animals compiled from ILAR journal, AntConc can be used to create the corpus of laboratory animals. For this reason, AntConc was found suitable for use for this corpus-based study.

3.2.2 VocabProfile

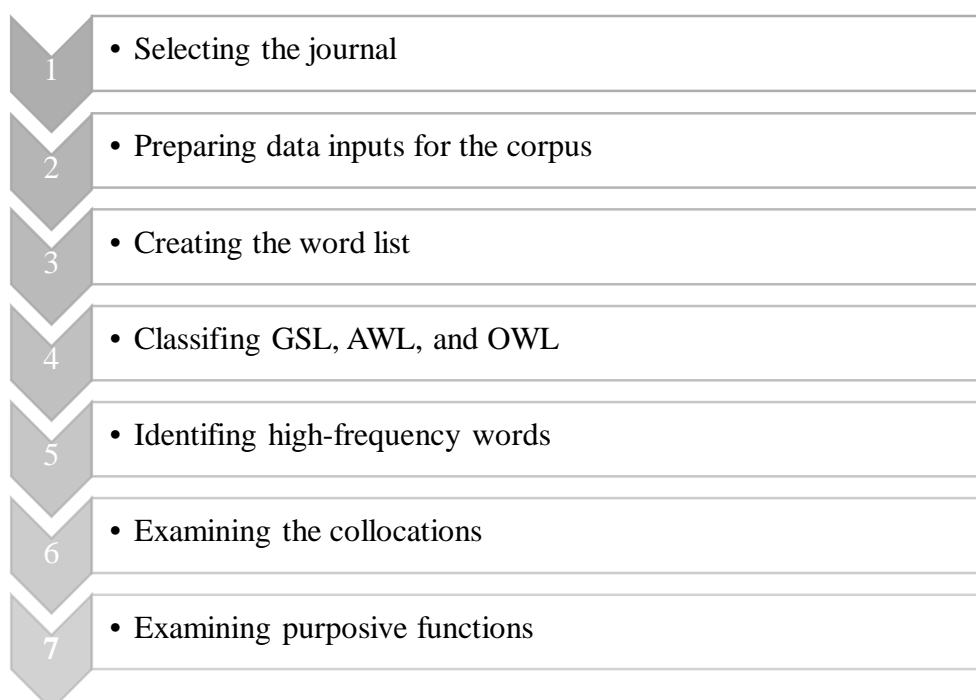
VocabProfile was created by Tom Cobb in 2003 and used to categorize the running words into GSL, AWL, and OWL. The program was applied because the database of the program was developed and updated from the new findings of GSL and AWL (Browne, Culligan & Phillips, 2013). Laufer and Nation (1995) validated the VocabProfile program and found that the program benefited students in term of vocabulary learning because the database of the program was suitable for all students'

lexical knowledge. Thus, the program divided vocabulary into levels, so students can learn vocabulary from the beginning to advanced level. In addition, the database of VocabProfile covered vocabulary in AWL that was needed by students (Cobb & Horst, 2001).

In short, VocabProfile and AntConc (3.4.4) were used as research instruments. Words in the study of laboratory animals were collected by AntConc, which generated the word list. The word list was categorized by VocabProfile into GSL, AWL, and OWL. Lastly, collocations and communicative functions were analyzed by using concordance lines.

3.3 Steps of the Study

The procedural steps of the study were planned to fulfill research objectives by seeking the answer for the research questions. Consequently, the research methodology design had the seven following steps:

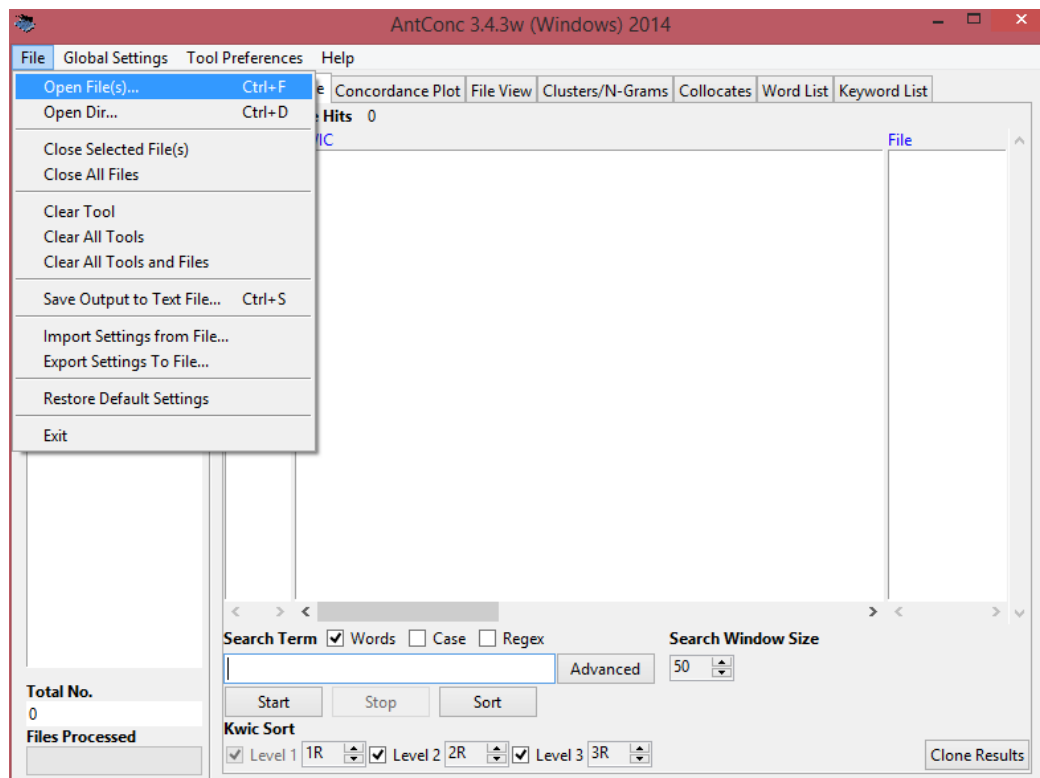


First, the ILAR journal was validated and chosen as suggested after interviewing the director of National Laboratory Animal Center. Then the journal was validated under six criteria (authentic material, specialist, Beall's list, peer review,

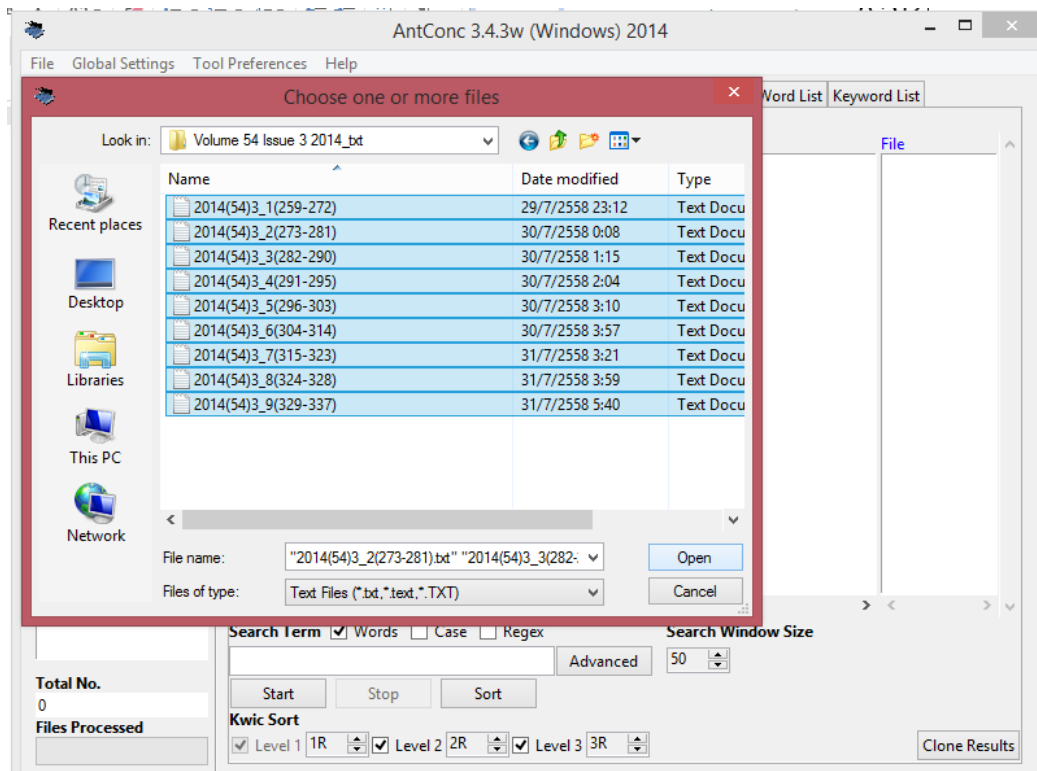
impact factor, and quartile). Thereafter, the 100 articles between 2010 and 2014 were randomized.

Second, the 100 selected articles were downloaded and prepared as the data input for the corpus study. After download, each article was saved as text file (.txt) because AntConc program can process only text files. Only introduction, literature review, methodology, result, and discussion sections were included. The labels and pictures were removed from the articles.

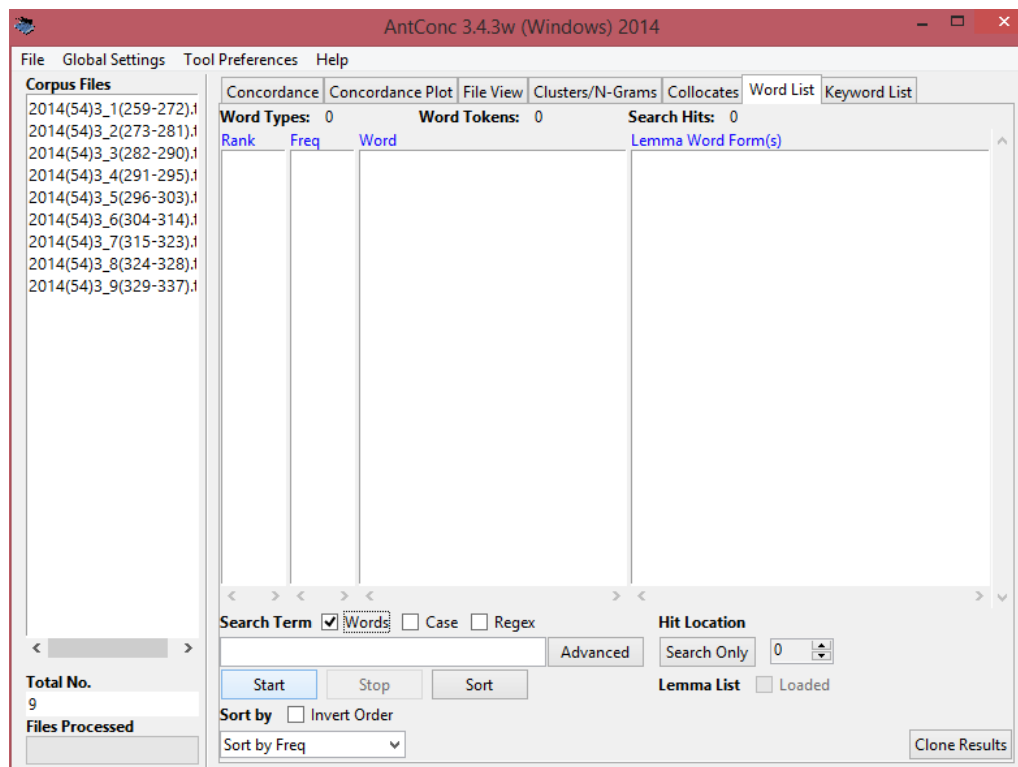
Third, a word list was created. All text files were processed by AntConc program to make a word list and identify the frequency of each word from high to low (see Figure 3.5).



Click 'File' > 'Open File(s)'



Select files > click 'Open'



Select program 'Word List' > click 'Start'
Figure 3.5 Creating a word list

After creating a word list of laboratory animals, the list contained both content and function words. The present study focused on only content words, and therefore, all function words were deleted from the list. The function words that were deleted were adverbial particles, modal verbs, prepositions, conjunctions, determiners, pronoun, and numbers (Nation, 2001). Also, the words such as proper nouns, personal names and non-words were removed manually from the list.

Fourth, General Service List (GSL), Academic Word List (AWL), and Outside Word List (OWL) were grouped by VocabProfile program. The content word list of laboratory animals was put into VocabProfile so as to divide the words into GSL, AWL, and OWL. See Figure 3.6.

The screenshot displays the VocabProfile web interface. At the top, it shows the navigation path: [home](#) > [VocabProfilers](#) > **Compleat** (CLASSIC; NGSL; BNL; BNC;+COCA; +Core; French). The main heading is **Compleat Web VP!** with the subtext "2 languages, 6 English list frameworks - on 1 interface for clear comparisons". A note states: "Note that BNL, Coca-Core, and Classic AWL are not full 1000-family lists; and that NGSL and French are Lemmas not Families". Below this, there are links for "How to make specific list framework comparisons? See Demo 8 [here](#)" and "Lex Frequency predicts Text Complexity? Check [these](#)".

The interface is divided into two main input methods:

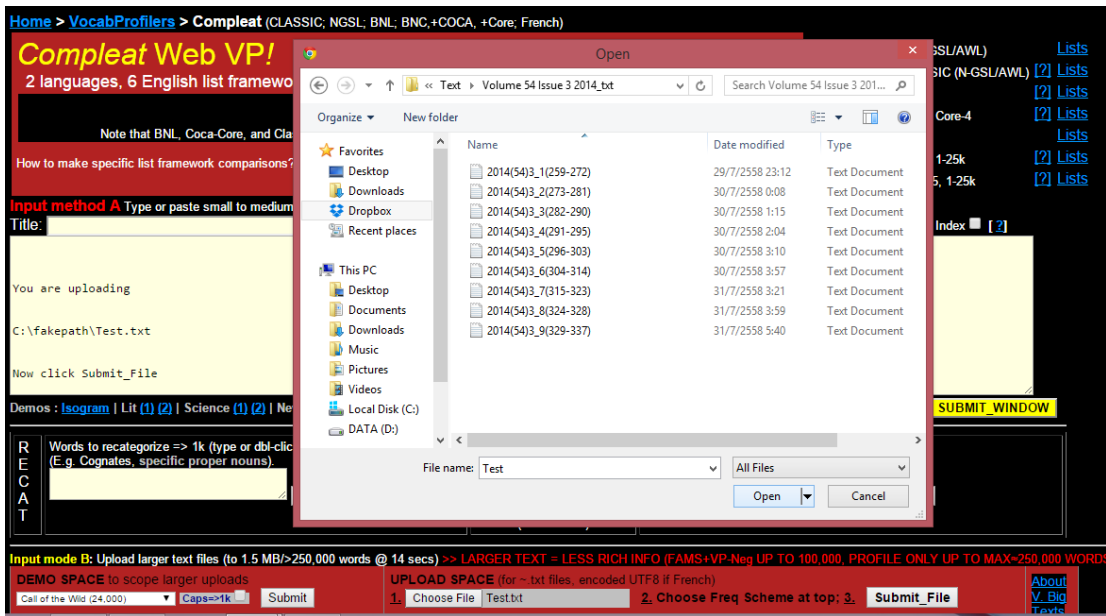
- Input method A:** Type or paste small to medium size text (350,000 chars - about 60,000 words - max) and click *Submit_window* for Frequency Profile. It includes a text input field, a "Submit Window" button, and checkboxes for "Eng:Fri Cognates", "Edit-to-a-Profile", "Sentence Count", "Bar Chan", and "Count Index".
- Input mode B:** Upload larger text files (to 1.5 MB/>250,000 words @ 14 secs) >> LARGER TEXT = LESS RICH INFO (FAMS+VP-Neq UP TO 100,000, PROFILE ONLY UP TO MAX=250,000 WORD). It includes an "Upload Space" section with a "Choose File" button and a "Submit File" button.

At the bottom, there are several configuration options for processing:

- RECAT:** Words to recategorize => 1k (type or dbl-click) (E.g. Cognates, specific proper nouns). Includes a "Select" button.
- ALL mid-sentence capped words:** Options to "=> 1-k" or "=> ignore" (new Jan '14).
- Plus stray proper at sentence boundary => 1k (E.g. " Paul Martin"):** Includes a "Select" button.

The bottom navigation bar includes links for "Demos", "Lit", "Science", "News", "Speech", "Adults Kids", "Rex M.", "AWL", "LEGAL", "New! French", "Highlight", "Count", "No returns", and "SUBMIT_WINDOW".

Select 'NEO CLASSIC (N GSL/AWL)'



Click 'Choose Files' > select file > click 'Open' > click 'Submit_Files'

Figure 3.6 Categorizing word list

Fifth, high-frequency words were identified. The words from OWL were ordered according to the frequency from high to low so as to identify high-frequency words by using quartile.

The sixth step was identifying high-frequency word. The highest-frequency word in OWL was examined for its collocation lines by using AntConc program (Figure 3.6). Then, all concordance lines of the highest-frequency word were copied and put into Microsoft Excel. Each concordance line was analyzed and grouped into patterns following grammatical and lexical collocations (Benson, 1985). In this step, the concordance lines of each collocation pattern were analyzed for grammatical feature. For example, the pattern of noun collocated with a preposition was analyzed to find the semantic function of the preposition. The pattern of noun collocated with a noun was plausible to generate a compound noun. The pattern of noun collocated with a verb was searched for active and passive voices. Finally, the pattern of adjective collocated with a noun was analyzed for the kind of descriptive adjective.

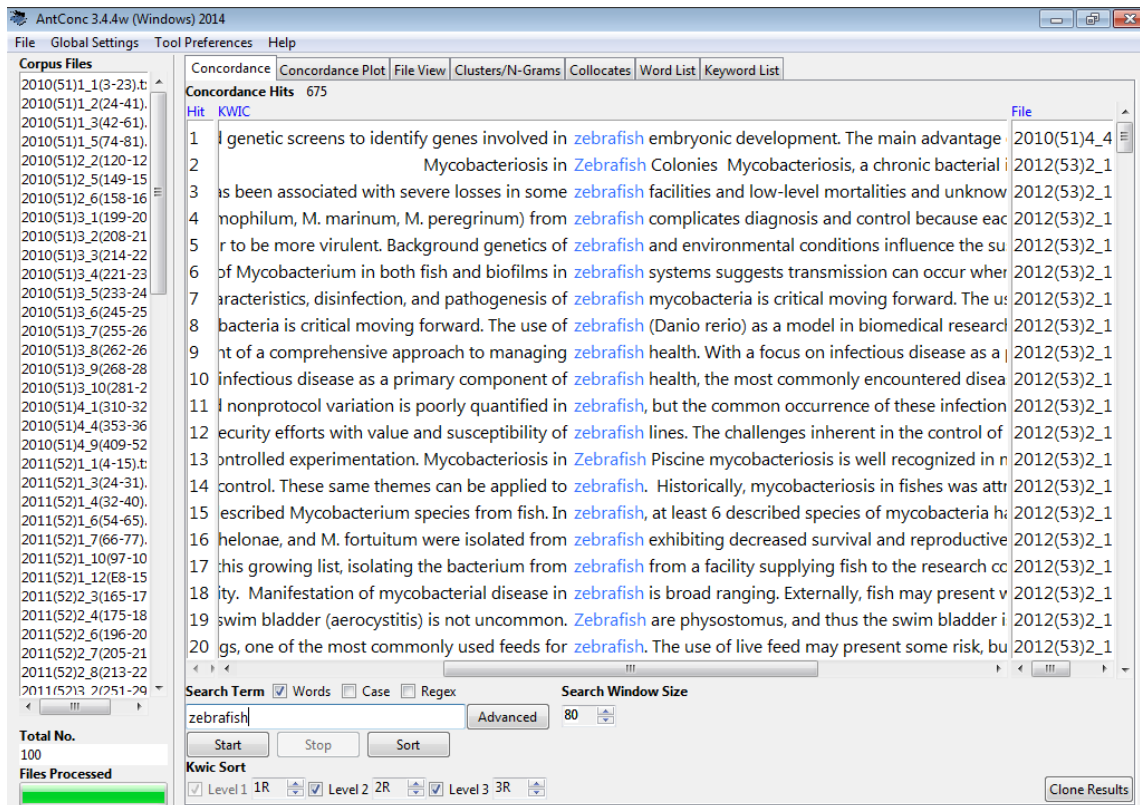


Figure 3.7 The concordance lines of the highest-frequency word

The last step was analyzing the purposive function. Concordance plot was used to find the article that the highest-frequency word belonged to (Figure 3.8). If the sentences of highest-frequency word state the purpose of article, the researcher keeps those sentences to analyze the patterns of purposive statement. The purposive sentences were grouped by patterns following Swale and Feak (2012) and Soranastaporn (2013).

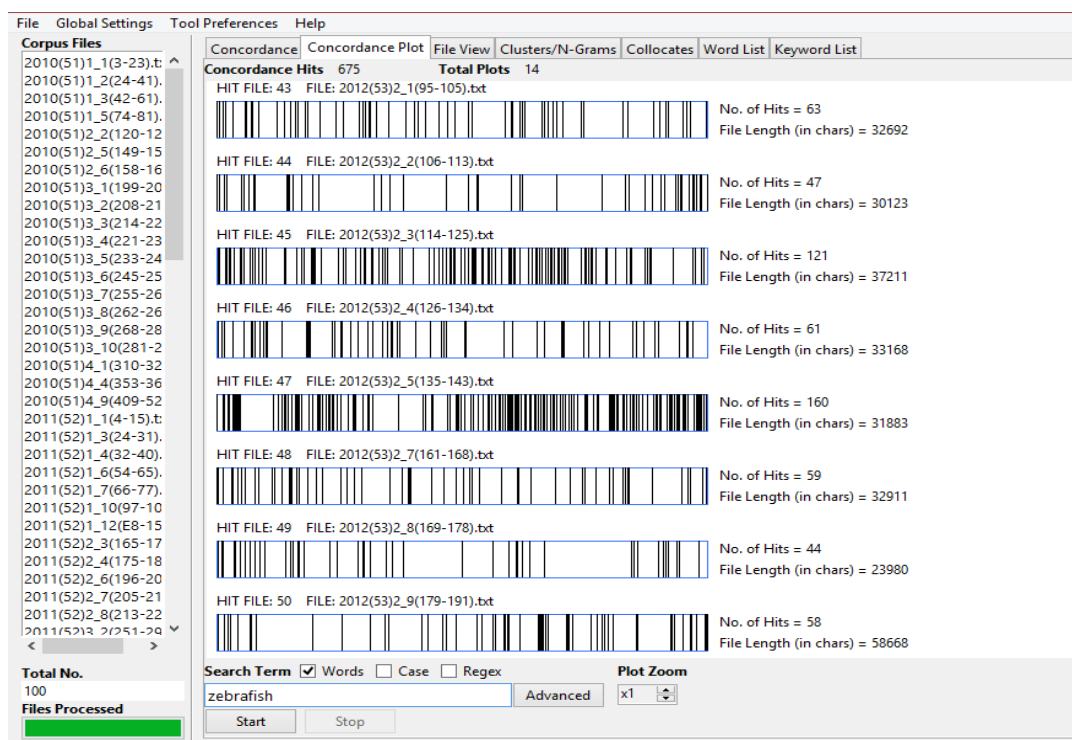


Figure 3.8 The concordance plot of the highest-frequency word

Thus, these seven steps were considered as the steps of the present study. The overall procedure began with preparing data for the corpus. Then the word list was created and divided into GAL, AWL, and OWL. Only OWL was examined for high-frequency words, collocations, and communicative functions.

3.4 Data Analysis

For the present study, the data analysis employed the framework of collocations and communicative functions. Thereafter, all analyzed data were presented in term of statistics. Thus, data analysis consisted of 1) rhetorical frameworks and 2) statistical devices.

3.4.1 Rhetorical frameworks

Two frameworks were used for data analysis: collocation framework and communicative functions.

3.4.1.1 Collocations

The collocation frameworks for this study were divided into two types: grammatical collocations and lexical collocations (Benson, 1985).

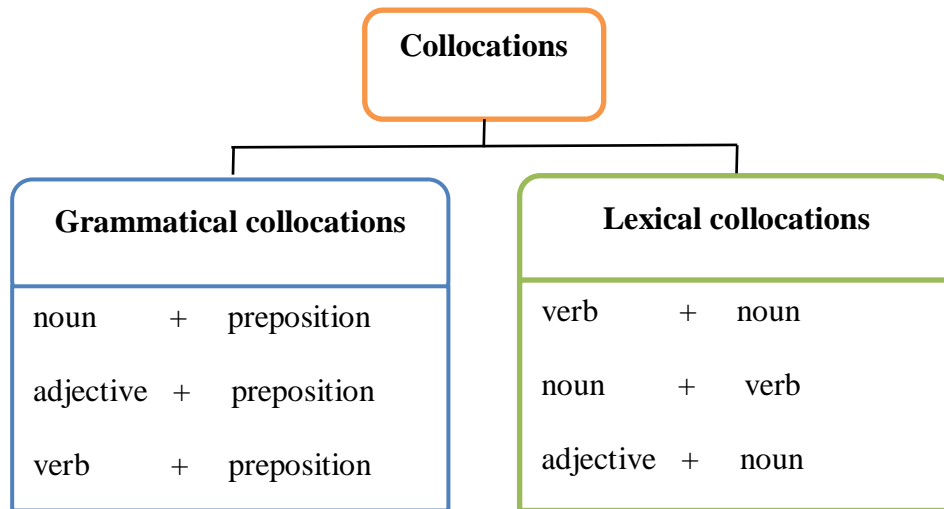


Figure 3.9 Grammatical and lexical collocations

Grammatical collocations were a noun, verb, or adjective followed by a preposition (Benson, 1985).

Table 3.2 Examples of Grammatical Collocations

Combination of collocations	Examples
noun + preposition	blockade against, apathy towards
adjective + preposition	be angry at, be fond of
verb + preposition	abide by, access to

On the other hand, *lexical collocations* were the two equal words combined together such as noun collocated with verb, verb collocated with noun, and adjective collocated with noun (Benson, 1985).

Table 3.3 Examples of Lexical Collocations

Combination of collocations			Examples
verb	+	noun	compose music, wind a watch, spend money
noun	+	verb	ball bounces, bee buzz, bombs explode
adjective	+	noun	strong tea, a sweeping generalization, cheerful expression

The present study used grammatical and lexical collocations (Benson, 1985) because it focused on words and language used in laboratory animal research which was academic writing. Hence, collocations such as phrasal verbs, idioms, quotations, or proverb rarely occurred in academic writing especially scientific research articles. Moreover, Benson (1985) identified the pattern of grammatical and lexical collocations by part of speech which can be beneficial for writing. The researcher can use the patterns of grammatical and lexical collocations to analyze their grammatical structure in terms of semantics of preposition, compound noun, voices, and descriptive adjective.

3.4.1.2 Purposive function

In addition, the patterns of the purposive function used to analyze purposive statements of the highest-frequency words adopt the framework of Swales and Feak (2012) and Soranastaporn (2013). The present study investigated communicative function in terms of a purposive function, i.e. the statement that showed the objective of the study, because the purpose statements are the compulsory requirement of all research articles. The purposive statement contains keyword that tells reader the central idea of the present study. Moreover, purposive statements contain specific words and patterns to represent the purpose of the study. Consequently, words and patterns that match with the framework (Swales & Feak, 2012; Soranastaporn, 2013) were considered as a key factor to form a purposive statement.

3.4.2 Statistical devices

The statistical devices used to analyze the results of the data were percentage, frequency, and quartile. The text coverage of word lists (GSL, AWL, and

OWL) and collocations were reported in percentages. In order to report the occurrences of words, frequencies were used to rank words from the highest to the lowest. Lastly, quartile was used to identify the high-frequency words in OWL. Words in the OWL were divided into quartile by their ranks. Then words in the first quartile were identified as high-frequency words.

Briefly, to analyze the data, the study adopted rhetorical frameworks and statistical devices. Collocations were analyzed following grammatical and lexical collocations (Bahns, 1993; Benson, 1985; & Biskup, 1992). In addition, the writing textbooks (Swales & Feak, 2012; Soranastaporn, 2013) were used as a framework to investigate purposive functions. Finally, the results of words and collocations were reported in percentage and frequency while high-frequency words were determined by quartile.

Chapter Summary

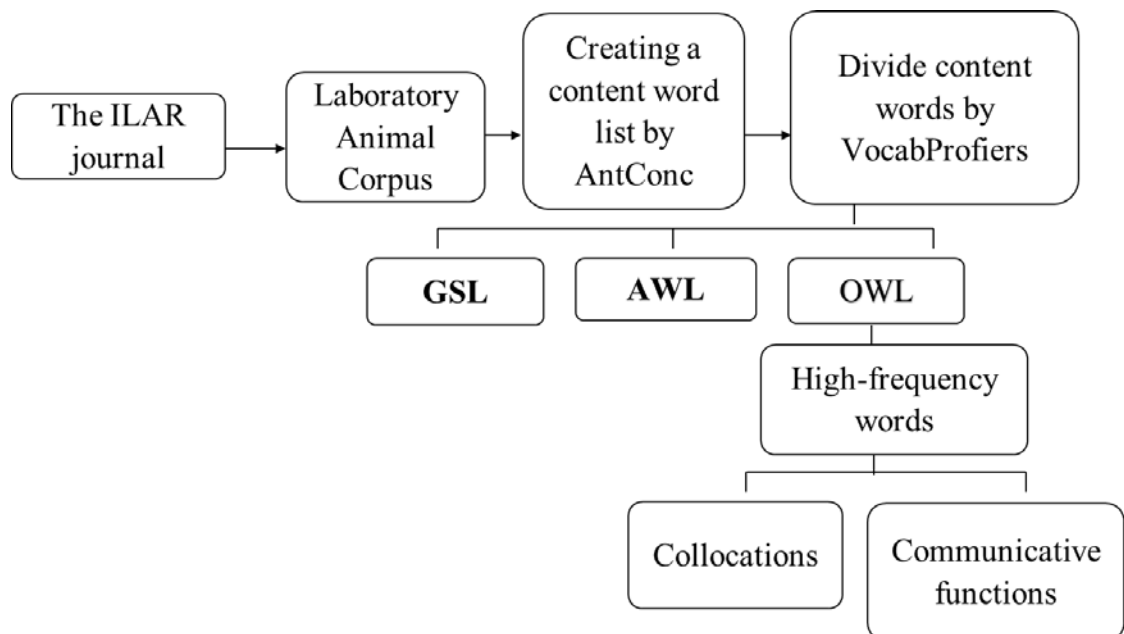


Figure 3.10 Summary of research methodology

In conclusion, this research methodology explained the steps of data collection and data analysis of the study. To compile a laboratory animal corpus, 100

review articles were selected by stratified random sampling technique. Then AntCon was used to create the content word list of laboratory animals. VocabProfile was applied to group the content words in the word list into General Service List (GSL), Academic Word List (AWL), and Outside Word List (OWL). The focus of the study was the words in OWL because it contained the specialized or technical words in the field of laboratory animals. Therefore, OWL was identified for high-frequency word by quartile. Finally, the highest-frequency word in OWL was analyzed for collocations (Benson, 1985) and purposive function (Swales & Feak, 2012; Soranastaporn, 2013).

CHAPTER IV

RESULTS

The present corpus study aims to investigate words, collocations, and communicative functions used in the review articles of laboratory animals. To achieve the research purposes, this section presents the results of the study which answered the four research questions (Figure 4.1).

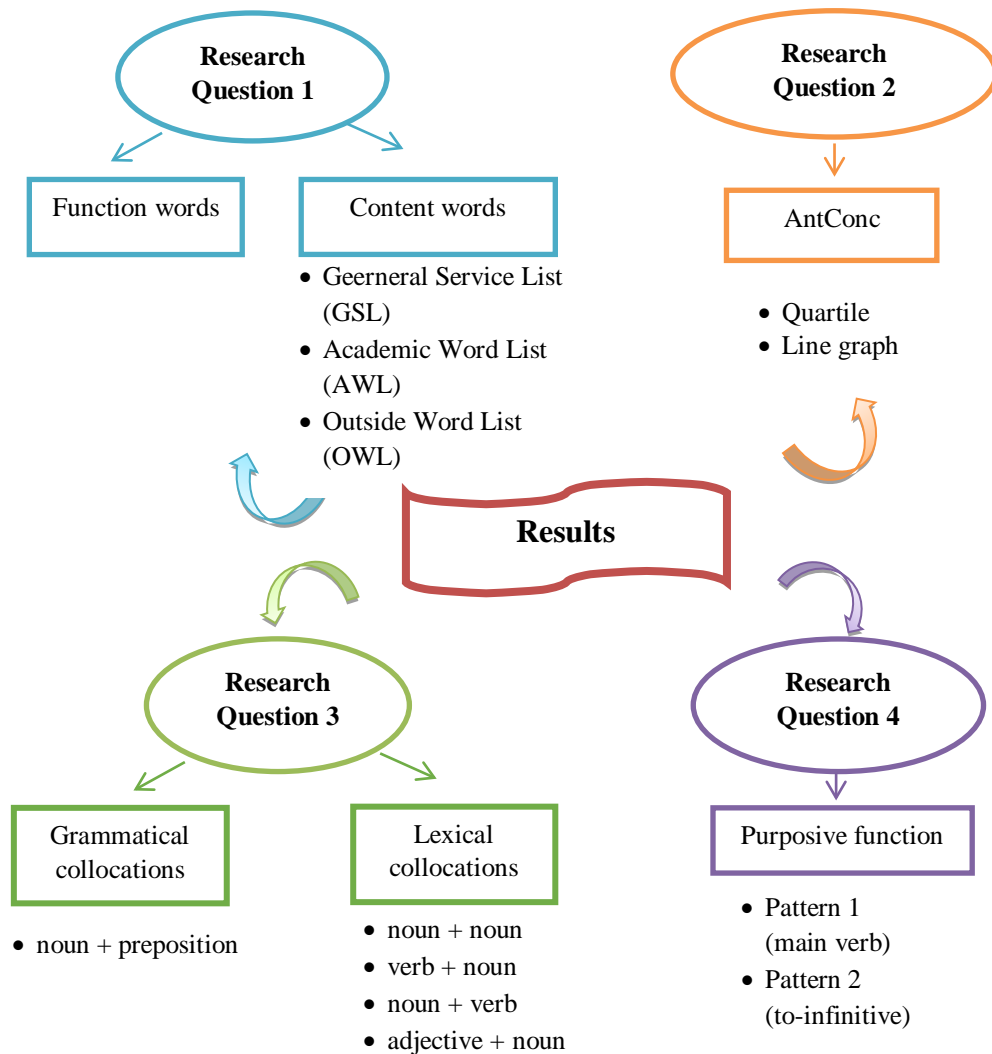


Figure 4.1 The overview of result chapter

4.1 The Content Words

Research question 1: What were the content words of GSL, AWL, and OWL used in the laboratory animal review articles?

The first question aimed to find content words that occurred in the field of laboratory animals. To answer this question, the source of corpus from ILAR journal was identified and compiled to make the laboratory animal’s word list. The framework of GSL and AWL revised by Browne, Culligan, and Phillips (2013) was used as reference word lists to identify GSL, AWL, and OWL of the laboratory animals.

A hundred review articles from the ILAR journal were selected by sample random sampling, and all of the random review articles were review articles. All the 100 random articles were processed by AntConc which showed that there were 555,526 word tokens and 21,609 word types. See Figure 4.2

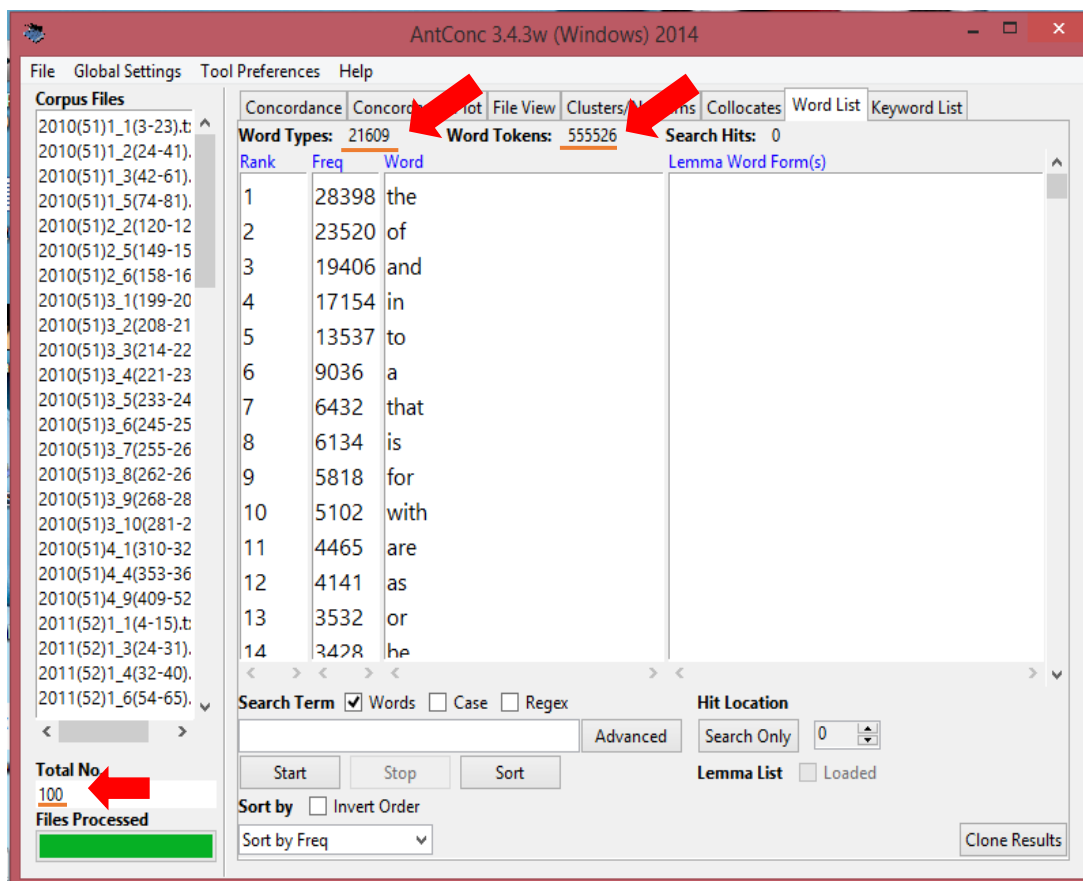


Figure 4.2 The total of word types and word tokens

The number of word types was 21,609 which were both function words and content words. As showed in Figure 4.3, content words in the corpus accounted for 12,220 (56.55%). The others, which accounted for 9,389 (43.45%), were function words, proper names, and abbreviations. As the the present study aimed to find only content words, all function words were removed from the word list and the present corpus.

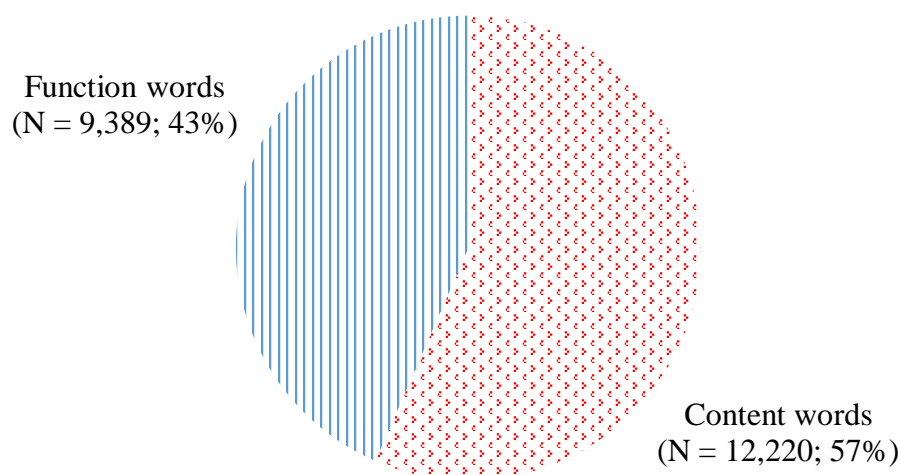


Figure 4.3 The proportion of content words and function words ($N_{\text{total}} = 21,609$)

Table 4.1 showed the results of the content words occurred in Laboratory Animal review articles. After creating the content word list, the list was processed by the VocabProfile program in order to identify GSL, AWL, and OWL. Thus, the results showed that the most occurrences were OWL, GSL, and AWL respectively.

Table 4.1 The Coverage of Word Lists

Word Types	Tokens	Percentage (%)	Cumul. Token (%)
GSL 1	2,141	11.54	11.54
GSL 2	1,775	9.57	21.11
GSL 3	1,167	6.29	27.40
AWL	1,248	6.73	34.13
OWL	12,220	65.87	99.99
Total	18,551	100	100.00

GSL occurred 27.40% of the whole corpus which had three levels. GSL 1 was the first 1000 highest-frequency word; while, GSL 2 was the second 1000 high-frequency word, and GSL 3 was the third highest occurrences of the remaining of 818 words in GSL (Browne, Culligan, and Phillips, 2013). According to all GSL levels, GSL 1 (11.54%) appeared to occur the most which was nearly twice as much as GSL 3 (6.29%). GSL 2 (9.57%) occurred almost as often as GSL 1. The findings of GSL 1, 2, and 3 were presented in Table 4.1, 4.2, and 4.3 below.

Table 4.2 The 100 Most Frequent Words in GSL 1

Rank	<i>f</i>	Words	Rank	<i>f</i>	Words
1	1838	studies	51	421	related
2	1695	animal	52	416	several
3	1649	animals	53	416	social
4	1487	human	54	415	potential
5	1433	research	55	413	age
6	1024	disease	56	413	differences
7	1018	models	57	393	drug
8	1001	model	58	384	patients
9	975	cells	59	381	include
10	952	used	60	372	long
11	946	study	61	372	years
12	919	humans	62	368	common
13	884	effects	63	367	diseases
14	825	use	64	366	effect
15	822	development	65	364	same
16	787	cell	66	359	likely
17	782	dogs	67	357	male
18	781	associated	68	349	less
19	662	levels	69	347	compared
20	658	water	70	338	female
21	644	health	71	333	shown
22	614	increased	72	331	evidence
23	613	specific	73	329	responses
24	604	using	74	328	activity
25	580	behavior	75	323	showed
26	577	risk	76	322	body
27	573	including	77	317	known
28	567	time	78	316	systems
29	548	response	79	312	provide
30	547	no	80	311	higher
31	538	high	81	311	role
32	532	important	82	308	days
33	524	changes	83	307	food
34	524	different	84	306	large
35	516	system	85	296	recent
36	514	fish	86	294	life
37	507	treatment	87	293	function
38	505	number	88	291	conditions
39	499	factors	89	290	available
40	499	results	90	290	level
41	482	data	91	290	possible
42	478	example	92	288	increase
43	476	reported	93	286	groups
44	475	early	94	285	cases
45	472	based	95	285	females
46	472	significant	96	285	small
47	470	group	97	279	care
48	465	new	98	279	low
49	455	control	99	278	result
50	452	similar	100	278	type

(*N* = 2,141)

Table 4.3 The 100 Most Frequent Words in GSL 2

Rank	f	Words	Rank	f	Words
1	730	expression	51	155	relevant
2	728	gene	52	154	scientific
3	683	mice	53	152	regulation
4	612	stress	54	148	outcomes
5	598	genes	55	145	importance
6	563	cancer	56	142	derived
7	552	brain	57	139	assessment
8	403	pain	58	139	injury
9	386	environmental	59	138	assess
10	339	mouse	60	138	context
11	336	colleagues	61	138	negative
12	319	observed	62	137	sequencing
13	315	mechanisms	63	137	trial
14	305	bone	64	136	sexual
15	302	administration	65	135	appropriate
16	292	complex	66	135	association
17	261	demonstrated	67	132	physical
18	248	sex	68	129	global
19	247	blood	69	128	phase
20	246	protein	70	126	combination
21	243	multiple	71	126	medical
22	234	significantly	72	123	dish
23	233	effective	73	123	status
24	233	exposed	74	121	defeat
25	222	decreased	75	119	cancers
26	222	experiments	76	119	initial
27	220	researchers	77	118	currently
28	217	trials	78	118	decrease
29	212	diet	79	117	directly
30	210	wild	80	115	contribute
31	206	novel	81	114	reduction
32	202	presence	82	112	techniques
33	199	proteins	83	111	essential
34	198	highly	84	111	outcome
35	189	approximately	85	110	concentrations
36	188	critical	86	110	severe
37	182	additional	87	107	components
38	180	contrast	88	105	sequences
39	179	procedures	89	105	specifically
40	174	host	90	104	temperature
41	171	birds	91	103	prevent
42	170	primary	92	103	procedure
43	169	sequence	93	102	capacity
44	168	interactions	94	102	characteristics
45	164	binding	95	102	conducted
46	163	signaling	96	102	enhanced
47	162	samples	97	102	exhibited
48	160	exhibit	98	101	mechanism
49	158	relatively	99	101	relative
50	157	consistent	100	100	monitoring

(N = 1,775)

Table 4.4 The 100 Most Frequent Words in GSL 3

Rank	f	Words	Rank	f	Words
1	1008	species	51	102	altered
2	694	exposure	52	101	harm
3	648	rats	53	101	therapies
4	621	genetic	54	99	transportation
5	533	clinical	55	98	sensitive
6	513	infection	56	97	resistance
7	510	found	57	96	deficits
8	429	laboratory	58	96	density
9	340	experimental	59	96	disorder
10	292	virus	60	93	intervention
11	287	tissue	61	91	nerves
12	280	infections	62	90	muscle
13	272	strains	63	89	frequency
14	238	therapy	64	88	evaluation
15	230	depression	65	88	respectively
16	211	nerve	66	85	extensive
17	198	survival	67	85	personnel
18	195	disorders	68	85	ratio
19	186	functional	69	84	interventions
20	186	typically	70	84	versus
21	184	tissues	71	83	distinct
22	180	platform	72	82	bias
23	174	breeding	73	81	laboratories
24	170	variation	74	81	rearing
25	168	characterized	75	81	tanks
26	164	welfare	76	80	evaluate
27	159	viruses	77	80	nervous
28	155	symptoms	78	79	isolated
29	153	biological	79	79	maintenance
30	150	rat	80	78	stem
31	149	dependent	81	76	detected
32	149	motor	82	75	hypothesis
33	144	naturally	83	73	carbon
34	140	subsequent	84	73	stroke
35	134	strain	85	72	pregnancy
36	132	united	86	71	evaluated
37	123	similarly	87	70	exposures
38	120	pigs	88	70	infant
39	120	rapid	89	68	alter
40	118	underlying	90	66	discovery
41	113	potentially	91	65	tail
42	112	infants	92	64	reared
43	111	formation	93	64	restriction
44	109	anxiety	94	64	ultimately
45	107	breeds	95	63	limitations
46	107	guidelines	96	63	tank
47	106	breed	97	62	apparent
48	103	numerous	98	62	involvement
49	103	recovery	99	62	regulate
50	103	secondary	100	62	surgery

(*N* = 1,167)

In addition, AWL, which accounted for 6.73%, occurred the least in the laboratory animal corpus. The total number of words found in AWL was 1,248. The most frequent word was 'behavioral' which occurred 485 times in the corpus (as shown in Table 4.5).

Table 4.5 The 100 Most Frequent Words in AWL

Rank	f	Words	Rank	f	Words
1	485	behavioral	51	105	criteria
2	388	induced	52	104	validity
3	338	tumor	53	103	spatial
4	303	maternal	54	102	receptors
5	282	monkeys	55	100	metabolism
6	256	infected	56	99	mortality
7	243	chronic	57	99	spontaneous
8	224	offspring	58	97	mutations
9	223	molecular	59	97	syndrome
10	203	immune	60	96	diagnostic
11	193	peripheral	61	94	chemotherapy
12	193	receptor	62	94	detection
13	190	tumors	63	94	duration
14	183	dose	64	92	liver
15	183	transmission	65	91	interval
16	171	cues	66	90	modifications
17	170	developmental	67	89	availability
18	167	conservation	68	89	diagnosis
19	167	impact	69	89	interestingly
20	161	neurons	70	88	incidence
21	159	acute	71	88	lung
22	159	traits	72	87	adverse
23	157	consumption	73	87	susceptible
24	154	acid	74	86	comparative
25	151	cognitive	75	86	onset
26	148	commonly	76	84	injection
27	147	ethical	77	84	separation
28	140	infectious	78	83	facilitate
29	139	differentiation	79	83	variants
30	139	induce	80	82	markers
31	139	prevalence	81	81	cortex
32	134	physiological	82	81	diverse
33	133	neural	83	80	colonies
34	133	pathways	84	79	candidate
35	131	biology	85	79	correlated
36	129	mediated	86	79	obtained
37	128	aspects	87	78	stimulus
38	126	sensory	88	77	colony
39	126	stimuli	89	75	pathway
40	124	genetics	90	74	respiratory
41	120	doses	91	72	published
42	119	transcription	92	72	surgical
43	118	parameters	93	71	organism
44	117	elevated	94	70	bacteria
45	117	genetically	95	70	differential
46	110	identification	96	69	induction
47	109	protocols	97	69	paradigm
48	109	sensitivity	98	69	protocol
49	106	fetal	99	69	readily
50	106	progression	100	68	antibodies

(*N* = 1,248)

Table 4.6 showed the finding of OWL in the review articles of laboratory animals. The proportion of OWL, which accounted for 65.87%, was the highest, and covered more than the half of the corpus. The most frequent occurrence was the word 'zebrafish', and the second occurrence was the word 'methylation'. The result showed that 'zebrafish' appeared 675 times in the corpus while 'methylation' appeared 533 times. This indicated that the word 'zebrafish' was used mostly and prominently in the review articles of laboratory animals because it appeared more frequently than the second and other occurrences.

Table 4.6 The 100 Most Frequent Words in OWL

Rank	<i>f</i>	Words	Rank	<i>f</i>	Words
1	675	zebrafish	51	115	prenatal
2	533	methylation	52	115	vivo
3	435	epigenetic	53	112	regulatory
4	372	cocaine	54	112	vitro
5	284	genome	55	110	cellular
6	254	canine	56	109	glucose
7	237	nicotine	57	107	mammalian
8	219	rodents	58	107	pathogens
9	216	histone	59	107	translational
10	206	wildlife	60	106	hippocampus
11	198	rhesus	61	106	inflammatory
12	196	neuropathy	62	106	proliferation
13	195	veterinary	63	104	neuronal
14	188	macaques	64	100	bladder
15	186	rodent	65	100	pathogenesis
16	185	primates	66	98	quail
17	185	swine	67	95	anesthetic
18	177	ketamine	68	95	efficacy
19	177	viral	69	94	avian
20	175	invertebrates	70	94	cortisol
21	175	nonhuman	71	94	macaque
22	170	genomic	72	93	administered
23	168	pet	73	93	phenotypes
24	165	influenza	74	91	chromatin
25	162	lesions	75	91	maze
26	161	primate	76	91	postnatal
27	160	bartonella	77	91	preclinical
28	159	vervet	78	90	allele
29	158	transgenic	79	89	degeneration
30	156	activation	80	89	distal
31	154	investigators	81	89	larval
32	147	promoter	82	89	somatic
33	145	extinction	83	89	testosterone
34	145	obesity	84	88	apes
35	143	metabolic	85	88	parrots
36	142	therapeutic	86	88	reproductive
37	141	alterations	87	88	sexually
38	141	diabetes	88	87	serum
39	141	insulin	89	87	spondylosis
40	134	plasma	90	85	mammals
41	128	anesthesia	91	85	neoplasia
42	128	baboons	92	85	withdrawal
43	128	lifespan	93	84	abnormalities
44	121	susceptibility	94	83	additionally
45	120	intake	95	82	marmosets
46	119	phenotype	96	82	zoonotic
47	118	addiction	97	81	covariates
48	118	biomedical	98	81	relapse
49	117	cage	99	81	stereotypies
50	115	imprinted	100	79	baseline

(*N* = 12,220)

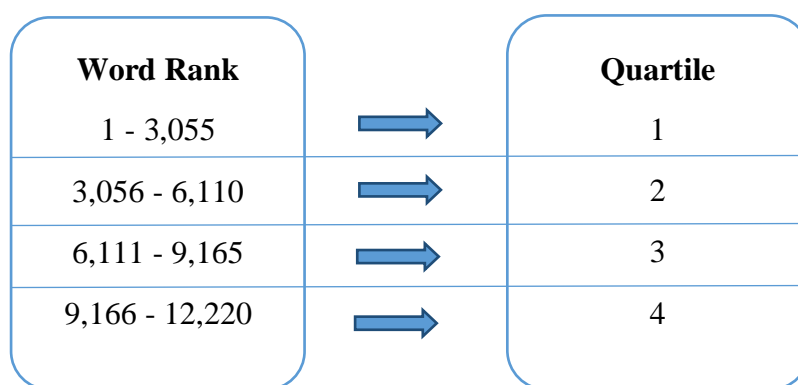
All in all, the results revealed the findings of the three word groups (GSL, AWL, and OWL) in laboratory animal review articles. The most occurrence was in the OWL which covered more than half of the corpus; while, the GSL took 27.40%, and the AWL covered the least.

4.2 The High-Frequency Content Words in OWL

Research question 2: What were the high-frequency content words in OWL used in the laboratory animal review articles?

The high-frequency words of the present corpus-based study were OWL because it was similar to technical words which occurred in the specific discipline (see Appendix C). Thus, only the OWL was considered for the high-frequency words of the present study.

Word frequency and quartile were the criteria for selecting high-frequency words in the Outside Word List (OWL). Nation (2001, p.14) stated that the frequency and ranking were crucial to identify high-frequency words. Therefore, the frequencies of OWL words were ranked from the highest to the lowest (see Appendix C). In addition, high-frequency words were determined by the quartile. The 1st quartile contains the highest word frequency while the 4th quartile contains the lowest frequency.



Total number of OWL is 12,220.

Figure 4.4 The quartile of OWL ranking

Figure 4.4 showed that the number of content words in OWL was 12,220. In order to group words into quartile, 12,220 was divided by 4, and the cutting point of the high-frequency words in each quartile was derived to be 3,055. From the results, the words in the 1st quartiles were ranked from 1 to 3,055. Therefore, the words appeared in the 1st quartile were considered as the high-frequency words.

Moreover, the results showed that by plotting graph, the top five highest word frequency of OWL decreased, dramatically. Unlike other words, the word frequency is slightly decreasing continuously. It means that these five words occurred mostly and prominently in the five recent years (2010-2014) more than the other words. The top five highest frequency words in OWL were *zebrafish*, *methylation*, *epigenetic*, *cocaine*, and *genome*. See Figure 4.5.

The frequency of the 100 most frequent words in OWL

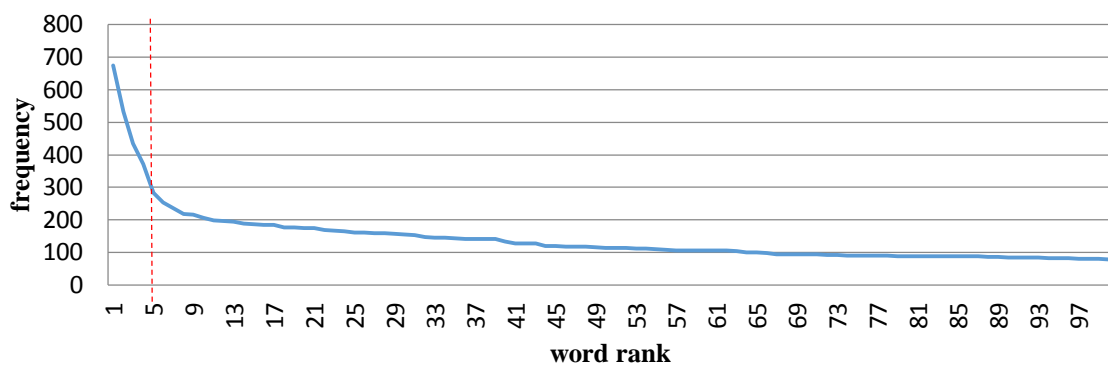


Figure 4.5 The frequency of the 100 most frequent words in OWL

In a summary, the high-frequency words of the present corpus were the content words from OWL. Frequency and quartile were used to consider for the high-frequency words. As a result, words that ranked from 1st to 3,055th were high frequency.

4.3 Collocations

Research question 3: What collocations of the highest-frequency content word in OWL were used in laboratory animal review articles?

The one purpose of the present study was to find the collocations of the highest-frequency content word in OWL. The word that occurred the most in the OWL was ‘zebrafish’; therefore, this word was analyzed for the collocations. The concordance program showed that there were 675 sentences with the word ‘zebrafish’ (Figure 4.6). These 675 sentences occurred in 14 articles. Ten out of 14 articles, where the word ‘zebrafish’ occurred, were published in 2012. Consequently, all the 675 sentences were analyzed for the collocations.

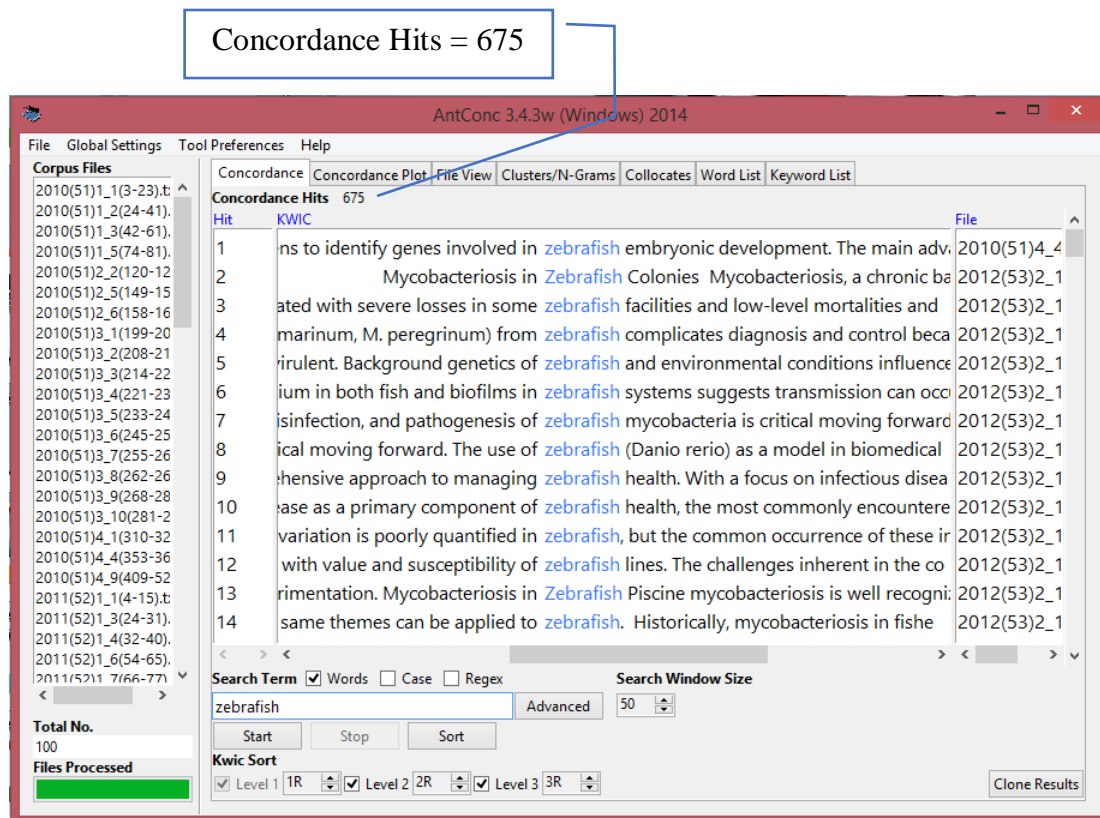


Figure 4.6 The concordance lines of ‘zebrafish’

According to Benson (1985), collocations normally were categorized into grammatical and lexical collocations. Grammatical collocations were a noun, verb, or adjective followed by a preposition while lexical collocations were the two equal

words combined together such as nouns, verbs, adverbs, and adjectives joined together. The results indicated that almost the collocations of 'zebrafish' were lexical collocation accounted for 494 (95.74 %) of collocations. Only 22 (4.26%) was of grammatical collocation. See Figure 4.7.

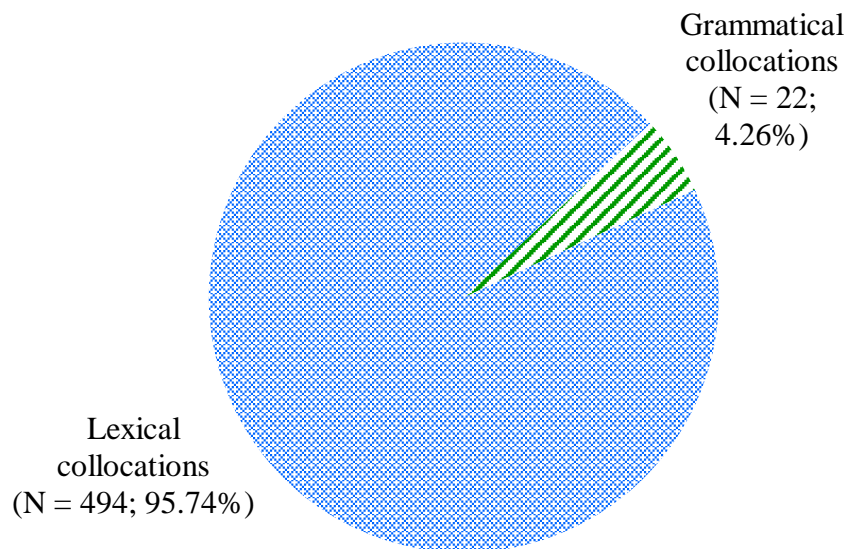


Figure 4.7 Types of collocations ($N_{\text{total}} = 516$)

4.3.1 Grammatical Collocation

The results found that there were 22 sentences of grammatical collocation (4.26%). The pattern of grammatical collocation found was *noun + preposition*. The word 'zebrafish' collocated the most with the preposition 'in', followed by 'with', and 'from'. For examples;

1. 'zebrafish' collocated with 'in' (found 11 sentences)

e.g.

(1) *The dramatic increase in the use of **zebrafish** (*Danio rerio*) **in** biomedical research has led to a corresponding increased interest in the diseases affecting this important biological model.*

(Sanders, Watral & Kent, 2012)

(2) *Severe hepatic megalocytosis was found in **zebrafish in** a new facility and was presumed to be associated with a toxicant from new plastics, glues, and the like.*

(Kent, Harper & Wolf, 2012)

(3) *For the foreseeable future, the use of **zebrafish in** biomedical research will continue to expand.*

(Matthews & Varga, 2012)

From the examples, the preposition 'in' was used to indicate space such as a location or place of 'zebrafish' in the experimental.

2. 'zebrafish' collocated with 'with' (found 10 sentences)

e.g.

(4) ***Zebrafish with** clinical microsporidiosis are emaciated and may exhibit skeletal deformities, but many infected fish are asymptomatic and the infection is often seen in sentinel fish.*

(Kent, Harper & Wolf, 2012)

(5) *Obviously, the control of this parasite in existing facilities is much more complex and requires systematic screening and isolation of **zebrafish with** known infections in order to eliminate or reduce the presence of *P. neurophilia* infections in the colony.*

(Sanders, Watral & Kent, 2012)

(6) *We are anxious to obtain live **zebrafish with** skin or fin papillomas because we believe that these papillomas are the best neoplasm candidates for harboring a tumorigenic virus of zebrafish.*

(Spitsbergen, Buhler & Peterson, 2012)

The result indicated that the preposition 'with' was used to express the relationship between a noun 'zebrafish' and other parts of the sentence. The examples

4 and 5 used the preposition 'with' to indicate being together or being involved. In addition, the example 6 showed that the preposition 'with' can be used to indicate "having".

3. 'zebrafish' collocated with 'from' (found 1 sentence)

e.g.

(7) *Watral and Kent added M. marinum to this growing list, isolating the bacterium from **zebrafish from** a facility supplying fish to the research community and experiencing low to moderate levels of mortality.*

(Whipps, Lieggi & Wagner, 2012)

The finding indicated that the preposition of 'from' was used to show the origin of 'zebrafish'. Moreover, the example showed that using 'from' together with 'to' can show the distance between two places from the start to the end.

In short, the grammatical collocation of the word 'zebrafish' accounted for 22 (4.26%) of the corpus. The result showed that the pattern of grammatical collocation was *noun + preposition* and preposition 'in' collocated mostly with 'zebrafish' to indicate location.

4.3.2 Lexical Collocation

The word 'zebrafish', which was the highest frequent word in OWL, was analyzed for lexical collocations. The results showed that the lexical collocations of 'zebrafish' accounted for 494 (95.74%) of the corpus. Moreover, the researcher found four patterns of the lexical collocations namely *noun + noun*, *noun + verb*, *adjective + noun*, and *verb + noun*. (as shown in Figure 4.8)

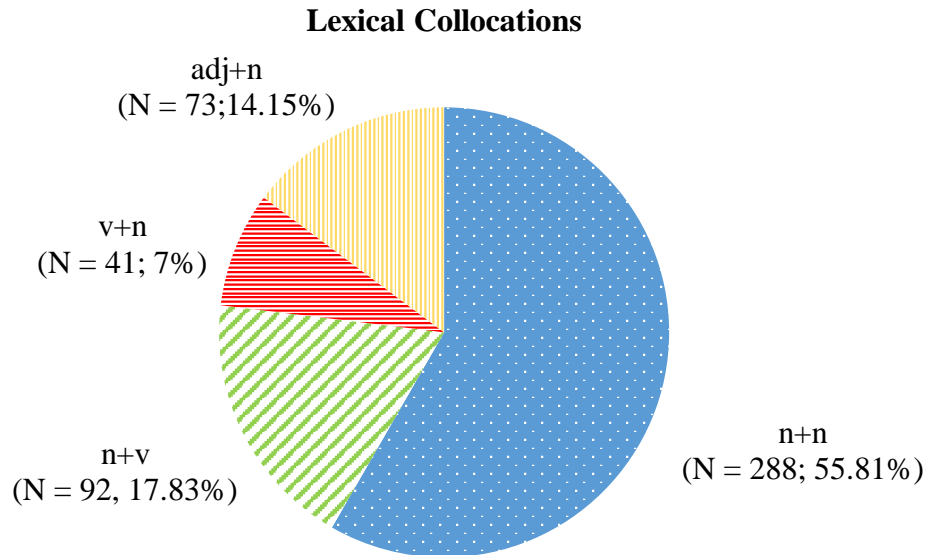


Figure 4.8 The proportion of lexical collocation (N_{total} = 494)

1. The pattern of ‘noun’ collocated with ‘noun’

The word ‘zebrafish’ collocated with a noun was found 288 times in (55.81% of) the corpus was had the most occurrence and appeared in more than the half of the present corpus. Most of them were formed as a compound noun such as *zebrafish model*, *zebrafish research*, *zebrafish facilities*, *zebrafish colonies*, *laboratory zebrafish*, *zebrafish system*, *zebrafish embryos*, *zebrafish husbandry*, *zebrafish adults*, *zebrafish lines*, *zebrafish reproduction*, and *zebrafish breeding*. See Table 4.7.

Table 4.7 The Examples of Lexical Collocations (noun + noun)

Collocation	<i>f</i>	Examples
Zebrafish model	22	<p>1. <i>Optimization of tools for inducible tissue-specific expression of transgenes and recently developed techniques for efficient targeted mutagenesis, such as zinc finger nucleases and transcription activator-like effector nucleases, now allow production of "custom-made" zebrafish models for precise histologic types of cancer affecting specific organs.</i> (Spitsbergen, Buhler & Peterson, 2012)</p> <p>2. <i>Because zebrafish models have expanded from developmental biology and genetics to include models for toxicology, aging, cancer, infection, and immunology, many kinds of research may be seriously and adversely affected by unidentified underlying viral infections.</i> (Crim & Riley, 2012)</p> <p>3. <i>However, advantages to zebrafish models for mammalian viral infections include the capacity for live imaging, whole-organism histopathology and immunohistochemistry, and temperature-shift experiments.</i> (Matthews & Varga, 2012)</p>
Zebrafish research	18	<p>1. <i>It is reasonable to assume that space-occupying or inflammatory lesions in the brain, spinal cord, meninges, or skeletal muscle could profoundly affect behavior, which is an emerging field of zebrafish research.</i> (Kent, Harper & Wolf, 2012)</p> <p>2. <i>Because no naturally occurring viral infections have been reported in zebrafish, it is not possible to provide specific examples of how naturally occurring viral infections have confounded zebrafish research.</i> (Crim & Riley, 2012)</p> <p>3. <i>However, zebrafish studies of infection and immunity, as well as other types of zebrafish research, are at risk of being confounded by unrecognized naturally occurring viral infections.</i> (Wilson, 2012)</p>

Table 4.7 The Examples of Lexical Collocations (noun + noun) (cont.)

Collocation	<i>f</i>	Examples
Laboratory zebrafish	14	<p>1. <i>This research not only demonstrates that laboratory zebrafish may still harbor a variety of pathogens from their origin in the pet trade but also provides evidence that zebrafish raised in outdoor ponds continue to directly enter laboratory colonies, providing a constant source of new pathogens from other pet and aquarium species.</i> (Crim & Riley, 2012)</p> <p>2. <i>Laboratory zebrafish typically attain sexual maturity at the 3rd month of their development.</i> (Nasiadka & Clark, 2012)</p>
Zebrafish system	11	<p>1. <i>Mycobacterium fortuitum infection appears to be a manageable disease but is not usually entirely eliminated from a large recirculation zebrafish system.</i> (Whipps, Lieggi & Wagner, 2012)</p> <p>2. <i>More sophisticated means of solids removal are now a possibility in commercially available zebrafish systems.</i> (Lawrence & Mason, 2012)</p>
Zebrafish husbandry	6	<p>1. <i>There are many areas of zebrafish husbandry that still require a great deal of research, and this includes larval rearing.</i> (Wilson, 2012)</p> <p>2. <i>Breeding is a key element in zebrafish husbandry.</i> (Nasiadka & Clark, 2012)</p>
Zebrafish adults	5	<p>1. <i>The need for embryos has fueled the development of specialized equipment designed to encourage zebrafish adults to breed.</i> (Lawrence & Mason, 2012)</p> <p>2. <i>The immobilization of zebrafish adults by submersion in ice water (2-4°C) immediately followed by cranial maceration via an in-sink garbage disposal causes immediate destruction of all brain function and activity.</i> (Matthews & Varga, 2012)</p>
<i>(N = 288)</i>		

From Table 4.7, the noun and nouns were put together to form a single noun called compound noun. The combinations of noun can generate new words. Moreover, the noun can be used as an adjective to modify the primary noun.

2. The pattern of 'noun' collocated with 'verb'

Table 4.8 showed the results of lexical collocations in a form of noun collocated with verb. There were 92 sentences which were accounted for 17.83%. The noun 'zebrafish' collocated with verb namely *used*, *raised*, *treated*, *maintained*, *exhibit*, *emerged*, *spawn*, *obtained*, *survive*, and *infected*. More than an half of the word 'zebrafish' collocated with verbs was written in the form of active voice (63.54%). 33 sentences or 36.46% was written in the form of passive voice.

Table 4.8 The Examples of Lexical Collocations (noun + verb)

Collocation	<i>f</i>	Examples
use	7	<p>1. <i>However, investigators may not be fully aware of the potential impact that underlying disease may have on their research when postlarval zebrafish are used in medium- or long-term experiments.</i> (Kent, Harper & Wolf, 2012)</p> <p>2. <i>Zebrafish and other fish larvae use cutaneous gas exchange for their oxygen demands while their gills are still developing.</i> (Matthews & Varga, 2012)</p> <p>3. <i>Such nonexperimental variation magnifies the importance of evaluating the influences these infections may have when zebrafish are used as models for studies on disease, immunology, ecotoxicology, and so on.</i> (Whipps, Lieggi & Wagner, 2012)</p>
raised	4	<p>1. <i>In contrast to purpose bred research mice, ornamental zebrafish are often raised together with other species that harbor numerous pathogens.</i> (Crim & Riley, 2012)</p> <p>2. <i>In addition, the diagnostic pathology service at the Zebrafish International Resource Center occasionally detects common aquarium fish helminth parasites, indicating that the submitted zebrafish were raised in outdoor ponds.</i> (Lawrence & Mason, 2012)</p>
treated	3	<p><i>Florida strain wild type zebrafish treated with N-nitroso-N-ethylurea in one study exhibited 100% incidence of cutaneous papillomas; however, no cutaneous papillomas were observed in similar experiments conducted at several other research institutions, suggesting the possibility of an unrecognized oncogenic virus.</i> (Crim & Riley, 2012)</p>
maintained	3	<p><i>For example, zebrafish are often maintained at approximately 28°C, whereas some mammalian viruses adapted to replicate at 37°C may not be pathogenic at 28°C.</i> (Wilson, 2012)</p>
spawn	2	<p><i>Sexually mature zebrafish can spawn in the laboratory continuously all year at a frequency of two or three times a week.</i> (Nasiadka & Clark, 2012)</p>
(N = 92)		

Table 4.8 showed the pattern of noun collocated with verb by using active and passive voices. Generally, the active voice was a more straight-forward and more

effective way to express action. Passive voice was used to emphasize the receiver of the action, or when the person or thing performing the action was unknown.

3. The pattern of ‘adjective’ collocated with ‘noun’

Another aspect of lexical collocation found in the present study was an adjective collated with a noun (see Table 4.9). The findings showed 73 sentences where adjectives collocated with noun (14.15%) of the corpus. The examples were *adult zebrafish*, *infected zebrafish*, *larval zebrafish*, *immunocompromised zebrafish*, and *pathogen-free zebrafish*.

Table 4.9 The Examples of Lexical Collocations (adjective + noun)

Collocation	<i>f</i>	Examples
Adult zebrafish	13	<p>1. <i>The brain of adult zebrafish is histologically quite distinct from that of mammals, with a much greater component of highly cellular areas composed of deeply basophilic embryonal cells surrounding the ventricular system of the forebrain, diencephalon, and myelencephalon.</i> (Spitsbergen, Buhler & Peterson, 2012)</p> <p>2. <i>A 2- to 5-g/ml dosage concentration is recommended to sedate adult zebrafish, and 60 to 100 ?g/ml is used for immersion anesthesia.</i> (Matthews & Varga, 2012)</p> <p>3. <i>Because this physical method produces rapid loss of consciousness and subsequent death, thereby minimizing pain, distress, or discomfort to the animal, we believe it to be an acceptable euthanasia technique for adult zebrafish.</i> (Nasiadka & Clark, 2012)</p> <p>4. <i>At midlarval stage, the feeding regimens become very disparate; although Artemia is a well-known foodstuff for both adult zebrafish and larval forms with corresponding gape size, usually from 8 to 9 DPF, some facilities do not provide it until 14 to 15 DPF.</i> (Wilson, 2012)</p>

Table 4.9 The Examples of Lexical Collocations (adjective + noun) (cont.)

Collocation	<i>f</i>	Examples
Larval zebrafish	5	<p>1. <i>At this point, some aspects of water quality do not seem as important because larval zebrafish seem more tolerant of poorer water quality with respect to free ammonia than adult fish, although temperature will still affect the rate of larval development and thus staging.</i> (Wilson, 2012)</p> <p>2. <i>Larval zebrafish are quite susceptible to <i>P. neurophilia</i> infections; hence, the infection could compromise developmental neurotoxicity testing and musculoskeletal development studies.</i> (Kent, Harper & Wolf, 2012)</p>
Infected zebrafish	4	<p>1. <i>Nevertheless, treatment of infected zebrafish may be appropriate when extremely valuable strains or populations are involved.</i> (Whipps, Lieggi & Wagner, 2012)</p> <p>2. <i>Further evidence for the maternal transmission of <i>P. neurophilia</i> was observed in the experiment described by Sanders and Kent, where parasite DNA was detected in the eggs and water from a group spawn of infected zebrafish.</i> (Sanders, Watral & Kent, 2012)</p>
Immunocompromised zebrafish	3	<p>1. <i>As with other species, immunocompromised zebrafish are likely to be more susceptible to viral infections than wild-type fish, exhibiting higher mortality, higher morbidity, more clinical signs, and more severe histopathologic lesions.</i> (Crim & Riley, 2012)</p> <p>2. <i>In recent years, researchers have developed immunocompromised zebrafish to aid in the study of hematopoiesis, tumorigenesis, infection, and immunity.</i> (Crim & Riley, 2012)</p>
Pathogen-free zebrafish	2	<p>1. <i>The only current effort to generate pathogen-free zebrafish is being conducted at the Sinnhuber Aquatic Resources Laboratory at Oregon State University.</i> (Matthews & Varga, 2012)</p> <p>2. <i>For development of pathogen-free zebrafish to be successful, however, naturally occurring viral pathogens affecting zebrafish colonies must be identified and diagnostic assays to detect viral infection must be developed to facilitate the elimination of viral pathogens.</i> (Matthews & Varga, 2012)</p>

(N = 73)

Table 4.9 showed that most descriptive adjectives used to modify 'zebrafish' were the subjective or evaluative adjective and measurement adjective. The subjective or evaluative adjective was used to state the quality of zebrafish such as *important*, *pathogen-free* and *standard*. The evaluative adjective also used to describe the age of 'zebrafish' such as *adult* and *larval*. Moreover, the measurement adjective in this study represented size of 'zebrafish' such as *large*. In addition, the use of adjectives made the sentence more vivid and descriptive.

4 The pattern of 'verb' collocated with 'noun'

The least occurrences of lexical collocation found in the corpus were verb collocated with noun. The total number of sentences found was 41 or 7% of the corpus. Verbs that collocated with the word 'zebrafish' were *using*, *predispose*, *infect*, *describe*, and *sedate*. See Table 4.10.

Table 4.10 The Examples of Lexical Collocation (verb + noun)

Collocation	<i>f</i>	Example
sing	6	<p>1. <i>Oncogenesis due to unrecognized viruses may therefore be a confounding variable in cancer research using zebrafish.</i> (Crim & Riley, 2012)</p> <p>2. <i>Many of the institutions using zebrafish (<i>Danio rerio</i>) for research, testing, or teaching are funded by the Public Health Service (PHS) and accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International.</i> (Matthews & Varga, 2012)</p>
predispose	2	<p><i>These abundant embryonal periventricular cells may predispose zebrafish to develop more embryonal neoplasms of the central nervous system, resembling those seen in pediatric cases in humans.</i> (Spitsbergen, Buhler & Peterson, 2012)</p>
infect	2	<p><i>The lack of information about naturally occurring viral infections in zebrafish reflects a lack of investigation in this area rather than an inability of viral pathogens to infect zebrafish, as evidenced by studies documenting the experimental infection of zebrafish with viruses isolated from other fish species and the presence of multiple endogenous retroviruses, retrotransposons, and other retroid agents in the zebrafish genome.</i> (Crim & Riley, 2012)</p>
sedate	2	<p><i>A 2- to 5-g/ml dosage concentration is recommended to sedate adult zebrafish, and 60 to 100 g/ml is used for immersion anesthesia.</i> (Matthews & Varga, 2012)</p>

(N = 41)

From the results, the writer used action verb to represent behavior or physical actions. The verb that collocated with noun was transitive verb. The transitive verb was used when its action is directed toward something, which is the object of the verb. For results of this study, the object of the verb was the word 'zebrafish'.

In conclusion, the highest-frequency word in OWL of this study was purposively analyzed for grammatical and lexical collocation. The most frequent word occurrence in OWL was 'zebrafish' which collocated mostly as lexical collocations. The results showed that there were four patterns of lexical collocations found in the corpus. Those patterns were *noun + noun*, *noun + verb*, *adjective + noun*, and *verb +*

noun. The results also showed a a grammatical collocation pattern as *noun + preposition*.

4.4 Purposive Function

Research question 4: What were the patterns used to state the purposes in the laboratory animal review articles?

This corpus-based study also aimed to find the patterns used to state purposive function of the word ‘zebrafish’ which was found as the highest-frequency word in OWL. Regarding the word ‘zebrafish’, the concordance plot showed that the 675 sentences came from 14 review articles of ILAR journal (Figure 4.9).

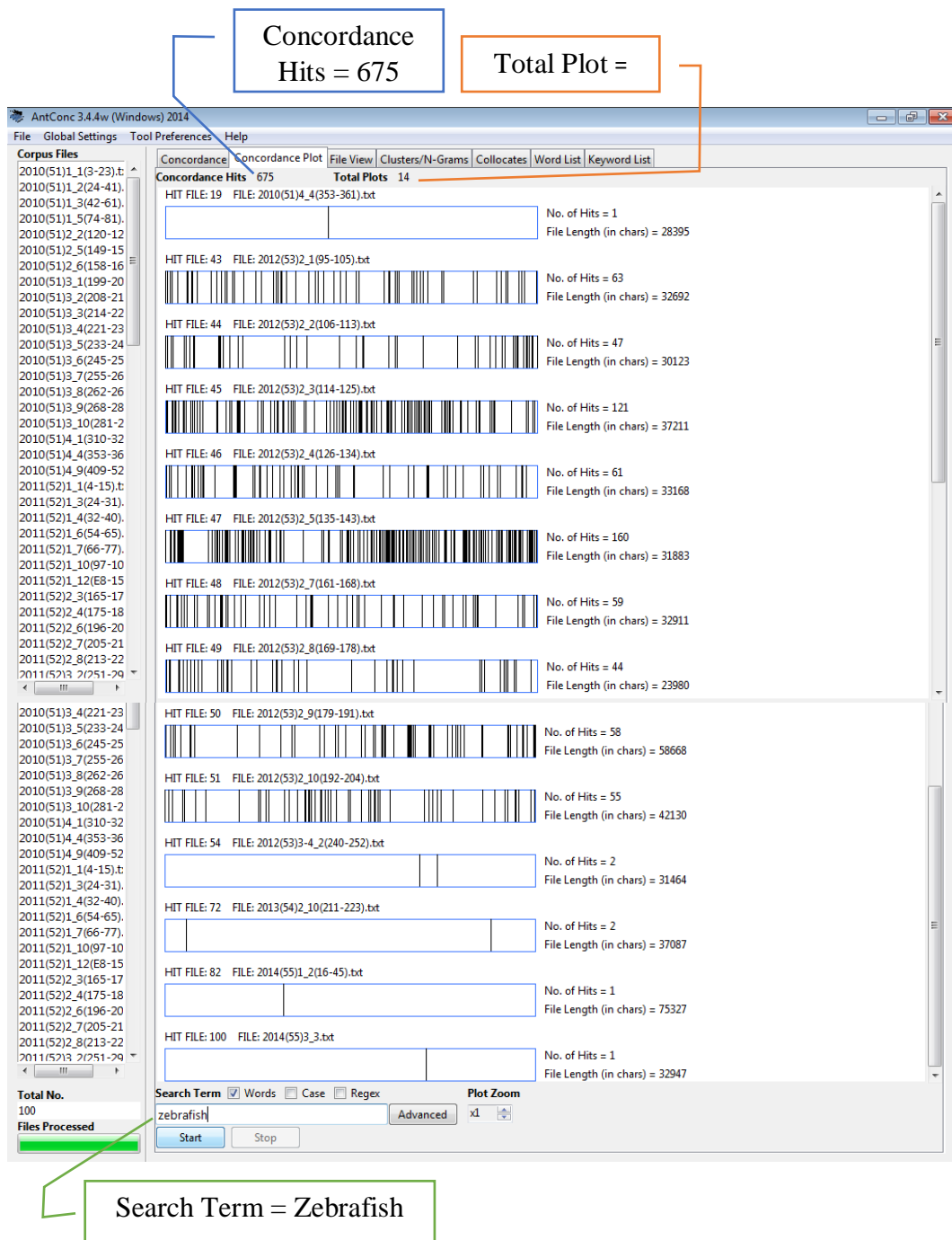


Figure 4.9 Concordance plot of the word ‘zebrafish’

Therefore, these 14 review articles were investigated in the introductory section in order to find the purposive statement which included the word ‘zebrafish’. The results indicated that there were six purposive sentences with the word

'zebrafish'. These six purposive sentences were written in two different patterns as following.

Pattern1

Impersonal subject	Main verb	Phrase / Clause
This (present) review	highlights	noun phrase or noun clause
This studies	focuses on	
This article	addresses outlines	

The finding showed that there were five sentences used the main verb to state the purpose of the articles. The main verbs were written in present simple and present future tenses.

Example 1



*The present review focuses specifically on the transmission and control of microsporidia in **zebrafish** facilities.*

(Sanders, Watral & Kent, 2012)

Example 2



*These studies highlight the need for careful consideration of diet and husbandry in order to ensure valid and reproducible data in research using the **zebrafish** model.*

(Spitsbergen, Buhler & Peterson, 2012)

Example 3

*This article addresses the importance of identifying and characterizing the viral diseases of **zebrafish** as the scope of **zebrafish** models expands into new research areas and also briefly addresses **zebrafish** susceptibility to experimental viral infection and the utility of the **zebrafish** as an infection and immunology model.*

(Crim & Riley, 2012)

Example 4

*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**.*

(Matthews & Varga, 2012)

Example 5

*Although model systems such as drosophila, C. elegans and **zebrafish** have played important roles in biomedicine, such as the discovery of the regulatory role of noncoding RNA in healthy and disease conditions (Ambros, 2003; Ambros, 2008), this review will focus on the mammalian model systems, specifically nonhuman primates (NHPs).*

(Chan, 2013)

Pattern 2

Subject (signal words of purposes)			Verb to be	to + V.inf	Noun phrase
The goal (s)	of	this review	is	to provide	Noun details of objective(s)

Another purposive pattern used to state a purpose was to-infinitive. The sentence was written in the present simple tense followed by to-infinitive in order to indicate the details of the objectives which can be written in the noun phrase.

Example 1

signal word
to V.inf
*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.*

(Lawrence & Mason, 2012)

In summary, the purposes of the study in review articles can be found in their introductory section. Moreover, the sentences that showed purposive function can be stated by the main verbs that expressed the objective and to-infinitive.

Summary Chapter

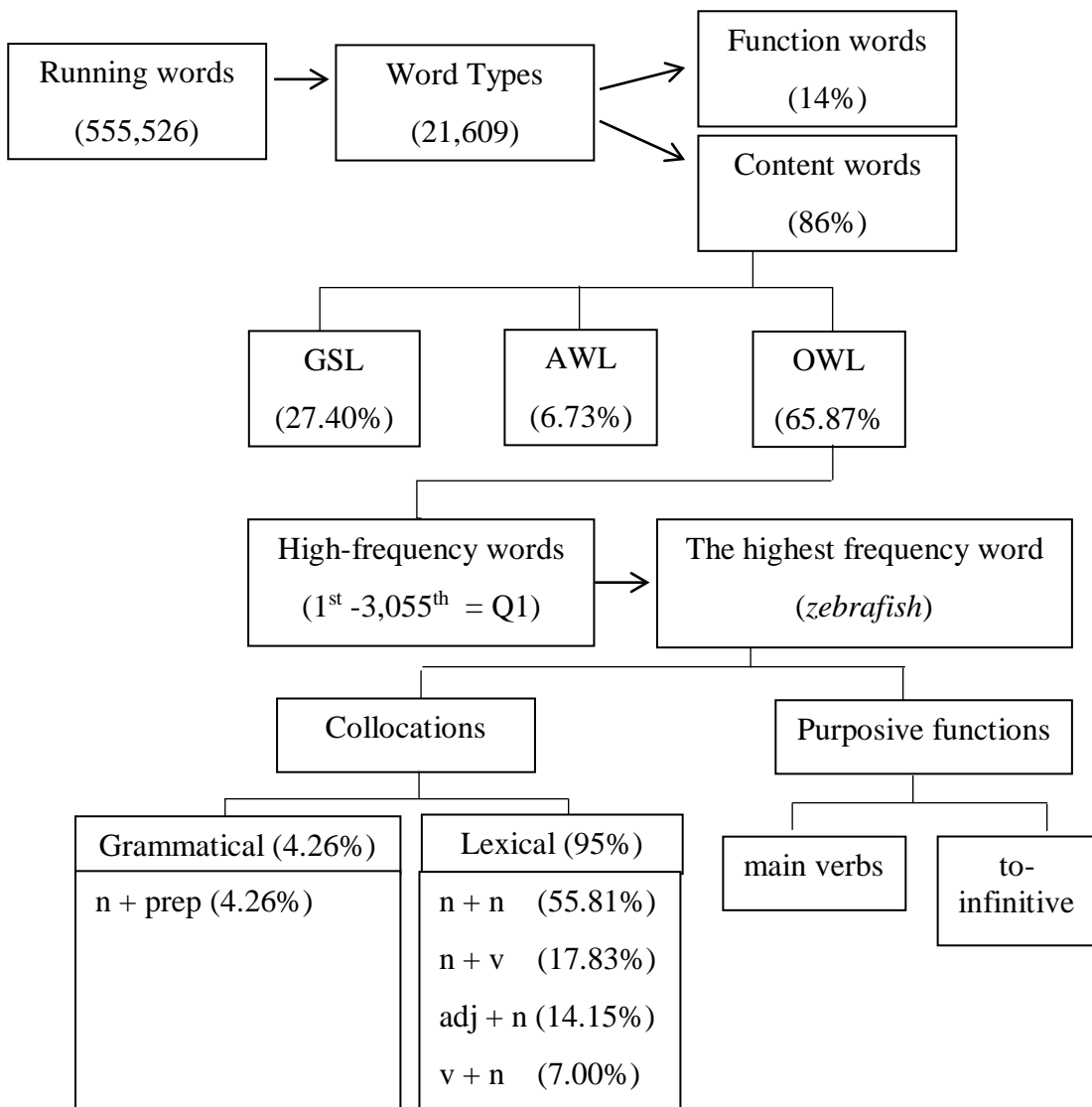


Figure 4.10 The results of the study

In conclusion, the present corpus of laboratory animal review articles comprised of 555,526 running words. The word types of the corpus accounted for 21,609 words which had both function words (14%) and content words (86%). As the focal words in the present study were content words, the content words were categorized into GSL, AWL, and OWL. Words in the laboratory corpus were OWL (65.87%) followed by GSL (27.40%) and AWL (6.73%) categories. The OWL was also analyzed for high-frequency words by using quartile. The results showed that 3,055 words belonged to the 1st quartile, and thus, were considered as the high-frequency words. The frequencies of top five highest frequency words (*zebrafish*,

methylation, epigenetic, cocaine, and genome) had dropped as shown by plotting graph (see Figure 4.4). Therefore, these high-frequency words distinctively occurred and used more than others in the laboratory animal review articles. In addition, the highest-frequency word namely 'zebrafish' was analyzed for its grammatical and lexical collocations. The collocations were lexical collocation (95.74%) which represented five patterns: 1) noun + noun (55.81%), 2) noun + verb (17.83%), 3) adjective + noun (14.15%), and 5) verb + noun (7.00%). Lastly, the results of purposive functions of the word 'zebrafish' showed that they used the main verbs (*highlight, focus on, addresses, and outlines*) to state the purpose of the research.

CHAPTER V

DISCUSSION

This chapter is to explain, interpret, and discuss the findings of the research that relates to the objectives of the study. The main findings of the content words, collocations, and communicative function are discussed in detail. Later, the results of the study are discussed for their implications. Thus, this chapter consists of five main areas.

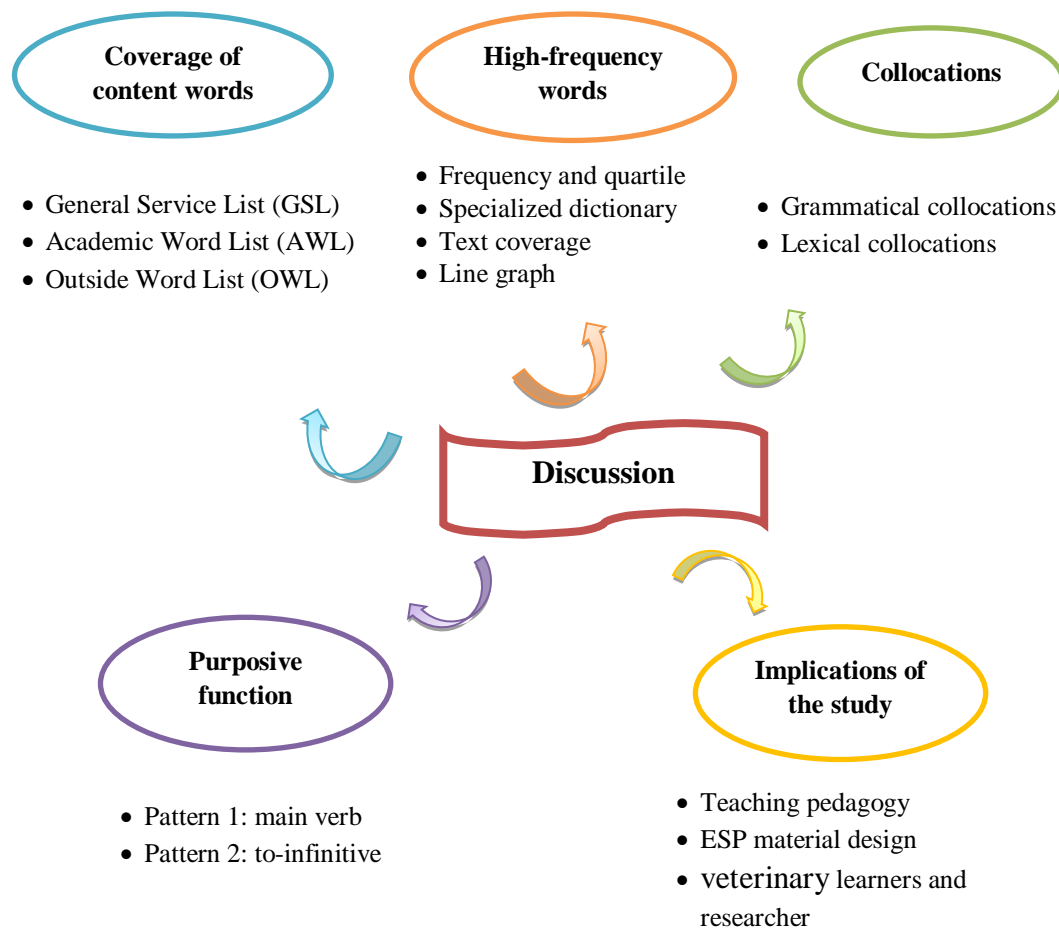


Figure 5.1 The overview of discussion

5.1 Content Words

After presenting the findings of the study, the results of content words in GSL, AWL, and OWL in the laboratory animal review articles will be discussed in this section. The analysis of content words and high-frequency words in OWL will be discussed in terms of the research questions.

5.1.1 The content words of GSL, AWL, and OWL used in the laboratory animal review articles

The research question 1 was, “What were the content words of GSL, AWL, and OWL used in the laboratory animal review articles?” As shown in Figure 5.1, the coverage of AWL in laboratory animal review was lowest (6.73%), the highest was in OWL (65.87%), and the coverage of GSL was 27.4%. Regarding the investigation of text types and text coverage, the academic word list (AWL) covered approximately 9% of academic texts (Coxhead, 2000). The most frequent 2,000 words of English or GSL investigated by West (1953) provided 80% of text coverage. Therefore, the results of GSL and AWL found in the present study were less than in previous studies.

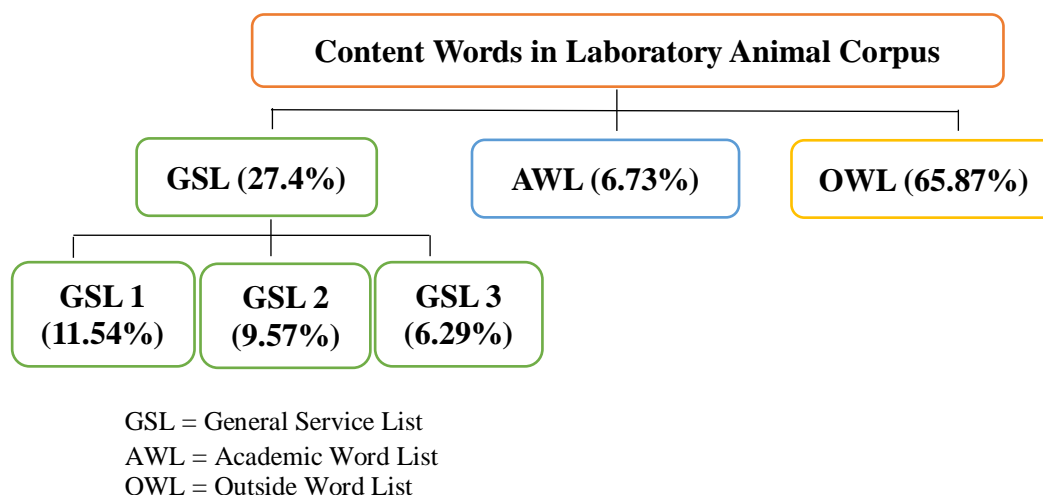


Figure 5.2 The results of word lists

5.1.1.1 The coverage of AWL

The result of AWL accounted for 6.73% of running words in the present corpus, which was much less than Coxhead’s AWL (9%) and other previous AWL studies in scientific research. Nation (2001) mentioned that the AWL coverage of academic texts generally covered 9% to 10%. For example, Wang, Liang, and Ge (2008) investigating a corpus study in the medical field found an AWL coverage of 10.07%. Martinez, Beck, and Panza (2009) conducted a corpus study in agricultural science, and the found an AWL coverage of 9.06%. Valipouri and Nassaji’s chemistry corpus (2013) found that the AWL coverage was 9.06%. The coverage of AWL (6.73%) in the present study was smaller than in other previous studies because the source of the corpus was not large (Figure 5.3).

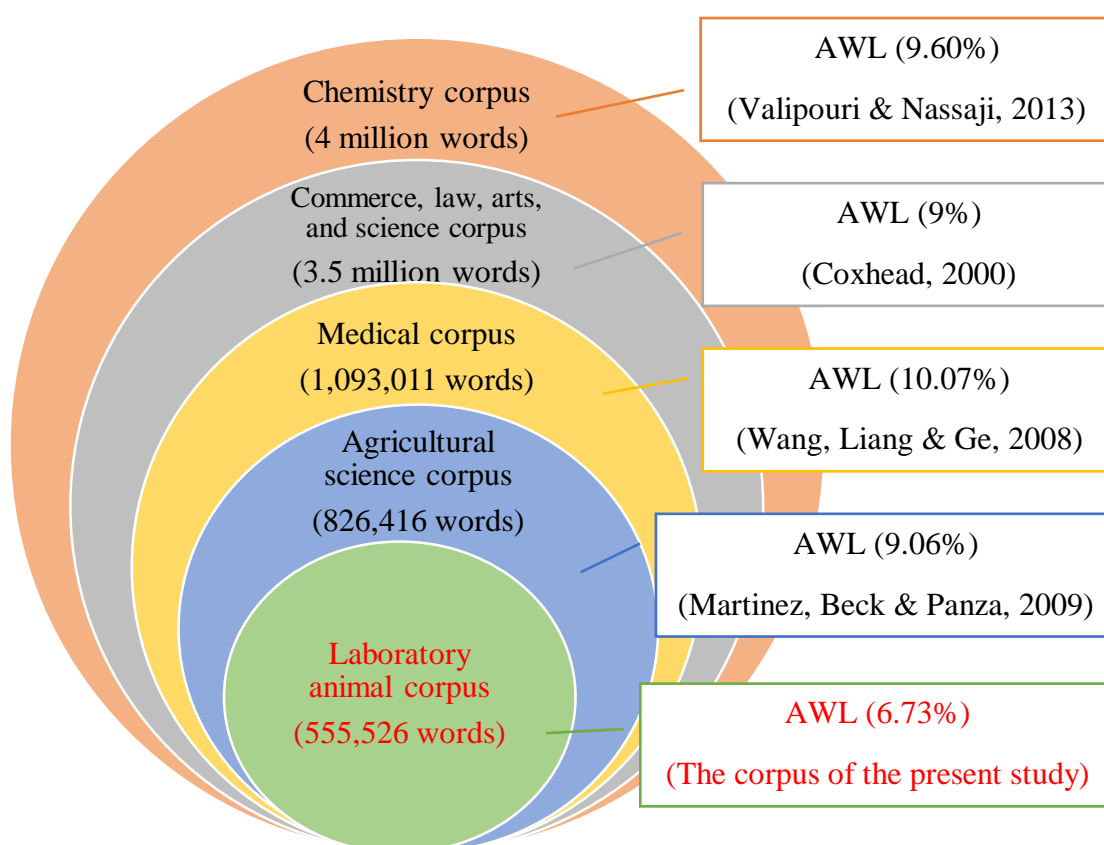


Figure 5.3 The size of AWL corpus in the previous scientific studies

Moreover, this laboratory animal corpus was collected from only one specialized branch in science, and source of corpus was gathered from only

review articles in the ILAR journal. The chance that AWL appeared in this corpus was the words from only science discipline. Unlike Coxhead's corpus (2000), Coxhead made a corpus from four disciplines: commerce, law, arts, and science. AWL can occur in all these disciplines. In brief, the result of AWL in the current study was less than the previous findings due to the small size and source of corpus.

5.1.1.2 The coverage of GSL and OWL

The proportion of GSL (27.4%) in this corpus was much smaller than West's GSL (1953). The most frequent 2,000 GSL words (West, 1953) included both content and function words, and those function words occurred frequently in a corpus as adverbial particles, auxiliary verbs, prepositions, conjunctions, determiners, pronouns, and numbers. It is contrastive to the present study that the function words were excluded. Therefore, the proportion of GSL was lower. Only content words were selected because they provided lexical meanings rather than structural functions, and because they were extensive to increase lexical knowledge. It was stated by Fries (1952) that function words or the closed class of words in grammar was rarely intended to have new functions. Unlike content words or the open class of the words, new words may be generated all the time, as for example, slang words, technical terms, and adoptions and adaptations of foreign words. Moreover, West's GSL was selected from large corpora in various disciplines such as humanity, law, science, and commerce. In contrast, this study focused only on the specific field of laboratory animals. SCImago Journal & Country Rank (SJR), the journals and country scientific indicators, showed that ILAR, which was the source of the study, was categorized in the field of Animal. As the source of the present corpus was the ILAR journal, which was in the specialized field of laboratory animals, the specialized words mostly occurred as OWL. For this reason, the present study considered OWL as high-frequency words because OWL had the specialized words similar to technical words occurred in the specific field.

In short, the discussion explained the findings of three word groups. The result of AWL and GSL were less than Coxhead's AWL (2000) and West's GSL (1953) because function words were excluded from GSL. In addition, source of the present corpus was from the specialized field in laboratory animals. Thus, OWL occurred most frequently in this corpus.

5.1.2 The high-frequency content words in OWL used in the laboratory animal review articles

Research question 2 was, “What were the high-frequency content words in OWL used in the laboratory animal review articles?” The findings of OWL in the current study were analyzed for high-frequency words. The findings show that the total number of OWL words was 12,220, out of which 3,055 words were considered as high frequency. There were four criteria used to explain the findings of high-frequency words (as shown in Figure 5.3).



Figure 5.4 The criteria of high-frequency words

Frequency and Quartile

The results of OWL showed and ranked high-frequency words from 1st to 3,055th (1st quartile). According to Nation (2001), the agreement of high-frequency words was accounted for by almost 80% of the running words. High-frequency words in the present corpus were less than 80%, because they referred to OWL words used particularly in own discipline, which contained specialized meanings. Unlike GSL and AWL, they occurred frequently in the general fields and general academic fields. As the focal words in the present study were OWL and generally occurred with low frequency, the quartile was employed as criteria to determine high-frequency words in OWL.

Dictionary

The findings from Scimago Journal and Country Rank (SJC) categorize the ILAR journal in three disciplines: 1) Animal Science and Zoology, 2) Biochemistry, and 3) Genetics and Molecular Biology. Nation (2001, p.201) stated that specialized or technical words can be verified by consulting specialized dictionaries. Consequently, the findings of high-frequency word in OWL were checked against two dictionaries, Oxford Dictionary of Zoology and Oxford Dictionary of Biochemistry and Molecular Biology. The study found that the high-

frequency words of OWL namely *zebrafish*, *methylation*, *epigenetic*, *cocaine*, and *genome* were found in the two specialized dictionaries. It can be seen that these words were specialized or technical words, and they were highly topic-related and directly reflected in the corpus of laboratory animal review articles. The findings also supported Nation's claim (2001) that technical words were words that were closely related to the subject discipline of the text. Moreover, it is similar to Kaewphanngam's study (2002) which comprised technical words in psychology corpus, and found that psychology technical words appeared in The Dictionary of Psychology. In Pitukwong's study (2012), she identified technical words in business corpus, and employed a business dictionary to check for business technical words. All in all, high-frequency words in OWL refer to technical words because OWL is the list of words that belong to neither GSL nor AWL, and high-frequency words in OWL also appear in specialized dictionaries.

Text Coverage

In addition, text coverage was used to assure that the high-frequency words found appeared in wide and different range. The top five high-frequency words were checked for their text coverage, and found to have occurred in more than eight different articles and were accounted for 8% of text coverage (Table 5.1).

Table 5.1 The Coverage of the Top Five High-frequency Words in OWL

Rank	Word	Frequency	Text coverage (%)
1	zebrafish	675	14
2	methylation	533	16
3	epigenetic	435	22
4	cocaine	372	9
5	genome	284	36
Total		2,299	100

Likely, to Nation (2001, p.16), he stated that high-frequency words covered approximately 8 to 15% of different ranges. As the length of pages in any article could result in high word frequency, the high-frequency words were verified by range or text coverage in order to assure that they occurred in different texts or articles. Briefly, the text coverage of high-frequency words indicated that *zebrafish*,

methylation, *epigenetic*, *cocaine*, and *genome* were used in a wide range in the discipline.

Line graph

Generally, high-frequency words can be determined by the above-mentioned criteria such as frequency, dictionary, and text coverage (Nation, 2001). However, the results of present study indicated that line graph can demonstrate a clearer use of word frequencies. Plotting the graph gave a better view of the word frequency because the line graph represented the tendency of whole numbers. Normally, the word frequency decreases slightly. However, the graph (Figure 5.5) showed that the frequencies of the top five highest-frequency words in OWL dropped, dramatically.

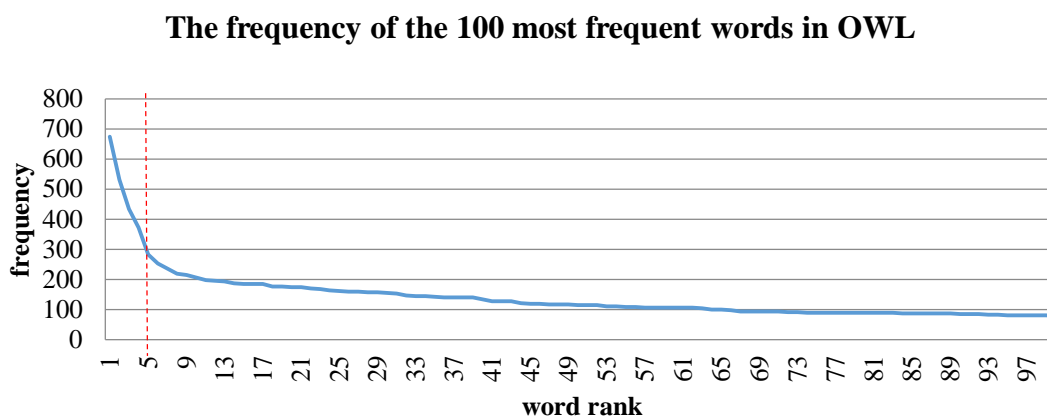


Figure 5.5 The Frequency of the high-frequency words in OWL

The line graph indicated that there was a wide gap between the word groups in OWL. In other words, the group of the top five highest-frequency words was used mostly and distinctively in the five recent years (2010-2014) more than the others. As a result, the top five highest-frequency words of OWL namely *zebrafish*, *methylation*, *epigenetic*, *cocaine*, and *genome* were used prominently in the laboratory animal corpus. These five words were distinctively used in laboratory review articles because the theme of articles in 2012 highlighted on Neurobiology of Addiction-like Behaviors (volume 1), Zebrafish Health and Husbandry (volume 2), and Epigenetics (volume 3 and 4). Moreover, the number of articles in 2012 published more than other

years. Consequently, the word ‘zebrafish’ occurred highly in the corpus because zebrafish was the target of investigation in the review of laboratory animals. The other high-frequency words were *methylation*, *epigenetic*, and *genome* because the main topic of volume 3 focused on epigenetics. Therefore, these three words occurring with high-frequency related to gene and DNA. As the main topic of volume 1 was ‘Addiction-like Behaviors’, researchers rely on experiments using animals to test how drugs promote abuse and produce addiction which affect the brain and body. Thus, ‘cocaine’ occurred highly because it was one of the addictive drug used in experiments. In short, the high-frequency words occurred in line with the focal topic or theme of the journal published in each year.

Overall, high-frequency words of OWL were discussed by frequency and quartile, line graph, specialized dictionary, and text coverage. The results of 3,055 high-frequency words were determined by word frequency and quartile. Furthermore, the top five highest-frequency words (*zebrafish*, *methylation*, *epigenetic*, *cocaine*, and *genome*) were the specialized words that used most widely in the field of laboratory animals.

5.2 Collocations

The research question 3 was, “what collocations of the highest-frequency content word in OWL were used in laboratory animal review articles?” Most collocations in the OWL found in this corpus were two-word combinations; for example, *zebrafish colonies*, *laboratory zebrafish*, *zebrafish embryos*, *zebrafish husbandry*, *adult zebrafish*, *infected zebrafish*, and *larval zebrafish*. Collocations were separated into various types. According to Benson (1985), collocations can categorize into grammatical collocations and lexical collocations. The grammatical collocations referred to a noun, verb, or adjective joined by a preposition while lexical collocations were the two-word combination of nouns, verbs, and adjectives. Due to the source of the present corpus, the findings of word collocations in this study followed the framework of grammatical and lexical collocations (Benson, 1985). Source of the corpus was collected from review articles that were academic texts, and written informal language. Thus, idioms and phrasal verbs, which were normally used in a

spoken language, do not appear in the current corpus. As a result, the collocations of the OWL in the laboratory corpus were analyzed by two-word combinations as known as grammatical collocations and lexical collocations (Table 5.6).

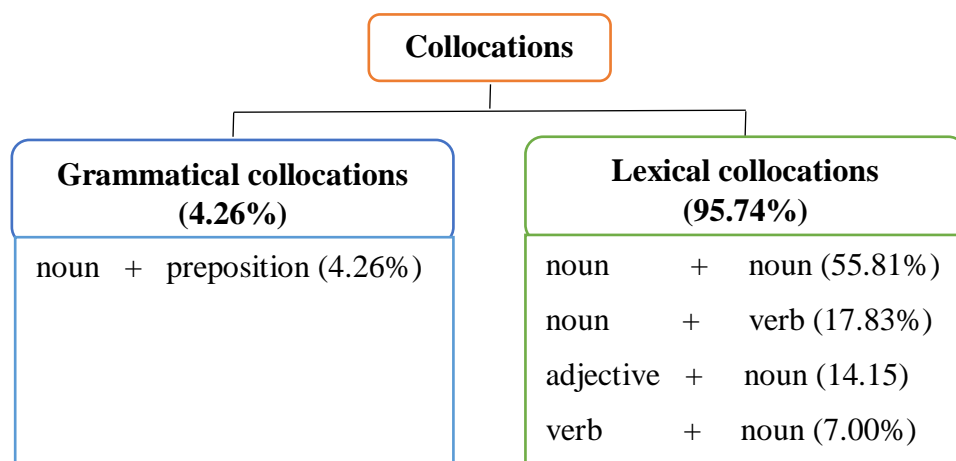


Figure 5.6 The results of collocations

5.2.1 Grammatical collocations

The results showed that the grammatical collocation accounted for only 4.26% of the corpus. From the finding, the pattern of grammatical collocations was noun collocated with preposition ‘in’, ‘with’, and ‘from’. The prepositions ‘in’ and ‘from’ were prepositions of place which were used to express the direction of the object. The preposition ‘with’ was used to express the relationship of the word ‘zebrafish’. The result was contrastive to Buakaew’s (2015). She investigated collocation usage in food and beverage advertisements and found that grammatical collocations in forms of verb + preposition occurred the most (44.06%), followed by preposition + noun (15.25%), and noun + preposition (13.56%). The result of grammatical collocations in the present study was less than Buakaew’s because the source of corpus and the genre of writing were different. As previously mentioned, grammatical collocations especially phrasal verb (verb + preposition) were generally used in spoken text or informal language such as advertisements (Buakaew, 2015), so they rarely occurred in this corpus which compiled review articles. Furthermore, the framework of grammatical collocation adopted for this study referred to noun, verb, or adjective following by preposition (Benson, 1985). It means that only one pattern, which is *noun + preposition*, can collocate as a grammatical collocation because the

target word (*zebrafish*) was a noun. Overall, owing to the limitations of source of corpus and restriction of investigated word, grammatical collocations were few in number.

5.2.2 Lexical collocations

Almost all the collocations found in the present corpus were lexical collocation (95.74%). The word ‘zebrafish’ co-occurred with noun, verb, and adjective. Consequently, there were four patterns found such as *noun + noun*, *noun + verb*, *adjective + noun*, and *verb + noun*. Three patterns of *noun + verb*, *adjective + noun*, and *verb + noun* were consistent with Benson (1985). However, this present study also found different patterns of lexical collocation such as *noun + noun*. Lexical collocations were found in various patterns because the target word (*zebrafish*) was a noun, which had more chance to collocate with other parts of speech such as noun, verb, and adjective. In addition, the most common patterns in the laboratory animal review articles was in the form of *noun + noun* which accounted for more than a half of the corpus (55%). Similar to Ackermann and Chen’s academic collocation list (2013), noun combinations were most found in the corpus (74.3%). This co-occurrence of one noun collocated with another noun made compound nouns, and thus the compound noun can either create a new word or describe the primary noun. In other words, the noun combinations were made up of two nouns, and those two nouns contained the own core meanings to modify each other. When the same type of lexical or part of speech joined together, it means that the two lexicals were equally treated. Unlike an adjective that uses to modify a noun, the core meaning of the message is a head noun. Thus, the reader would understand the main message even if the adjective was omitted. In addition, new words are likely to occur in scientific field by using a compound noun because some words cannot add morpheme or suffix to change the form to an adjective. For example, there is no an adjective for for the word ‘zebrafish’; therefore, the writer uses the noun ‘zebrafish’ as an adjective to modify the head noun.

In conclusion, this laboratory animal corpus comprised of a pattern of grammatical collocations and four patterns of lexical collocations. As the word collocations found in the study were collected from the authentic texts, they were used generally in the field of laboratory animal research.

5.3 Communicative Function

The research question 4 was, “How was the highest-frequency content word in OWL used for purposive functions in the laboratory animal review articles?” The results revealed that the purposive statements were found in the introduction. They were similar to the findings of discourse analysis by Trimble (1985), who claimed that the objective of the discourse was usually found in the introductory section of the articles. Moreover, research genre analysis by Swales (2004) proposed a typical model of research articles with four main sections: introduction, method, results, and discussion. He mentioned that the introduction was a section that stated the writer’s purpose or article’s objective. Therefore, to find the purposive functions in the research articles, the researcher focused on the introduction section. From the findings, the purposive statements were categorized into two patterns.

Pattern 1

Subject	Main verb	Phrase / Clause
This (present) review	highlight	
This studies	focuses on	noun phrase or
This article	address	noun clause
	outline	

The results showed that five purposive sentences were written following the pattern 1. The sentences were stated by using main verb, which expressed the purpose, scope, and direction of the paper. According to Bailey (2011) and Swales and Feak (2012), the statement of purpose can be written by using present simple tense. This kind of statement called the orientation toward the report itself, which refers to the thesis that communicates the information about the research. Moreover, the articles investigated in the present study were review articles. The writer aimed to emphasize the objective of the writing more than the experiments of the past. Therefore, the purposive sentences were written in present simple tenses in order to show his intention of writing.

Pattern 2

Subject (signal words of purposes)			Verb to be	to + V.inf	Noun phrase
The goal (s)	of	this review	is	to provide	Noun details of objective(s)

(Soranastaporn, 2013, p.122)

Another purposive pattern found in the present study used ‘to-infinitive’ to state the purpose. This pattern was similar to the research writing textbooks (Pyrzczak & Bruce, 2000; Swales & Feak, 2012; McMillan, 2012; Soranastaporn, 2013) stated that a purposive sentence can be written by using signal words of purpose and ‘to-infinitive’. The signal words of purpose such as *purpose*, *goal*, *aim*, *objective*, and *reason* while the examples of verb expressing the purpose of study were *investigate*, *explore*, *discover*, *find*, *determine*, and *develop* (Soranastaporn, 2013). The result showed that the word ‘goal’ was the signal word of purpose, and the word ‘provide’ was the verb expressing the purpose of the research. Generally, ‘to-infinitive’ can be used to express the purpose or objective of the research in the sentence. From the result, it showed that the purposive statement explicitly told by using not only ‘to-infinitive’ but also, using the signal word of purpose ‘goal’. Importantly, the result of purposive statement was also emphasized by using verb to be ‘is’. Rather than using ‘provide’ as a main verb like the first pattern mentioned above, the sentence was written by adding ‘verb to be’ to intensify the purpose of the study. In short, this pattern of the purposive function was similar to Swales and Feak (2012) and Soranastaporn (2013) which was used to emphasize the purpose of the study.

In conclusion, the results and patterns found in this corpus-based study were according to the typical model. The purpose of the study was generally stated at the end of introductory section. Furthermore, the sentence of purpose can be written by using either ‘to-infinitive’ or main verbs that directly express the objective of the study.

5.4 Implications of the Study

The findings of word lists and high-frequency words will contribute to teaching pedagogy. The word lists of GSL, AWL, and OWL can be benefit for ESP courses. As the word lists were collected from the ILAR journal, they would be advantageous to scientific students. Therefore, teachers can select the high-frequency words for the three lists, especially OWL to teach veterinary and scientific students, because this OWL is the word list which commonly occurs in the zoology and biochemistry field and matches with veterinary and scientific students' study field. Moreover, the teaching words or vocabulary should be demonstrated for their actual uses. To put in another way, the concordance lines of words should be exemplified for students. Word collocations also help students form sentences, naturally (Nation, 2001). Finally, the findings of communicative function, which is used to state the purpose or objective, are beneficial for students and researchers who need to write an English article. They can use the patterns found as a guideline to write the purpose of the study. In brief, word lists in this corpus could facilitate scientific teaching and learning especially in the field of laboratory animals.

CHAPTER VI

CONCLUSION

The use of English as a tool for international communication has been increasing, nowadays. Therefore, the demand for English plays a crucial role and affect all subject disciplines especially scientific fields because the majority of well-known scientific research publishers accept only English papers. It is important that students need to learn English which relates to their particular area such as English for Science and Technology (EST). In order to design an EST course, a corpus was employed because corpus is a collection of written and spoken texts that represent the real language used in a specific area.

6.1 Conclusions

This corpus-based study was conducted to find the content words of the General Service List (GSL), the Academic Word List (AWL), and the Outside Word List (OWL) used in laboratory animal review articles. Moreover, the most frequent words in OWL were investigated for collocations and purposive functions. This study explored the collocations in terms of grammatical and lexical collocations.

The source of the corpus was collected from 100 review articles of the ILAR journal between 2010 and 2014 by stratified random sampling. The present corpus included 555,526 running words that were used in laboratory animal review articles.

All 555,526 running words were computed by the AntConc program in order to be categorized into word types and make a word list from laboratory animal review articles. The results from AntConc showed that the word list of the present corpus composed of 21,609 word types out of 555,526 running words. Most of the laboratory animal corpus were content words (86%) while function words were accounted for only 14%. The purpose of the present study was to collect only content words because they contain the core meanings of the message. Moreover, language can be changed, especially content words which can be formed and adapted to new meanings (Fries, 1952). Therefore, content words can increase students' lexical knowledge. The results of content words are summarized as follows:

Research question 1: What were the content words of GSL, AWL, and OWL used in the laboratory animal review articles?

The content words found in the present study were classified into three groups: GSL, AWL, and OWL. The revised versions of GSL and AWL created by Browne, Culligan, and Phillips (2013) were used as reference word lists to separate word into the three groups. The results revealed that GSL accounted for 27.4% of the corpus. The high-frequency content words in GSL were such as *studies, animal, human, research, disease, model, cell, species, genes, dogs, health, behavior, cancer, brain, stress, mechanisms, bone, and mice*. West (1953) and Nation (2000) stated that the GSL or the 2000 most frequent words normally covered 80% of the corpus. However, the coverage of GSL in this study was much less than West's and Nation's because function words, which occurred with high frequency, were excluded. The present study investigated only content words.

The second word group was AWL. The coverage of AWL was the least proportion in the corpus (6.73%). AWL found in this corpus namely *behavioral, induced, tumor, monkeys, infected, chronic, offspring, molecular, immune, dose, transmission, cues, impact, neurons, acute, traits, consumption, and acid*. According to Coxhead (2000), AWL coverage generally covered 9 to 10% of the whole corpus. The results of AWL was less than Coxhead's since the present corpus, which aimed to investigate only in a specialized field (laboratory animal), was one-seventh of and much smaller than Coxhead's corpus.

In addition, the third word group was OWL which had the primary focus words in the present study because OWL had the specialized or technical words occurring in a specific discipline. The result indicated that OWL accounted for 65.8% of the corpus, which was the most coverage of this corpus. The words in OWL were highly used in laboratory animal's discipline; for example, *zebrafish*, *methylation*, *epigenetic*, *cocaine*, *genome*, *canine*, *nicotine*, *rodents*, *histone*, *neuropathy*, *veterinary*, *macaques*, *rodent*, and *swine*.

Research question 2: What were the high-frequency content words in OWL used in the laboratory animal review articles?

The OWL included 12,220 words which were highly related to the three sub-fields of sciences: animal science and zoology, biochemistry, genetics and molecular biology. The high-frequency words were determined by frequency and quartile. The OWL was ranked from the highest to lowest frequency, and then quartile was used to divide the words into four groups. The results indicated that the 3,055 words, which belonged to the first quartile, were high-frequency words. Moreover, the line graph showed that the top-five frequent words in OWL sharply declined. This indicated that the five words were used most prominently in the field of laboratory animals. The top-five frequent words in OWL were *zebrafish*, *methylation*, *epigenetic*, *cocaine*, and *genome*.

Research question 3: What collocations of the highest-frequency content word in OWL were used in laboratory animal review articles?

The word 'zebrafish' occurred with the highest frequency in OWL, and therefore, was analyzed for grammatical and lexical collocations. The grammatical collocation found in the study was noun collocated with preposition 'in', 'with', 'to', and 'from'. The grammatical collocations found in the present study accounted for 5% of the corpus while the other collocations were lexical collocations (95%). There were five patterns of lexical collocations: noun + noun, noun + verb, adjective + noun, and verb + noun. The examples of lexical collocations were *zebrafish model*, *zebrafish research*, *zebrafish facilities*, *zebrafish colonies*, *laboratory zebrafish*, *zebrafish*

embryos, zebrafish husbandry, zebrafish lines, zebrafish reproduction, zebrafish breeding, adult zebrafish, infected zebrafish, and larval zebrafish.

Research question 4: What were the patterns used to state the purposes in the laboratory animal review articles?

The statement used to mention the purpose of the study tended to have the following two patterns. For the first pattern, the writer used main verb to state the purpose of the articles. The main verbs such as *highlights, focuses on, addresses, and outlines* were written in present simple in order to show writer's intention of writing. Another pattern was the to-infinitive. The sentence was written in the present simple tense followed by the to-infinitive in order to indicate the details of the objective which can be explained in noun phrase. Overall, purposive sentences can be written by using main verbs that express the purpose of the study, or stated by using the to-infinitive.

In conclusion, the present study represented a lexical profile, collocations, and purposive functions used in laboratory animal review articles. This is an important contribution to the field of laboratory animals since no study has attempted to compile words in this field before. Hence, this corpus-based study has shed light on the vocabulary and the uses of words in the study of laboratory animals. In addition, the new revisions of GSL and AWL (Browne, Culligan, and Phillips, 2013) were used as the reference word lists to categorize the words into three groups: GSL, AWL, and OWL. These updated versions of GSL and OWL provided more 5% of text coverage, and thus, they gave results that are more reliable.

6.2 Recommendations for Further Studies

Due to the limitations of the present corpus-based study, the recommendations for further studies were suggested to obtain a deep understanding about vocabulary and the uses of the words.

1. Size of corpus and similar sub fields should be expanded for investigation. The future study can increase the number of running words in order to

find more common used words in the discipline. Moreover, similar sub fields such as biology can be examined to compare the results.

2. Future studies should collect the data from different sources. Apart from review articles, analysis should use textbooks and related materials used in learning .

3. The present study analyzed the collocations and purposive functions of the highest-frequency word. Further studies can investigate the collocations and purposive functions of the other high-frequency words in the corpus.

4. Apart from purposive functions, other communicative functions such as detailing method and making a conclusion can be carried out.

5. Future studies can explore and elucidate the features of the review articles such as genre analysis.

6. The data and findings of the conducted research can be used to develop the material. The word lists found in the study can be further experimented in the real classroom in order to measure the achievement of the students' lexical knowledge.

BIBLIOGRAPHY

- Ackermann, K., & Chen, Y. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *English For Academic Purposes*, 12(4), 235-247.
- Alderson, C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum International Publishing.
- Anthony, L. (2014). AntConc (version 3.4.3) [software]. Available from <http://www.laurenceanthony.net>
- Association of Southeast Asian Nations, (2008). Association of Southeast Asian Nations: Annual report 2007-2008. Jakarta, Indonesia: ASEAN Secretariat.
- Aston, G., Bernardini, S., & Stewart, D. (2009). *Corpora and language learners*. Amsterdam: John Benjamins.
- Bailey, S. (2011). *Academic Writing: A handbook for international students*. New York: Routledge.
- Baker, W. (2012). English as a lingua franca in Thailand: Characterizations and implications. *Englishes in Practice*, 1(1), 18-27.
- Barfield, A., & Gyllstad, H. (2009). Introduction: Researching L2 collocation knowledge and development. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp.1-20). Basingstoke: Palgrave Macmillan.
- Barlow, M. (2002). *MonoConc Pro (Version 2.2)*. Houston: Athelstn.
- Basturkmen, H. (2010). *Developing courses in English for specific purposes*. Basingstoke: Palgrave Macmillan.
- Belcher, D. (2006). English for specific purposes: Teaching to perceived needs and imagined futures in worlds of work, study, and everyday life. *TESOL Quarterly*, 40, 133-156.
- Belcher, D., Johns, A., & Paltridge, B. (2011). *New directions in English for Specific Purposes Research*. Michigan: The University of Michigan Press.

- Benson, M. (1985). *Collocations and idioms*. In R. Ison (Ed.), *Dictionaries, lexicography and language learning*. British Council: Pergamon Press.
- Bhatia, V., & Bremner, S. (2012). English for Business Communication. *Language Teaching*, 45(4), 410-445.
- Biber, D., Conrad, S., & Reppen, R. (2000). *Corpus linguistics*. Cambridge: Cambridge University Press.
- Boswood, T. (1997). *New ways of using computers in language teaching*. Alexandria: TESOL.
- Boulton, A., Thomas, S., & Jolivet, E. (2012). *Corpus-informed research and learning in ESP*. Amsterdam: John Benjamins Pub. Co.
- Bowker, L. & Pearson, J. (2002). *Working with Specialized Language: A practical guide to using corpora*. New York: Routledge.
- Breen, M. (1995). Authenticity in the Language Classroom. *Applied Linguistics*, 6, 60-70.
- Brezina, V., & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36, 1-22.
- Bromley, K. (2007). Nine things every teacher should know about words and vocabulary instruction. *Journal of Adolescent & Adult Literacy*, 50 (1), 528-536.
- Broughton, M. (1999). The important of the teaching of English in Thai educational system. *Studies in Language and Language Teaching*, 7(1), 1-10.
- Browne, C., Culligan, B., & Phillips, J. (2013). The New General Service List: Celebrating 60 years of vocabulary learning. *The Language Teacher*, 37 (4), 13-16.
- Buakaew, R. (2015). A study of collocation usage in food and beverage advertisements. *Panyapiwat Journal*, 7 (2), 232-244.
- Cabré, T. (1999). *Terminology – Theory, methods and applications*. Amsterdam: John Benjamins Publishing Company.
- Carstairs-McCarthy, A. (2002). *An introduction to English morphology*. Edinburgh: Edinburgh University Press.

- Celce-Murcia, M. (1980). Contextual analysis of English: Application to TESL. In D. Larsen-Freeman (Ed.), *Discourse analysis in second language research* (pp. 41-55). Rowley, MA: Newbury House.
- Celce-Murcia, M., & Olshtain, E. (2000). *Discourse and context in language teaching*. Cambridge: Cambridge University Press.
- Chan, A. (2013). Progress and prospects for genetic modification of nonhuman primate models in biomedical research. *ILAR Journal*, 54(2), 211-223.
- Charles, M. (2012). Proper vocabulary and juicy collocations. EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31 (2), 93-102.
- Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes*, 26, 502-514.
- Cheng, L. (2012). The power of English and the power of Asia: English as lingua franca and in bilingual and multilingual education. *Journal of Multilingual and Multicultural Development*, 33(4), 327-330.
- Childs, P., & O'Farrell, F. (2003) Learning science through English: An Investigation of the vocabulary skills of native and non-native English speakers in international school. *Chem. Euc. Res. Pract*, 4(3), 233-247.
- Chujo, K., & Utiyama, M. (2006). Selecting level-specific specialized vocabulary using statistical measures. *System*, 34(2), 255-269.
- Chung, T., & Nation, P. (2004). Identifying technical vocabulary. *System*, 32(2), 251-263.
- Cobb, T., & Horst, M. (2001). Reading academic English: Carrying learners across the lexical threshold. In J. Flowerdew & M. Peacock (Eds.), *The English for academic purposes curriculum* (pp. 315-329). London: Cambridge University Press.
- Cohn, M. (2010). *Alternatives to animal testing gaining ground: Researchers, regulators develop new systems for experiments*. Baltimore: The Baltimore Sun.

- Conzett, J. (2000). Integrating collocation into a reading and writing course. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 70-87). UK: Language Teaching Publications.
- Cosby, A. (1987). Does a scientist need foreign language?. *Foreign Language Annals*, 20(2), 181-183.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead, A., & Hirsh, D. (2007). A pilot science word list for EAP. *Revue Française de linguistique appliquée* xii, 2, 65-78.
- Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp.252-267). Cambridge University Press.
- Crim, M., & Riley, L. (2012). Viral diseases in zebrafish: What is known and unknown. *ILAR Journal*, 53(2), 135-143.
- Crystal, D. (2003). *English as a global language*. Germany: Ernst Klett Sprachen.
- Diniz, L. (2005). Comparative Review: TextSTAT 2.5, AntConc 3.0 and Compleat Lexical Tutor 4.0. *Language Learning and Technology*, 9(3), 22-24.
- Donley, K., & Peppen, R. (2001). Using corpus tools to highlight academic vocabulary in sustained content language teaching. *TESOL Journal*, 10(2), 7-12.
- Dovey, T. (2006). What purposes, specifically? Re-thinking purposes and specificity in the context of the “new vocationalism”. *English for Specific Purposes*, 25 (4), 387-402.
- Dudley-Evans, T., & John, M. (1998). *Developments in ESP*. UK: Cambridge University Press.
- Durrant, P. (2014). Corpus frequency and second language learners’ knowledge of collocations: A meta-analysis. *International Journal Of Corpus Linguistics*, 19(4), 443-477.
- Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188.
- Fernández, T. & Faber, P. (2011). The representation of multidimensionality in a bilingualized English-Spanish thesaurus for learners in architecture and building construction. *International Journal of Lexicography*, 24(2), 198-225.

- Flowerdew, J. & Peacock, M. (2001). *Research perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press.
- Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, 21(2), 231–244.
- Fries, C. (1952). *The structure of English*. New York: Harcourt Brace.
- Froehlich, H. (2015). Corpus analysis with Antconc programming historian. Retrieved from <http://programminghistorian.org/lessons/corpus-analysis-with-antconc.html>
- 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, 54-58.
- Gotti, M., & Giannoni, D. (2014). *Corpus analysis for descriptive and pedagogical purposes: ESP perspectives*. Bern: Peter Lang.
- Grabowski, L. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English For Specific Purposes*, 38, 23-33.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. Cowie (Ed.), *Phraseology: theory, analysis and applications* (145- 160). Oxford: Oxford University Press.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp.145-160). London: Oxford University Press.
- Halliday, M., & Webster, J. (2002). *Linguistic studies of text and discourse*. London: Continuum.
- Halliday, M., McIntosh, A., & Stevens, P. (1965). *The linguistic sciences and language teaching*. Bloomington: Indiana University Press.
- Harmer, J. (1994). *The practice of english language teaching*. London: Longman.
- Hasman, A. (2000). The role of English in the 21st century. *the English teaching Forum*, 38 (1), 2-5.

- Hayes, D. (2010). Language learning, teaching and educational reform in rural Thailand: An English teacher's perspective. *Asia Pacific Journal of Education*, 30(3), 305-319.
- Hewings, M & Hewings, A. (2002). "It is interesting to note that..." : A comparative study of anticipatory 'it' in student and published writing. *English for Specific Purposes*, 21, 367-383.
- Hill, J. (2000). Revising priorities: from grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching Collocation: Further Development in the Lexical Approach* (pp. 47-69). Hove: Language Teaching Publications.
- Honna, N. (2012). English as a multicultural language for international communication in Asia. In C. B. Della, J. Scott, & C. Hinton (Eds.), *Languages in a global world: Learning for better cultural understanding* (pp. 247-256). Paris: OECD Publishing.
- Hornby, A.S. (2005). *Oxford advanced learner's dictionary of current English*. London: Oxford University Press.
- Hu, C., & Li, Y. (2015). Discourse connectives in L1 and L2 argumentative writing. *Higher Education Studies*, 5(4), 34-43.
- Hunston, S. (2011). *Corpus approaches to evaluation*. New York: Routledge.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centered approach*. New York: Cambridge University Press.
- Hyland, K. (2006) *English for academic purposes: An advanced resource book*. London: Routledge.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English For Specific Purposes*, 27(1), 4-21.
- Intaraprasit, N. (2014). *Investigation of Collocations Occurrence in English Novels*. In 6th Rajamangala University of Technology International Conference (pp. 1110-1116). Bangkok, Thailand: Rajamangala University of Technology.
- Internet World Stats. (2015). *Top ten languages in the Internet in millions of user – November 2015*. Retrieved from <http://www.internetworldstats.com>

- Jitendra, A., Edwards, L., Sacks, G., & Jacobson, L. (2004). What research says about vocabulary instruction for students with learning disabilities. *Exceptional Children, 70*, 299-322.
- Jordan, R. (2002). The growth of EAP in Britain. *English for Academic Purposes, 1*, 69-78.
- Kaewphanngam, C. (2002). A corpus analysis of psychology texts as a basis for the development of teaching materials in English for academic purposes (Unpublished master dissertation). Mahidol University, Thailand.
- Kent, M., Harper, C., & Wolf, J. (2012). Documented and potential research impacts of subclinical diseases in zebrafish. *ILAR Journal, 53*(2), 126-134.
- Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.
- Kirkpatrick, A. (2012). English as an international language in Asia: Implications for language education. In A. Kirkpatrick & A. Sussex (Eds.), *English as an international language in Asia: Implications for language education* (pp. 29-44). London: Springer.
- Kwary, D. (2011). A hybrid method for determining technical vocabulary. *System, 39*, 175-185.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines*. Clevedon: Multilingual Matters.
- Laufer, B. (2014). A problem in vocabulary learning—synophones. *ELT Journal, 34*, 294-300.
- Laufer, B., & Nation, P. (1995). Vocabulary size & use: Lexical richness in L2 written productions. *Applied Linguistics, 16* (3), 307-322.
- Lawrence, C., & Mason, T. (2012). Zebrafish housing systems: A review of basic operating principles and considerations for design and functionality. *ILAR Journal, 53*(2), 179-191.
- Lewis, M. (2000). *Teaching collocation: Further development in the lexical approach*. Hove: Language Teaching Publications.

- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31, 25-35.
- Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English For Specific Purposes*, 39, 1-11.
- Low, E. & Hashim, A. (2012) *English in Southeast Asia: Features, policy and language in use*. Amsterdam: John Benjamins Publishing.
- Lubis, S. (2013). Collocation as source of translation unacceptability: Indonesian students' experiences. *International Journal of English Linguistics*, 3(5).
- Maher, A. (2008). Acquiring or enhancing a translation specialism: the monolingual corpus-guided approach. *The Journal of Specialised Translation*, 4(10), 12-24.
- Mahlberg, M. (2007). Corpus stylistics: bridging the gap between linguistic and literary studies. In Hoey, M., Mahlberg, M., Stubbs, M., Teubert, W. Text (Ed.), *Discourse and corpora*. London: Continuum.
- Martin-Chang, S., & Gould, O. (2008). Revisiting print exposure: Exploring differential links to vocabulary, comprehension and reading rate. *Journal of Research in Reading*, 31, 273–284.
- Martinez, I., Beck, S. & Panza, C. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, 183-198.
- Marukatat, S. (2012, January 8). Poor English skills could leave Thais out in the cold. Bangkok post. Retrieved from <http://www.bangkokpost.com>
- Matthews, M., & Varga, Z. (2012). Anesthesia and euthanasia in zebrafish. *ILAR Journal*, 53(2), 192-204.
- Mavor, S., & Trayner, B. (2001). Aligning genre and practice with learning in higher education: an interdisciplinary perspective for course design and teaching. *English for Specific Purposes*, 20(4), 345-366.
- McCarthy, A. (2002). *An introduction to English morphology*. Edinburgh: Edinburgh University Press.
- McEnery, T., & Wilson, A. (2010). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.

- McMillan, J. (2012). *Educational research fundamentals for the consumer* (6th ed.). New York: Longman.
- Ministry of Education. (2001). *Basic education curriculum 2001*. Bangkok: Kurusapa Ladprao Publishing.
- Montgomery, S., & Crystal, D. (2013). *Does science need a global language?: English and the future of research*. Chicago: University Press.
- Muñoz, V. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English For Specific Purposes*, 39, 26-44.
- Murphy, M. (2010). *Lexical meaning*. Cambridge: Cambridge University Press.
- Nasiadka, A., & Clark, M. (2012). Zebrafish breeding in the laboratory environment. *ILAR Journal*, 53(2)(7), 161-168.
- Nation, I. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- National Academy of Sciences. (2004). Guide for care laboratory animals. Retrieved from <https://grants.nih.gov/grants/olaw/Guide-for-the-Care-and-use-of-laboratory-animals.pdf>
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence. *International Journal Of Corpus Linguistics*, 10(2), 266-270.
- Noguchi, J. (2006). The science review article: An opportune genre in the construction of science. *Linguistic Insights*, 11(1), 16-30.
- Northcott, J., & Brown, G. (2006). Legal translator training: Partnership between teachers of English for legal purposes and legal specialists. *English for Specific Purposes*, 25, 358-375.
- Orr, T. (1999). ESP for Japanese universities: A guide for intelligent reform. *The Language Teacher*, 22 (11), 19-31.
- Otte, J. (2006). Real language to real people: A descriptive and exploratory case study of the outcomes of aural authentic texts on the listening comprehension of adult ESL students enrolled in an advanced ESL listening course. *Dissertation Abstracts International*, 67(4), 12-46.

- Paltridge, B. (2002). Thesis and dissertation writing: An examination of published advice and actual practice. *English for Specific Purposes*, 21(3), 125–143.
- Paltridge, B., & Starfield, S. (2014). Research in English for specific purposes. In Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning (Vol. 2)*, (pp. 106-121). New York: Routledge.
- Paquot, M. (2010). *Academic vocabulary in learner writing: From extraction to analysis (corpus and discourse)*. London: Continuum.
- Parkinson, J. (2000). Acquiring scientific literacy through content and genre. *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 (Unpublished master dissertation). Mahidol University, Thailand.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J. Richards & R. Schmidt. *Language and communication* (pp. 191-226). London: Longman.
- Peacock, M. (2001). The effect of authentic materials on the motivation of EFL learners. *ELT Journal*, 51(2), 144-156.
- Pearson, J. (1998). *Terms in context – studies in corpus linguistics*. Amsterdam: John Benjamins Publishing Company.
- Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, 10(2), 231-250.
- Prawanpath, J. (1984). A study of general vocabulary knowledge of the pre-medical science students at Mahidol University (Unpublished master dissertation). Mahidol University, Thailand.
- Pyczak, F., & Bruce, R. (2000). *Writing empirical research reports*. Los Angeles, CA: Pyczak Pub.
- Qian, D. (2005). Measuring lexical richness in business English writing: A study of Chinese learners. *The Hong Kong Linguist: Jubilee Version*, 25, 36-42.
- Richards, J., & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics*. Harlow: Longman.

- Ridley, D. (2008). *The literature review – a step-by-step guide for students*. London: Sage Publications.
- Römer, U., & Wulff, S. (2010). Applying corpus methods to writing research: Explorations of MICUSP. *Journal of Writing Research*, 2(2), 99–127.
- Rooth, O., & Saarela, J. (2007). Native language and immigrant labor market outcomes: An alternative approach to measuring. *International Migration & Integration*, 8(1), 207-221.
- Samraj, B. (2005). An exploration of a genre set: research article abstracts and introductions in two disciplines. *English for Specific Purposes*, 24, 141-156.
- Sanders, J., Watral, V., & Kent, M. (2012). Microsporidiosis in zebrafish research facilities. *ILAR Journal*, 53(2), 106-113.
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schoonen, R., Gelderen, A., Stoel, R., Hulstijn, J., & Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning*, 61, 31–79.
- Scott, M. (1996). *WordSmith Tools*, Oxford: Oxford University Press.
- Sharifian, F. (2009). *English as an international language: Perspectives and pedagogical issues*. Bristol, UK: Multilingual Matters.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Soranastaporn, S. (2013). *Effective reading & writing english texts*. Bangkok: Wang Aksorn Press.
- Spitsbergen, J., Buhler, D., & Peterson, T. (2012). Neoplasia and neoplasm-associated lesions in laboratory colonies of zebrafish emphasizing key influences of diet and aquaculture system design. *ILAR Journal*, 53(2), 114-125.
- Staples, S. (2015). Examining the linguistic needs of internationally educated nurses: A corpus-based study of lexico-grammatical features in nurse–patient interactions. *English For Specific Purposes*, 37, 122-136.
- Strevens, A., Butler, A., & Castellon, M. (2000). *Academic language and content assessment: Measuring the progress of ELLS*. Los Angeles: University of

California, National Center for Research on Evaluation, Standards and Student Testing (CRESST).

- Strevens, P. (1988). ESP after twenty years: A re-appraisal. in Tickoo, M. (Ed.), *ESP: State of the Art* (pp. 1-13). Singapore: SEAMEO Regional Centre.
- Stubbs, M. (2001). Text, corpora, and problems of interpretation: A response to Windowson. *Applied Linguistics*, 22, 149-172.
- Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP?: A corpus based case study. *RELC Journal*, 25, 34-50.
- Swales, J. M. (1981). *Aspects of article Introductions*. Birmingham, UK: University of Aston, Language Studies Unit.
- Swales, J. (1990). *Genre analysis*. Cambridge: Cambridge University Press.
- Swales, J. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Ed.), *Academic discourse* (pp. 150–164). Harlow: Longman.
- Swales, J. (2005). *Genre analysis*. UK: Cambridge University Press.
- Swales, J. & Feak, C. (2012). *Academic writing for graduate students* (3rd ed.). Ann Arbor, Mich.: University of Michigan Press.
- Tarone, E. (1981). Some thoughts on the notion of communication strategy. *TESOL Quarterly*, 15(3), 285-295.
- Teubert, W. (1996). Why corpus linguistics? [Online]. Retrieved from <http://www.lb.u-tokai.ac.jp/tono/resources.html>.
- The Royal Society. (2004). The use of non-human animals in research: a guide for scientists. Retrieved from <https://royalsociety.org>
- Thorndike, E., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College.
- Tribble, C. (2012). Teaching and Language Corpora: Quo Vadis?. In L. Anthony. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research* 30(2), 141-161.
- Tribble, C., & Jones, G. (1990). *Concordance in the classroom: a resource book for teachers*. England: Longman.
- Trimble, L. (1985). *English for science and technology*. Cambridge: Cambridge University Press.

- Ulrich, A. (1996). The European union status change of English during the last fifty years. In J. Fishman, et.al. (Eds.), *Post-imperial English: Status change in former British and American colonies, 1940-1990* (p.241-267). Berlin: Mouton de Gruyter.
- Underwood, M. (1989). *Teaching listening*. London: Longman.
- User language preferences online. (2011). *User language preferences online: Analytical report*. Hungary: The Gallup Organization.
- 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.
- Vanderweide, T., Rees-Miller, J., Aronoff, M., & O'Grady, W. (2010). *Contemporary linguistics*. Boston: Bedford, St. Martin's.
- Vongpumivitch, V. Huang, J & Chang, Y. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28, 33-41.
- Walker, C. (2011). How a corpus-based study of the factors which influence collocation can help in the teaching of business English. *English For Specific Purposes*, 30(2), 101-112.
- Wallace, M. & Wray, A. (2011). *Critical reading and writing for postgraduate students*. London: Sage Publications Ltd.
- Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27(4), 442-458.
- Wang, Y. (2014). A corpus-based contrastive study on academic vocabulary of RA abstracts in applied linguistics. *AMM*, 631-632.
- Waring, R. & Nation, P. (1997) Vocabulary size, text coverage and word lists [Online]. Retrieved from <http://www.harenet.ne.jp/~waring/papers/cup.html>
- Weigle, S. (2002). *Assessing Writing*. Cambridge, England: Cambridge University Press.
- West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.
- Whipps, C., Lieggi, C., & Wagner, R. (2012). Mycobacteriosis in zebrafish colonies. *ILAR Journal*, 53(2), 95-105.
- Wilkins, D. (1972). *Linguistics in language teaching*. London: Edward Arnold.

- William, O., Archibald, J., Aronoff, M., Rees-Miller, J. (2001). *Contemporary linguistics: An introduction (4th ed.)*. Boston: Bedford.
- Wilson, C. (2012). Aspects of larval rearing. *ILAR Journal*, 53(2), 169-178.
- Wong, V. (1995). The use of authentic materials at tertiary level. *ELT Journal*, 49(4), 318-322.
- Wood, A. (2001). International scientific English: the language of research scientists. In J. Flowerdew & M. Peacock (Eds). *Research Perspectives on English for Academic Purposes* (pp. 71-83). Cambridge: Cambridge University Press.
- Wood, D. (2010). *Perspectives on formulaic language: Acquisition and communication*. London, NY: Continuum.
- Worthington, D. & Nation, P. (1996). Using texts to sequence the introduction of new vocabulary in an EAP course. *RELC Journal*, 27(2), 1-11.
- Wray, A. (2002). *Formulaic language and the lexicon*. UK: Cambridge University Press.
- Yang, M. (2015). A nursing academic word list. *English For Specific Purposes*, 37, 27-38.
- Yuan, H. (2014). A corpus-based study on the influence of L1 on EFL learners' use of prepositions. *TPLS*, 4(12).

APPENDICES

APPENDIX A**LIST OF GSL 1 IN THE LABORATORY ANIMAL CORPUS**

Rank	<i>f</i>	Word
1	1838	studies
2	1695	animal
3	1649	animals
4	1487	human
5	1433	research
6	1024	disease
7	1018	models
8	1001	model
9	975	cells
10	952	used
11	946	study
12	919	humans
13	884	effects
14	825	use
15	822	development
16	787	cell
17	782	dogs
18	781	associated
19	662	levels
20	658	water
21	644	health
22	614	increased
23	613	specific
24	604	using
25	580	behavior
26	577	risk
27	573	including
28	567	time
29	548	response
30	547	no
31	538	high
32	532	important
33	524	changes
34	524	different

Rank	<i>f</i>	Word
35	516	system
36	514	fish
37	507	treatment
38	505	number
39	499	factors
40	499	results
41	482	data
42	478	example
43	476	reported
44	475	early
45	472	based
46	472	significant
47	470	group
48	465	new
49	455	control
50	452	similar
51	421	related
52	416	several
53	416	social
54	415	potential
55	413	age
56	413	differences
57	393	drug
58	384	patients
59	381	include
60	372	long
61	372	years
62	368	common
63	367	diseases
64	366	effect
65	364	same
66	359	likely
67	357	male
68	349	less
69	347	compared

Rank	<i>f</i>	Word
70	338	female
71	333	shown
72	331	evidence
73	329	responses
74	328	activity
75	323	showed
76	322	body
77	317	known
78	316	systems
79	312	provide
80	311	higher
81	311	role
82	308	days
83	307	food
84	306	large
85	296	recent
86	294	life
87	293	function
88	291	conditions
89	290	available
90	290	level
91	290	possible
92	288	increase
93	286	groups
94	285	cases
95	285	females
96	285	small
97	279	care
98	279	low
99	278	result
100	278	type
101	277	analysis
102	271	involved
103	271	males
104	269	information
105	267	identified
106	263	behaviors
107	262	environment
108	261	growth
109	260	weight
110	259	ability
111	257	learning
112	254	populations

Rank	<i>f</i>	Word
113	252	individual
114	251	understanding
115	250	like
116	250	methods
117	250	population
118	249	described
119	247	factor
120	246	adult
121	246	approach
122	244	single
123	243	loss
124	240	day
125	240	term
126	237	experience
127	237	individuals
128	236	regions
129	235	reduced
130	233	field
131	230	findings
132	226	cause
133	226	present
134	224	review
135	222	period
136	221	given
137	220	types
138	218	test
139	217	develop
140	217	normal
141	217	work
142	215	lines
143	215	range
144	215	show
145	213	aging
146	213	lower
147	212	greater
148	212	major
149	210	various
150	206	production
151	204	limited
152	204	memory
153	203	fear
154	203	weeks
155	201	people

Rank	<i>f</i>	Word
156	200	affect
157	199	process
158	196	occur
159	195	considered
160	194	later
161	194	size
162	193	agents
163	192	change
164	192	necessary
165	192	need
166	189	areas
167	189	knowledge
168	189	recently
169	186	dog
170	184	established
171	184	least
172	181	especially
173	180	support
174	179	lead
175	178	particular
176	178	state
177	177	following
178	177	required
179	176	area
180	176	better
181	175	developed
182	175	natural
183	174	patterns
184	174	useful
185	172	seen
186	171	children
187	170	particularly
188	170	provided
189	170	testing
190	169	quality
191	168	produce
192	165	difficult
193	165	drugs
194	165	rate
195	165	states
196	164	housing
197	164	method
198	163	identify

Rank	<i>f</i>	Word
199	163	require
200	162	measures
201	162	suggest
202	161	case
203	161	part
204	160	become
205	158	rates
206	157	affected
207	157	site
208	157	target
209	156	facilities
210	155	determine
211	154	still
212	153	general
213	151	reduce
214	150	influence
215	150	lack
216	149	made
217	148	developing
218	148	place
219	147	against
220	147	occurring
221	146	certain
222	145	access
223	145	clear
224	145	key
225	144	variety
226	143	design
227	143	future
228	142	direct
229	142	months
230	142	positive
231	141	expressed
232	141	reports
233	141	strategies
234	139	features
235	139	occurs
236	138	death
237	137	make
238	136	now
239	135	current
240	135	requires
241	135	subjects

Rank	<i>f</i>	Word
242	134	approaches
243	134	damage
244	134	short
245	133	old
246	133	processes
247	133	sites
248	132	controls
249	132	great
250	132	tested
251	131	events
252	131	focus
253	131	public
254	130	increases
255	130	performed
256	130	value
257	130	world
258	129	generally
259	129	indicate
260	128	increasing
261	128	needed
262	128	report
263	128	way
264	127	performance
265	127	suggesting
266	126	studied
267	125	efforts
268	124	fact
269	124	source
270	123	appear
271	123	wide
272	122	involving
273	122	region
274	122	tests
275	121	contact
276	121	management
277	121	set
278	120	caused
279	120	concerns
280	120	local
281	120	standard
282	120	suggests
283	119	order
284	118	challenge

Rank	<i>f</i>	Word
285	118	defined
286	118	play
287	118	provides
288	118	resulting
289	117	clearly
290	116	point
291	116	treatments
292	115	extended
293	115	hours
294	115	interest
295	114	agent
296	114	complete
297	114	noted
298	113	best
299	113	designed
300	113	resulted
301	112	light
302	112	treated
303	111	end
304	111	facility
305	111	larger
306	111	resources
307	110	degree
308	110	measured
309	110	numbers
310	110	training
311	107	culture
312	107	stages
313	106	central
314	106	followed
315	106	working
316	105	around
317	105	article
318	105	course
319	105	stage
320	104	determined
321	104	difference
322	104	line
323	103	application
324	103	improve
325	103	produced
326	102	activities
327	102	allow

Rank	<i>f</i>	Word
328	102	despite
329	102	issues
330	102	national
331	102	once
332	102	pattern
333	102	proposed
334	102	relationship
335	101	according
336	101	led
337	101	questions
338	100	aged
339	100	members
340	100	signs
341	99	basis
342	99	daily
343	99	mean
344	98	appears
345	98	good
346	98	help
347	98	included
348	98	suggested
349	98	table
350	97	able
351	97	choice
352	97	concern
353	97	controlled
354	97	means
355	97	total
356	96	act
357	96	causes
358	96	remains
359	95	action
360	95	attention
361	95	longer
362	94	center
363	94	generation
364	94	strong
365	94	structure
366	93	hand
367	92	benefit
368	92	challenges
369	92	free
370	92	imaging

Rank	<i>f</i>	Word
371	92	indeed
372	92	marks
373	92	remain
374	92	serve
375	92	whole
376	91	housed
377	91	involves
378	90	examples
379	90	maintain
380	90	successful
381	89	linked
382	89	taken
383	88	amount
384	88	last
385	88	leads
386	88	nature
387	88	project
388	88	properties
389	88	risks
390	88	staff
391	88	task
392	87	cross
393	86	analyses
394	86	expected
395	86	targeted
396	86	targets
397	85	finding
398	85	mothers
399	85	year
400	84	basic
401	84	improved
402	84	involve
403	84	values
404	84	young
405	83	average
406	83	condition
407	83	consider
408	83	poor
409	83	power
410	83	release
411	83	subject
412	82	applied
413	82	avoid

Rank	<i>f</i>	Word
414	82	discussed
415	82	learned
416	82	live
417	82	practice
418	82	take
419	81	advantage
420	81	plan
421	80	called
422	80	documented
423	80	functions
424	80	goal
425	80	providing
426	79	community
427	79	older
428	79	possibility
429	79	question
430	78	account
431	78	experiences
432	78	problems
433	78	times
434	77	along
435	77	leading
436	77	program
437	77	received
438	76	below
439	76	mother
440	76	strategy
441	76	week
442	75	placed
443	75	practices
444	75	represent
445	74	allows
446	74	main
447	74	measure
448	74	women
449	73	indicated
450	73	problem
451	72	affects
452	72	authors
453	72	conditioned
454	72	conditioning
455	72	earlier
456	72	establish

Rank	<i>f</i>	Word
457	72	instead
458	72	parts
459	72	regarding
460	72	targeting
461	71	address
462	71	countries
463	71	making
464	71	needs
465	71	occurred
466	70	arm
467	70	continue
468	70	gain
469	70	history
470	70	material
471	70	periods
472	70	sources
473	70	understand
474	69	adults
475	69	focused
476	69	international
477	69	reporting
478	68	advantages
479	68	done
480	68	full
481	68	home
482	68	quite
483	68	ways
484	67	family
485	67	includes
486	67	perhaps
487	66	depending
488	66	forms
489	66	late
490	66	opportunity
491	66	previous
492	66	seems
493	66	smaller
494	65	event
495	65	reasons
496	65	separate
497	64	affecting
498	64	cost
499	64	find

Rank	<i>f</i>	Word
500	64	mentioned
501	64	minutes
502	64	programs
503	64	success
504	63	living
505	63	resource
506	63	side
507	63	structures
508	62	again
509	62	express
510	62	patient
511	62	products
512	62	start
513	61	achieved
514	61	heart
515	61	issue
516	61	maintained
517	61	organization
518	61	science
519	60	law
520	60	roles
521	60	turn
522	59	benefits
523	59	describe
524	59	discuss
525	59	growing
526	59	indicating
527	59	lasting
528	59	simple
529	59	situation
530	59	university
531	58	beyond
532	58	communication
533	58	modeling
534	58	own
535	58	reducing
536	57	close
537	57	containing
538	57	offer
539	57	pressure
540	57	ranging
541	57	see
542	57	step

Rank	<i>f</i>	Word
543	56	create
544	56	final
545	56	instance
546	56	points
547	56	reviews
548	55	back
549	55	indicates
550	55	makes
551	55	probably
552	55	reflect
553	55	reviewed
554	55	room
555	55	setting
556	55	trade
557	54	determining
558	54	outside
559	54	recognized
560	54	reference
561	54	respect
562	54	standards
563	54	studying
564	54	technology
565	53	date
566	53	open
567	53	space
568	53	supply
569	53	traditional
570	53	understood
571	52	exist
572	52	figure
573	52	movement
574	52	note
575	52	section
576	52	seeking
577	52	society
578	51	air
579	51	contain
580	51	easily
581	51	reduces
582	51	serious
583	50	directed
584	50	except
585	50	existing

Rank	<i>f</i>	Word
586	50	produces
587	50	raised
588	50	share
589	50	view
590	49	achieve
591	49	completely
592	49	maintaining
593	49	processing
594	49	programming
595	49	situations
596	48	environments
597	48	estimated
598	48	head
599	48	immediately
600	48	policy
601	48	projects
602	48	purposes
603	48	treat
604	48	white
605	47	actions
606	47	attempts
607	47	creating
608	47	energy
609	47	highest
610	47	identifying
611	47	marked
612	47	materials
613	47	search
614	47	sections
615	47	sometimes
616	47	tasks
617	47	tend
618	47	unclear
619	46	allowing
620	46	base
621	46	effort
622	46	feature
623	46	generations
624	46	interesting
625	46	simply
626	45	advanced
627	45	challenging
628	45	define

Rank	<i>f</i>	Word
629	45	fewer
630	45	follow
631	45	house
632	45	month
633	45	networks
634	45	position
635	45	presented
636	45	shows
637	45	sizes
638	45	theory
639	44	beginning
640	44	depends
641	44	detailed
642	44	learn
643	44	perform
644	44	proven
645	44	technologies
646	44	uses
647	43	continued
648	43	give
649	43	quickly
650	43	showing
651	43	themselves
652	42	advances
653	42	ages
654	42	applications
655	42	easy
656	42	influences
657	42	opportunities
658	42	reach
659	42	reason
660	42	shorter
661	41	abilities
662	41	allowed
663	41	centers
664	41	conclusions
665	41	discussion
666	41	establishing
667	41	explain
668	41	gained
669	41	managers
670	41	men
671	41	options

Rank	<i>f</i>	Word
672	41	special
673	40	began
674	40	class
675	40	fire
676	40	move
677	40	original
678	40	positions
679	40	return
680	40	screening
681	39	became
682	39	competition
683	39	difficulty
684	39	face
685	39	flies
686	39	forward
687	39	limit
688	39	network
689	39	protect
690	39	relationships
691	39	shared
692	39	subjected
693	39	supported
694	39	weights
695	38	causing
696	38	completed
697	38	costs
698	38	enough
699	38	half
700	38	plans
701	38	product
702	38	right
703	37	depend
704	37	designs
705	37	fixed
706	37	greatest
707	37	moving
708	37	post
709	37	receive
710	37	referred
711	37	rise
712	37	seconds
713	36	arms
714	36	come

Rank	<i>f</i>	Word
715	36	idea
716	36	link
717	36	matter
718	36	rights
719	36	seem
720	36	suffering
721	35	certainly
722	35	discovered
723	35	failed
724	35	fast
725	35	hospital
726	35	presents
727	35	represents
728	35	separated
729	35	supplies
730	35	supporting
731	35	team
732	34	added
733	34	carried
734	34	club
735	34	created
736	34	decisions
737	34	direction
738	34	list
739	34	remaining
740	34	settings
741	34	version
742	33	agencies
743	33	decision
744	33	economic
745	33	education
746	33	experienced
747	33	extend
748	33	hour
749	33	instances
750	33	introduced
751	33	purpose
752	33	regard
753	33	taking
754	32	appeared
755	32	balance
756	32	building
757	32	functioning

Rank	<i>f</i>	Word
758	32	goals
759	32	lost
760	32	parent
761	32	person
762	32	producing
763	32	professionals
764	32	promising
765	32	teaching
766	31	considering
767	31	controlling
768	31	develops
769	31	formed
770	31	improving
771	31	influenced
772	31	keeping
773	31	kept
774	31	sense
775	31	steps
776	30	advance
777	30	becoming
778	30	begin
779	30	believe
780	30	changing
781	30	cold
782	30	conclusion
783	30	contains
784	30	expressing
785	30	industry
786	30	organizations
787	30	rocking
788	30	service
789	30	sets
790	30	shape
791	30	trained
792	30	wall

APPENDIX B**LIST OF GSL 2 IN THE LABORATORY ANIMAL CORPUS**

Rank	<i>f</i>	Word
1	730	expression
2	728	gene
3	683	mice
4	612	stress
5	598	genes
6	563	cancer
7	552	brain
8	403	pain
9	386	environmental
10	339	mouse
11	336	colleagues
12	319	observed
13	315	mechanisms
14	305	bone
15	302	administration
16	292	complex
17	261	demonstrated
18	248	sex
19	247	blood
20	246	protein
21	243	multiple
22	234	significantly
23	233	effective
24	233	exposed
25	222	decreased
26	222	experiments
27	220	researchers
28	217	trials
29	212	diet
30	210	wild
31	206	novel
32	202	presence
33	199	proteins

Rank	<i>f</i>	Word
34	198	highly
35	189	approximately
36	188	critical
37	182	additional
38	180	contrast
39	179	procedures
40	174	host
41	171	birds
42	170	primary
43	169	sequence
44	168	interactions
45	164	binding
46	163	signaling
47	162	samples
48	160	exhibit
49	158	relatively
50	157	consistent
51	155	relevant
52	154	scientific
53	152	regulation
54	148	outcomes
55	145	importance
56	142	derived
57	139	assessment
58	139	injury
59	138	assess
60	138	context
61	138	negative
62	137	sequencing
63	137	trial
64	136	sexual
65	135	appropriate
66	135	association

Rank	<i>f</i>	Word
67	132	physical
68	129	global
69	128	phase
70	126	combination
71	126	medical
72	123	dish
73	123	status
74	121	defeat
75	119	cancers
76	119	initial
77	118	currently
78	118	decrease
79	117	directly
80	115	contribute
81	114	reduction
82	112	techniques
83	111	essential
84	111	outcome
85	110	concentrations
86	110	severe
87	107	components
88	105	sequences
89	105	specifically
90	104	temperature
91	103	prevent
92	103	procedure
93	102	capacity
94	102	characteristics
95	102	conducted
96	102	enhanced
97	102	exhibited
98	101	mechanism
99	101	relative
100	100	monitoring
101	100	previously
102	99	interaction
103	99	literature
104	99	sample
105	98	selected
106	97	consequences
107	97	medicine

Rank	<i>f</i>	Word
108	97	revealed
109	95	scientists
110	95	unique
111	95	widely
112	94	eggs
113	94	frequently
114	94	responsible
115	92	assessed
116	92	conduct
117	92	investigation
118	91	healthy
119	91	surface
120	90	selection
121	89	ensure
122	89	safety
123	87	delivery
124	87	fat
125	86	alternative
126	86	extent
127	86	tools
128	85	active
129	85	skin
130	84	closely
131	84	observations
132	83	abuse
133	83	mass
134	83	pool
135	83	protection
136	82	birth
137	82	experiment
138	82	origin
139	81	display
140	81	smoke
141	80	equipment
142	80	primarily
143	79	chemical
144	79	combined
145	79	elements
146	79	sufficient
147	79	volume
148	78	examined

Rank	<i>f</i>	Word
149	78	location
150	77	majority
151	76	decades
152	76	investigations
153	75	grade
154	74	introduction
155	74	removed
156	74	valuable
157	74	variable
158	73	emerging
159	73	increasingly
160	73	length
161	72	absence
162	72	rapidly
163	72	reward
164	71	independent
165	71	repeated
166	71	variables
167	70	comparison
168	70	concept
169	70	consideration
170	70	shock
171	69	demonstrate
172	69	internal
173	68	assessing
174	68	background
175	68	prior
176	68	scale
177	68	vary
178	67	feeding
179	67	promote
180	67	rare
181	66	cycle
182	66	smoking
183	66	stable
184	65	failure
185	65	latter
186	65	requirements
187	65	respond
188	65	typical
189	64	enhance

Rank	<i>f</i>	Word
190	64	technique
191	63	component
192	63	mixed
193	63	progress
194	62	flow
195	62	manner
196	62	spread
197	62	transfer
198	61	fed
199	60	concentration
200	60	confirmed
201	60	decline
202	60	generated
203	60	located
204	60	mapping
205	60	percentage
206	60	signal
207	59	improvement
208	58	broad
209	58	cats
210	58	institutions
211	57	collected
212	57	effectively
213	57	fully
214	57	observation
215	56	decreases
216	56	diets
217	56	emergency
218	55	considerations
219	55	generate
220	54	examine
221	54	rank
222	54	reaction
223	54	route
224	53	ideal
225	52	immediate
226	51	considerable
227	51	content
228	51	core
229	51	crucial
230	51	differ

Rank	<i>f</i>	Word
231	51	farm
232	51	investigated
233	51	largely
234	51	normally
235	50	chemicals
236	50	egg
237	50	pairs
238	50	safe
239	49	characteristic
240	49	emotional
241	49	employed
242	49	escape
243	49	left
244	49	map
245	49	transport
246	48	domestic
247	48	entire
248	48	possibly
249	47	assessments
250	47	commercial
251	47	extreme
252	47	practical
253	47	predict
254	47	presentation
255	47	recommended
256	47	signals
257	47	tool
258	47	unlike
259	46	actual
260	46	strongly
261	45	capable
262	45	horses
263	45	hosts
264	45	implications
265	45	integrated
266	45	interpretation
267	45	islands
268	45	moral
269	45	proper
270	45	shift
271	44	guide

Rank	<i>f</i>	Word
272	44	illness
273	44	institution
274	44	mental
275	44	pair
276	44	principle
277	43	efficient
278	43	predicted
279	42	accompanied
280	42	extremely
281	42	fundamental
282	42	transported
283	42	yield
284	41	creation
285	41	excellent
286	41	expanded
287	41	owners
288	41	reveal
289	41	silencing
290	41	threat
291	40	farms
292	40	handling
293	40	investigate
294	40	preventing
295	40	temperatures
296	39	committee
297	39	proportion
298	39	unfortunately
299	38	consequence
300	38	emerged
301	38	enable
302	38	eventually
303	38	examination
304	38	nuclear
305	38	principles
306	38	reactions
307	37	demonstrating
308	37	federal
309	37	hidden
310	37	path
311	37	peak
312	37	slow

Rank	<i>f</i>	Word
313	37	surrounding
314	36	acquired
315	36	approved
316	36	aware
317	36	contexts
318	36	gas
319	36	remove
320	36	repair
321	36	sampling
322	36	suitable
323	36	swim
324	35	aids
325	35	careful
326	35	cat
327	35	circumstances
328	35	examining
329	35	obvious
330	35	powerful
331	34	carefully
332	34	childhood
333	34	comparisons
334	34	substantial
335	34	swimming
336	33	copy
337	33	errors
338	33	inform
339	33	opposite
340	33	paired
341	33	phases
342	33	reasonable
343	33	reductions
344	33	regulations
345	33	repeats
346	33	yielded
347	32	chain
348	32	constant
349	32	decreasing
350	32	equal
351	32	locations
352	32	mainly
353	32	monitor

Rank	<i>f</i>	Word
354	32	repeat
355	32	responding
356	32	technical
357	31	bird
358	31	block
359	31	brains
360	31	collection
361	31	consists
362	31	contributes
363	31	contributing
364	31	displayed
365	31	distributed
366	31	double
367	31	engineered
368	31	feed
369	31	heat
370	31	illustrate
371	31	improvements
372	31	index
373	30	analyzed
374	30	associations
375	30	combinations
376	30	confirm
377	30	permanent

APPENDIX C

LIST OF GSL 3 IN THE LABORATORY ANIMAL CORPUS

Rank	<i>f</i>	Word
1	1008	species
2	694	exposure
3	648	rats
4	621	genetic
5	533	clinical
6	513	infection
7	510	found
8	429	laboratory
9	340	experimental
10	292	virus
11	287	tissue
12	280	infections
13	272	strains
14	238	therapy
15	230	depression
16	211	nerve
17	198	survival
18	195	disorders
19	186	functional
20	186	typically
21	184	tissues
22	180	platform
23	174	breeding
24	170	variation
25	168	characterized
26	164	welfare
27	159	viruses
28	155	symptoms
29	153	biological
30	150	rat
31	149	dependent
32	149	motor
33	144	naturally

Rank	<i>f</i>	Word
34	140	subsequent
35	134	strain
36	132	united
37	123	similarly
38	120	pigs
39	120	rapid
40	118	underlying
41	113	potentially
42	112	infants
43	111	formation
44	109	anxiety
45	107	breeds
46	107	guidelines
47	106	breed
48	103	numerous
49	103	recovery
50	103	secondary
51	102	altered
52	101	harm
53	101	therapies
54	99	transportation
55	98	sensitive
56	97	resistance
57	96	deficits
58	96	density
59	96	disorder
60	93	intervention
61	91	nerves
62	90	muscle
63	89	frequency
64	88	evaluation
65	88	respectively
66	85	extensive

Rank	<i>f</i>	Word
67	85	personnel
68	85	ratio
69	84	interventions
70	84	versus
71	83	distinct
72	82	bias
73	81	laboratories
74	81	rearing
75	81	tanks
76	80	evaluate
77	80	nervous
78	79	isolated
79	79	maintenance
80	78	stem
81	76	detected
82	75	hypothesis
83	73	carbon
84	73	stroke
85	72	pregnancy
86	71	evaluated
87	70	exposures
88	70	infant
89	68	alter
90	66	discovery
91	65	tail
92	64	reared
93	64	restriction
94	64	ultimately
95	63	limitations
96	63	tank
97	62	apparent
98	62	involvement
99	62	regulate
100	62	surgery
101	62	unknown
102	60	detect
103	60	preference
104	59	external
105	59	measurements
106	57	diversity
107	57	subsequently

Rank	<i>f</i>	Word
108	55	presumably
109	54	aggressive
110	54	significance
111	53	acquisition
112	53	insight
113	52	establishment
114	50	institutional
115	50	modified
116	50	pregnant
117	49	greatly
118	48	evolution
119	48	reliable
120	47	mating
121	47	million
122	47	profiles
123	47	regulated
124	47	successfully
125	46	comprehensive
126	45	disaster
127	44	complexity
128	43	mature
129	42	adequate
130	42	cluster
131	42	pig
132	41	composition
133	41	perceptions
134	40	accurate
135	40	characterize
136	40	chickens
137	40	frequent
138	40	motivation
139	40	visual
140	39	approval
141	39	compounds
142	39	profile
143	39	restricted
144	39	reverse
145	38	initially
146	38	newly
147	38	structural
148	37	moderate

Rank	<i>f</i>	Word
149	37	parallel
150	37	summary
151	36	evaluating
152	36	profiling
153	36	regulating
154	36	variations
155	35	acceptable
156	35	awareness
157	35	chicken
158	35	efficiency
159	35	insights
160	35	random
161	35	recognition
162	35	routine
163	34	determination
164	34	framework
165	34	isolates
166	34	scope
167	33	equivalent
168	32	attributed
169	32	continuous
170	32	intense
171	32	mate
172	32	mode
173	32	perception
174	31	chamber
175	31	cigarette
176	31	consumed
177	31	peer
178	31	solids
179	31	underlie
180	30	publication
181	30	unlikely

APPENDIX D**LIST OF AWL IN THE LABORATORY ANIMAL CORPUS**

Rank	<i>f</i>	Word
1	485	behavioral
2	388	induced
3	338	tumor
4	303	maternal
5	282	monkeys
6	256	infected
7	243	chronic
8	224	offspring
9	223	molecular
10	203	immune
11	193	peripheral
12	193	receptor
13	190	tumors
14	183	dose
15	183	transmission
16	171	cues
17	170	developmental
18	167	conservation
19	167	impact
20	161	neurons
21	159	acute
22	159	traits
23	157	consumption
24	154	acid
25	151	cognitive
26	148	commonly
27	147	ethical
28	140	infectious
29	139	differentiation
30	139	induce
31	139	prevalence
32	134	physiological
33	133	neural

Rank	<i>f</i>	Word
34	133	pathways
35	131	biology
36	129	mediated
37	128	aspects
38	126	sensory
39	126	stimuli
40	124	genetics
41	120	doses
42	119	transcription
43	118	parameters
44	117	elevated
45	117	genetically
46	110	identification
47	109	protocols
48	109	sensitivity
49	106	fetal
50	106	progression
51	105	criteria
52	104	validity
53	103	spatial
54	102	receptors
55	100	metabolism
56	99	mortality
57	99	spontaneous
58	97	mutations
59	97	syndrome
60	96	diagnostic
61	94	chemotherapy
62	94	detection
63	94	duration
64	92	liver
65	91	interval
66	90	modifications

Rank	<i>f</i>	Word
67	89	availability
68	89	diagnosis
69	89	interestingly
70	88	incidence
71	88	lung
72	87	adverse
73	87	susceptible
74	86	comparative
75	86	onset
76	84	injection
77	84	separation
78	83	facilitate
79	83	variants
80	82	markers
81	81	cortex
82	81	diverse
83	80	colonies
84	79	candidate
85	79	correlated
86	79	obtained
87	78	stimulus
88	77	colony
89	75	pathway
90	74	respiratory
91	72	published
92	72	surgical
93	71	organism
94	70	bacteria
95	70	differential
96	69	induction
97	69	paradigm
98	69	protocol
99	69	readily
100	68	antibodies
101	67	chromosome
102	66	cattle
103	66	distribution
104	65	hormone
105	65	statistical
106	65	variability
107	62	mechanical

Rank	<i>f</i>	Word
108	62	minimal
109	61	oxygen
110	61	replication
111	60	hormones
112	60	importantly
113	60	limb
114	59	axis
115	59	comparable
116	58	cue
117	58	manipulation
118	58	modification
119	58	organisms
120	58	tolerance
121	57	conduction
122	57	induces
123	57	similarities
124	57	vectors
125	56	activated
126	56	differentiate
127	56	probe
128	56	stereotyped
129	55	identical
130	54	characterization
131	54	dominant
132	54	optimal
133	53	migration
134	52	bacterial
135	52	circulating
136	52	injections
137	52	minimize
138	51	inhibition
139	51	integration
140	51	kidney
141	50	confounding
142	49	array
143	49	diagnosed
144	49	fibers
145	49	hepatitis
146	49	impacts
147	49	obtain
148	49	paradigms

Rank	<i>f</i>	Word
149	49	spectrum
150	48	psychiatric
151	47	diameter
152	47	systematic
153	46	consensus
154	46	conserved
155	46	enzymes
156	46	evident
157	46	intervals
158	46	progressive
159	46	replacement
160	45	nucleus
161	45	partial
162	45	toxic
163	45	transplanted
164	44	fiber
165	44	fluid
166	44	linear
167	44	locus
168	44	nutrient
169	44	radiation
170	43	emergence
171	43	longitudinal
172	43	statistically
173	43	trauma
174	43	utility
175	43	vector
176	43	vitamin
177	43	widespread
178	42	antibody
179	42	assembly
180	42	correlation
181	42	habitat
182	42	marker
183	42	molecules
184	42	mutation
185	42	resistant
186	42	selective
187	42	thermal
188	40	adaptive
189	40	ethics

Rank	<i>f</i>	Word
190	40	occurrence
191	40	painful
192	40	translation
193	39	differentiated
194	39	monkey
195	39	temporal
196	39	trait
197	38	cord
198	38	ecological
199	38	inhibit
200	38	reproduction
201	38	reservoir
202	38	sodium
203	37	dependence
204	37	donor
205	37	oral
206	37	organs
207	37	predator
208	37	sperm
209	36	interact
210	36	media
211	36	parental
212	35	aspect
213	35	calcium
214	35	complications
215	35	dynamic
216	35	incidences
217	35	innate
218	34	neuron
219	34	organ
220	34	profound
221	33	arrays
222	33	dosing
223	33	encoding
224	33	enzyme
225	33	likelihood
226	33	similarity
227	33	subset
228	32	classification
229	32	inducing
230	32	quantitative

Rank	<i>f</i>	Word
231	31	manipulations
232	31	standardized
233	31	tract
234	31	urine
235	30	adjacent
236	30	fever
237	30	prominent
238	30	rack
239	30	randomized
240	30	reactive

APPENDIX E**LIST OF OWL IN THE LABORATORY ANIMAL CORPUS**

Rank	<i>f</i>	Word
1	675	zebrafish
2	533	methylation
3	435	epigenetic
4	372	cocaine
5	284	genome
6	254	canine
7	237	nicotine
8	219	rodents
9	216	histone
10	206	wildlife
11	198	rhesus
12	196	neuropathy
13	195	veterinary
14	188	macaques
15	186	rodent
16	185	primates
17	185	swine
18	177	ketamine
19	177	viral
20	175	invertebrates
21	175	nonhuman
22	170	genomic
23	168	pet
24	165	influenza
25	162	lesions
26	161	primate
27	160	bartonella
28	159	vervet
29	158	transgenic
30	156	activation
31	154	investigators
32	147	promoter
33	145	extinction

Rank	<i>f</i>	Word
34	145	obesity
35	143	metabolic
36	142	therapeutic
37	141	alterations
38	141	diabetes
39	141	insulin
40	134	plasma
41	128	anesthesia
42	128	baboons
43	128	lifespan
44	121	susceptibility
45	120	intake
46	119	phenotype
47	118	addiction
48	118	biomedical
49	117	cage
50	115	imprinted
51	115	prenatal
52	115	vivo
53	112	regulatory
54	112	vitro
55	110	cellular
56	109	glucose
57	107	mammalian
58	107	pathogens
59	107	translational
60	106	hippocampus
61	106	inflammatory
62	106	proliferation
63	104	neuronal
64	100	bladder
65	100	pathogenesis
66	98	quail

Rank	<i>f</i>	Word
67	95	anesthetic
68	95	efficacy
69	94	avian
70	94	cortisol
71	94	macaque
72	93	administered
73	93	phenotypes
74	91	chromatin
75	91	maze
76	91	postnatal
77	91	preclinical
78	90	allele
79	89	degeneration
80	89	distal
81	89	larval
82	89	somatic
83	89	testosterone
84	88	apes
85	88	parrots
86	88	reproductive
87	88	sexually
88	87	serum
89	87	spondylosis
90	85	mammals
91	85	neoplasia
92	85	withdrawal
93	84	abnormalities
94	83	additionally
95	82	marmosets
96	82	zoonotic
97	81	covariates
98	81	relapse
99	81	stereotypies
100	79	baseline
101	79	genomes
102	78	captive
103	78	euthanasia
104	78	invasive
105	78	navigation
106	78	spinal
107	78	stressors

Rank	<i>f</i>	Word
108	77	opioid
109	76	hippocampal
110	76	leprosy
111	75	colonization
112	75	dysfunction
113	75	genotype
114	75	mirnas
115	75	severity
116	75	stimulation
117	74	ecosystem
118	73	abnormal
119	73	folate
120	73	methyl
121	72	gestation
122	72	pathogen
123	71	endocrine
124	71	enrichment
125	71	neurophilia
126	70	baboon
127	70	inbred
128	70	mrna
129	70	vaccine
130	69	poultry
131	69	transcriptional
132	68	embryonic
133	68	impaired
134	67	cages
135	67	inflammation
136	67	pets
137	67	surveillance
138	66	husbandry
139	65	dietary
140	65	implicated
141	65	methylated
142	65	serotonin
143	64	autism
144	63	clinically
145	63	hypomethylation
146	63	isolation
147	63	nutritional
148	63	removal

Rank	<i>f</i>	Word
149	62	depressive
150	62	hypermethylation
151	62	larvae
152	62	locomotor
153	62	repetitive
154	62	transplantation
155	62	veterinarians
156	61	assays
157	61	disruption
158	61	dorsal
159	61	invertebrate
160	60	pathology
161	60	physiologic
162	60	postweaning
163	60	prolonged
164	59	axons
165	59	distress
166	59	estradiol
167	59	etiology
168	59	imprintome
169	59	lambs
170	59	longevity
171	59	relevance
172	58	chimpanzees
173	58	imprinting
174	58	persistent
175	58	protective
176	58	stressful
177	57	african
178	57	epigenetics
179	57	mycobacteria
180	57	neuropathic
181	56	marmoset
182	56	zoos
183	55	adulthood
184	55	autoimmune
185	55	avoidance
186	55	endogenous
187	55	striatum
188	55	transcripts
189	54	armadillos

Rank	<i>f</i>	Word
190	54	neurological
191	54	pancreatic
192	54	prevention
193	53	embryos
194	53	stressor
195	52	antidepressant
196	52	cardiovascular
197	52	inhibitor
198	52	inhibitors
199	52	nutrition
200	52	pups
201	51	dopamine
202	51	epilepsy
203	51	ongoing
204	51	pathogenic
205	51	reversal
206	51	vervets
207	50	acetylation
208	50	apoptosis
209	50	carcinogenesis
210	50	feasible
211	50	gestational
212	50	grooming
213	50	prostate
214	50	reinstatement
215	50	robust
216	50	scaffolds
217	50	skeletal
218	50	spores
219	49	antigen
220	49	aquarium
221	49	composting
222	49	epithelial
223	49	extensively
224	49	mutant
225	49	outbreak
226	49	oxidative
227	49	plasticity
228	48	alpha
229	48	carcinoma
230	48	consistently

Rank	<i>f</i>	Word
231	48	diabetic
232	48	ecosystems
233	48	evacuation
234	48	glutamate
235	48	immunity
236	48	physiology
237	48	renewal
238	48	tobacco
239	47	american
240	47	axonal
241	47	cohort
242	47	mirna
243	47	parasite
244	47	pathological
245	47	polymorphisms
246	47	toxicity
247	47	vertebrate
248	47	zoo
249	46	genotypes
250	46	livestock
251	46	neonatal
252	46	phenotypic
253	46	proximal
254	46	rabbits
255	46	utero
256	46	vascular
257	46	vulnerability
258	45	appetitive
259	45	aquariums
260	45	aquatic
261	45	fecal
262	45	henselae
263	45	mild
264	45	necrosis
265	45	palate
266	44	experimentally
267	44	gwas
268	44	hypertension
269	44	inoculation
270	44	methionine
271	44	mycobacterium

Rank	<i>f</i>	Word
272	44	predictive
273	44	systemic
274	43	aberrant
275	43	embryo
276	43	ischemic
277	43	neurogenesis
278	43	noxious
279	43	outbreaks
280	43	pandemic
281	43	shedding
282	42	covariate
283	42	estrogen
284	42	hepatic
285	42	hypothesized
286	42	malignant
287	42	neoplasms
288	42	osteoblasts
289	42	paternal
290	42	vertebrates
291	42	windrows
292	41	captivity
293	41	compost
294	41	folic
295	41	ganglia
296	41	genomics
297	41	homeostasis
298	41	impairment
299	41	intact
300	41	odor
301	41	oncology
302	41	pharmacological
303	41	ram
304	41	sensitization
305	41	spawning
306	41	suppression
307	40	allocentric
308	40	armadillo
309	40	concea
310	40	corticosterone
311	40	crabs
312	40	deer

Rank	<i>f</i>	Word
313	40	deletion
314	40	dysregulation
315	40	endpoints
316	40	germline
317	40	pedigree
318	40	promoters
319	40	relocation
320	39	analgesia
321	39	ethanol
322	39	humane
323	39	integrity
324	39	olfactory
325	39	overexpression
326	39	regeneration
327	39	traumatic
328	38	axon
329	38	carcasses
330	38	feces
331	38	intravenous
332	38	juvenile
333	38	lactation
334	38	lineage
335	38	nociception
336	38	sentinel
337	38	steroid
338	38	subclinical
339	37	endothelial
340	37	kinase
341	37	mimic
342	37	striatal
343	37	vaccines
344	36	aquaculture
345	36	assay
346	36	aversive
347	36	cognition
348	36	cox
349	36	dendritic
350	36	drosophila
351	36	egocentric
352	36	lentiviral
353	36	maturation

Rank	<i>f</i>	Word
354	36	median
355	36	mesenchymal
356	36	metastatic
357	36	murine
358	36	nucleotide
359	36	potent
360	35	aggression
361	35	compulsive
362	35	contamination
363	35	disinfection
364	35	extinguished
365	35	morbidity
366	35	neuropathies
367	35	transient
368	34	biochemical
369	34	cynomolgus
370	34	escalation
371	34	institute
372	34	macrophages
373	34	manifestations
374	34	microsporidian
375	34	refinement
376	34	remodeling
377	34	reproducibility
378	34	supplementation
379	34	suppressor
380	34	veterinarian
381	33	antagonists
382	33	ape
383	33	europe
384	33	excessive
385	33	fetus
386	33	heterogeneity
387	33	microbiome
388	33	myelin
389	33	prospective
390	33	renal
391	33	ventral
392	32	conversely
393	32	dopaminergic
394	32	histologic

Rank	<i>f</i>	Word
395	32	linkage
396	32	metastasis
397	32	microbial
398	32	parrot
399	32	upregulation
400	31	carcinogen
401	31	causal
402	31	chelonae
403	31	differentially
404	31	dimorphic
405	31	exogenous
406	31	guidance
407	31	inherent
408	31	lethal
409	31	mazes
410	31	recirculating
411	31	reflex
412	31	repression
413	31	reprogramming
414	31	responsive
415	31	sprague
416	31	tandem
417	30	anesthetics
418	30	antitumor
419	30	autonomic
420	30	beneficial
421	30	caribbean
422	30	endangered
423	30	epidemiological
424	30	intestinal
425	30	leptin
426	30	morally
427	30	negatively
428	30	neurobiological
429	30	persistence

Rank	<i>f</i>	Word
430	30	prefrontal
431	30	rabbit
432	30	sera
433	30	spine

APPENDIX F
THE CONCORDANCES LINE OF GRAMMATICAL
COLLOCATIONS (NOUN + PREPOSITION)

ID	Concordance Line			Collocation
70	The dramatic increase in the use of	zebrafish	(Danio rerio) in biomedical research has led to a corresponding increased interest in the diseases affecting this important biological model.	n+p
513	Aspects of Larval Rearing Fish-and	zebrafish	(Danio rerio) in particular-are now the second-most used biomedical model in the United Kingdom.	n+p
281	Severe hepatic megalocytosis was found in	zebrafish	in a new facility and was presumed to be associated with a toxicant from new plastics, glues, and the like.	n+p
249	Listed are both documented effects and our interpretation of potential effects on	zebrafish	in a research setting.	adj+n , n+p
37	This method has been applied to M. chelonae from	zebrafish	in a single facility, distinguishing multiple strains, and we continue to use this method as a routine part of our diagnostic screening of cultures from both fish and environmental sources.	n+p
276		Zebrafish	in advanced stages of EAIF exhibit swollen abdomens, but mild forms of the condition may be subclinical.	n+p
667	For the foreseeable future, the use of	zebrafish	in biomedical research will continue to expand.	n+p
246	The two most common infections of	zebrafish	in research laboratories are microsporidiosis, caused by Pseudoloma neurophilia, and mycobacteriosis.	n+p
490		Zebrafish	in the laboratory can occasionally breed throughout the day.	n+p
382	Although knowledge of the viruses that infect the aquarium fish species that are raised with	zebrafish	in the pet trade is limited, the viruses that have been described include herpesviruses, iridoviruses, a nodavirus, and a reovirus.	n+p

ID	Concordance Line			Collocation
525	Studies of	zebrafish	in their natural habitat can be used to understand their general environment.	n+p
Frequency of collocation (zebrafish in) = 11				
107	Because the treatment of	zebrafish	with antimicrosporidial drugs may exacerbate impacts on research outcomes, the only effective method of controlling P. neurophilia infections in zebrafish is identification and removal of infected fish and avoidance of introduction of the parasite by proper quarantine and screening of incoming fish.	n+p
253		Zebrafish	with clinical microsporidiosis are emaciated and may exhibit skeletal deformities, but many infected fish are asymptomatic and the infection is often seen in sentinel fish.	n+p
98	Obviously, the control of this parasite in existing facilities is much more complex and requires systematic screening and isolation of	zebrafish	with known infections in order to eliminate or reduce the presence of P. neurophilia infections in the colony.	n+p
203	In female	zebrafish	with normal functional adult ovaries, we have seen sperm-producing testicular tissue in pancreas of WT lines.	adj+n, n+p
106	We also strongly recommend that zebrafish be obtained from suppliers who do not maintain	zebrafish	with other aquarium fish species.	v+n , n+p
143	We have recently shown that the specific protocol for infection of	zebrafish	with P. tomentosa is critical for optimal tumor induction and promotion in carcinogen studies.	n+p
155	We are anxious to obtain live	zebrafish	with skin or fin papillomas because we believe that these papillomas are the best neoplasm candidates for harboring a tumorigenic virus of zebrafish.	adj+n, n+p
77	As with other animals used in research, experiments using	zebrafish	with these infections may be subject to nonexperimental variation, potentially confounding results, as has been described in laboratory colonies of rabbits and mice infected with the microsporidian parasite Encephalitozoon cuniculi.	v+n , n+p

ID	Concordance Line			Collocation
265	A wide variety of chronic inflammatory lesions are observed in clinically normal	zebrafish	with underlying Mycobacterium spp. infections, and common targets are abdominal organs such as the kidney, spleen, ovary, and swim bladder (diffuse, chronic aerocystitis).	adj+n , n+p
319	The lack of information about naturally occurring viral infections in zebrafish reflects a lack of investigation in this area rather than an inability of viral pathogens to infect zebrafish, as evidenced by studies documenting the experimental infection of	zebrafish	with viruses isolated from other fish species and the presence of multiple endogenous retroviruses, retrotransposons, and other retroviral agents in the zebrafish genome.	n+p
Frequency of collocation (zebrafish with) = 10				

APPENDIX G
THE CONCORDANCE LINES OF LEXICAL COLLOCATIONS
(NOUN + NOUN)

ID	Concordance Line			Collocation
126	Despite the high level of sophistication of genetic and genomic tools available for the	zebrafish	model, basic pathology data for this species still lag far behind the data available for most mammalian laboratory and domestic animal species.	n+n
136	These studies highlight the need for careful consideration of diet and husbandry in order to ensure valid and reproducible data in research using the	zebrafish	model.	n+n
175	Many	zebrafish	models with inactivating mutation in tumor suppressor genes, including tp53, mlh1, msh2, and msh6, and ribosomal genes, show high incidences of malignant peripheral nerve sheath tumors when the human or rodent cancer spectrum from inactivating mutations in the orthologous tumor suppressor genes causes a much wider range of neoplasms in mesenchymal, epithelial, or lymphomyeloid tissues.	adj+n , n+n
222	One of the most urgent issues in the rapidly growing field of cancer research using the	zebrafish	model is the need to optimize aquaculture systems and diets to eliminate or minimize the natural carcinogens and possible tumor promoters that confound research in many recirculating systems using commercial diets.	n+n

ID	Concordance Line			Collocation
232	The	zebrafish	model can play a unique role in discovery of novel contrast agents for tumor imaging as well as in development of innovative anticancer drugs and more effective delivery methods, such as use of nanoparticles to deliver drug combinations in a tissue-targeted fashion.	n+n
297	Despite the ignorance about naturally occurring zebrafish viruses,	zebrafish	models are rapidly expanding in areas of biomedical research where the confounding effects of unknown infectious agents present a serious concern.	n+n
310	The growing importance and expansion of	zebrafish	models in biomedical research necessitate improvements in the standards of husbandry and biosecurity in laboratory zebrafish colonies.	n+n
326	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.	n+n
340	Because	zebrafish	models have expanded from developmental biology and genetics to include models for toxicology, aging, cancer, infection, and immunology, many kinds of research may be seriously and adversely affected by unidentified underlying viral infections.	n+n
392	The utility of the	zebrafish	model for infection and immunity experiments has been extensively reviewed and hinges on the functional similarity of the zebrafish immune system and the mammalian immune system.	n+n
424	However, advantages to	zebrafish	models for mammalian viral infections include the capacity for live imaging, whole-organism histopathology and immunohistochemistry, and temperature-shift experiments.	n+n
Frequency of collocations (zebrafish model) = 22				

ID	Concordance Line			Collocation
218	Critical Role of Genetic Background as an Influence on Neoplasia in Zebrafish and Other Species A factor that has received little attention in	zebrafish	research until the past 5 years is the critical role of genetic background in determining tumor incidences as well as other physiological parameters, such as disease resistance, immune responses, and other endpoints, in response to toxicant exposure.	n+n
258	It is reasonable to assume that space-occupying or inflammatory lesions in the brain, spinal cord, meninges, or skeletal muscle could profoundly affect behavior, which is an emerging field of	zebrafish	research.	n+n
270	We are not aware of other examples in	zebrafish	research in which confounding effects of mycobacteriosis have been documented.	n+n
338	The potential cost of undiagnosed viral infections may be increasing as the nature of	zebrafish	research changes and as zebrafish facilities become more centralized.	n+n
342	Because no naturally occurring viral infections have been reported in zebrafish, it is not possible to provide specific examples of how naturally occurring viral infections have confounded	zebrafish	research.	v+n , n+n
439	However, zebrafish studies of infection and immunity, as well as other types of	zebrafish	research, are at risk of being confounded by unrecognized naturally occurring viral infections.	n+n
586	In truth, the	zebrafish	research drives the need for the animal housing and that need can be satisfied by anything from a single glass aquarium on a lab bench to a large housing room with rows of racks and thousands of small plastic tanks.	n+n
587	With three decades of exponential growth in	zebrafish	research and its associated need for animal housing, commercial companies have emerged to fill the zebrafish housing need and though some innovative ideas are present, the standard for zebrafish housing in the research lab is the rack and tank system that, from a distance, might resemble books on shelves in a library.	n+n

ID	Concordance Line			Collocation
598	As in rack construction, materials toxicity in tank construction is a concern in	zebrafish	research.	n+n
Frequency of collocations (zebrafish research) = 18				
3	Mycobacteriosis in Zebrafish Colonies Mycobacteriosis, a chronic bacterial infection, has been associated with severe losses in some	zebrafish	facilities and low-level mortalities and unknown impacts in others.	n+n
67	Another microsporidium, Pleistophora hyphessobryconis, has recently been detected in a few	zebrafish	facilities.	adj+n, n+n
72	The present review focuses specifically on the transmission and control of microsporidia in	zebrafish	facilities.	n+n
96	Chlorine is commonly used to disinfect tanks and other equipment in	zebrafish	facilities.	n+n
264	This demonstrates that opportunistic Mycobacterium spp. are ubiquitous in	zebrafish	facilities, even those with pathogen control.	n+n
339	The potential cost of undiagnosed viral infections may be increasing as the nature of zebrafish research changes and as	zebrafish	facilities become more centralized.	n+n
360	Most commonly reported are spermatocytic seminomas affecting the testes, spindle cell sarcomas (malignant nerve sheath tumors), ultimobranchial gland tumors, and gastrointestinal tumors, which may have a neoplastic or preneoplastic prevalence of greater than 30% in some	zebrafish	facilities.	n+n
370	Because naturally occurring viral infections can confound research and therefore pose risks to	zebrafish	facilities, it is important to prevent the introduction of new viruses into naive fish colonies and to identify and manage existing pathogens within a colony.	n+n

ID	Concordance Line			Collocation
377	A recent example of pathogen transmission from the pet trade to laboratory zebrafish is neon tetra disease, caused by the microsporidian <i>Pleistophora hyphessobryconis</i> , which was detected in three unrelated laboratory	zebrafish	facilities.	n+n
449		Zebrafish	facilities should exclusively use independent sources of purpose-bred laboratory zebrafish, such as the Zebrafish International Resource Center and the Sinnhuber Aquatic Resources Laboratory.	n+n
535	However, there are many factors governing the overall running of	zebrafish	facilities, not all of which relate directly to meeting the biological needs of the fish.	n+n
Frequency of collocations (zebrafish facilities) = 20				
2	Mycobacteriosis in	Zebrafish	Colonies Mycobacteriosis, a chronic bacterial infection, has been associated with severe losses in some zebrafish facilities and low-level mortalities and unknown impacts in others.	n+n
146	The mycobacterial strain that most often infects several	zebrafish	colonies in Oregon is <i>Mycobacterium chelonae</i> , a relatively nonpathogenic agent that typically causes mild focal lesions in and around the gas bladder.	adj+n , n+n
147	This strain does not appear to increase the incidence of neoplasia in	zebrafish	colonies with or without carcinogen treatment.	n+n
148	In contrast, the more pathogenic strain, <i>Mycobacterium haemophilum</i> , which has occurred in	zebrafish	colonies in Singapore, causes severe diffuse inflammation throughout most visceral organs.	n+n
312	Improved biosecurity of	zebrafish	colonies requires a better understanding of natural occurring zebrafish pathogens, including viruses, their diagnosis, and how they are transmitted.	n+n
381	The often unrestricted relationship between the pet trade and research colonies undoubtedly facilitates the introduction of viral pathogens, as well as parasites and bacteria, into laboratory	zebrafish	colonies.	n+n

ID	Concordance Line			Collocation
387	Moreover, many newer zebrafish facilities are centralized; therefore, an epizootic event caused by breaches in biosecurity has the potential to be devastating by exposing the	zebrafish	colonies of multiple investigators to pathogens during an outbreak of disease.	n+n
Frequency of collocations (zebrafish colonies) = 17				
94	In fact, the method of Whipps and Kent was used to screen eggs and larval fish in the development of a P. neurophilia SPF	zebrafish	colony at Oregon State University.	n+n , n+n
331	The lack of information about viral infections in zebrafish can be partially explained by the fact that, in the absence of an epizootic event, chronic morbidity and mortality in a	zebrafish	colony generally have not prevented researchers from collecting a sufficient number of zebrafish embryos to conduct their studies. Consequently, many researchers are accustomed to accepting a level of morbidity and mortality.	n+n
Frequency of collocations (zebrafish colony) = 2				
558		Zebrafish	housing systems have evolved considerably in the 30 years since Dr. Streisinger published his seminal work, "Production of Clones of Homozygous Diploid Zebra Fish (<i>Brachydanio rerio</i>)," in Nature.	n+n
563	The elemental operating goal of any	zebrafish	housing system is to provide a stable and favorable macroenvironment for the animals housed within it.	n+n
565	The great majority of	zebrafish	housing systems are recirculating, so this review will focus on the major treatment steps employed in this application: solids removal, biological filtration, chemical filtration, aeration, and disinfection.	n+n
585		Zebrafish	housing in the research laboratory is often referred to as "racks and tanks." In truth, the zebrafish research drives the need for the animal housing and that need can be satisfied by anything from a single glass aquarium on a lab bench to a large housing room with rows of racks and thousands of small plastic tanks.	n+n

ID	Concordance Line			Collocation
588	With three decades of exponential growth in zebrafish research and its associated need for animal housing, commercial companies have emerged to fill the	zebrafish	housing need and though some innovative ideas are present, the standard for zebrafish housing in the research lab is the rack and tank system that, from a distance, might resemble books on shelves in a library.	n+n
589	With three decades of exponential growth in zebrafish research and its associated need for animal housing, commercial companies have emerged to fill the zebrafish housing need and though some innovative ideas are present, the standard for	zebrafish	housing in the research lab is the rack and tank system that, from a distance, might resemble books on shelves in a library.	n+n
590	In its most basic, generic form, a rack in the context of	zebrafish	housing is a structure that safely holds tanks.	n+n
Frequency of collocations (zebrafish housing) = 17				
295	Whereas the viral diseases of research rodents are well characterized and closely monitored, no naturally occurring viral infections have been characterized for the laboratory	zebrafish	(Danio rerio), an increasingly important biomedical research model.	n+n
309	The viral infections of laboratory rodents have been extensively studied, yet virtually nothing is known about naturally occurring viruses of laboratory	zebrafish	.	n+n
350	Mycobacterium spp. are the most important bacterial pathogens of laboratory	zebrafish	and cause a wide array of chronic inflammatory lesions that complicate histopathologic interpretation.	n+n
359	Spontaneously occurring neoplastic lesions are relatively common in laboratory	zebrafish	.	n+n
369	Importance of Biosecurity for Laboratory	Zebrafish		n+n
371	Biosecurity of laboratory	zebrafish	is affected not only by the current lack of knowledge regarding naturally occurring zebrafish viruses but also by the introduction of zebrafish raised by the pet and aquarium trade into laboratory zebrafish facilities.	n+n

ID	Concordance Line			Collocation
376	A recent example of pathogen transmission from the pet trade to laboratory	zebrafish	is neon tetra disease, caused by the microsporidian <i>Pleistophora hyphessobryconis</i> , which was detected in three unrelated laboratory zebrafish facilities.	n+n
Frequency of collocations (laboratory zebrafish) = 14				
6	Occurrence of identical strains of <i>Mycobacterium</i> in both fish and biofilms in	zebrafish	systems suggests transmission can occur when fish feed on infected tissues or tank detritus containing mycobacteria.	n+n
53	<i>Mycobacterium fortuitum</i> infection appears to be a manageable disease but is not usually entirely eliminated from a large recirculation	zebrafish	system.	n+n
582	To maximize energy use for growth and reproductive purposes,	zebrafish	systems monitor and regulate salts, principally sodium chloride, in the fish water to try to minimize the cost of osmoregulation to the fish.	n+n
583	Salinity levels frequently vary in	zebrafish	systems whenever fresh water is added to the system or when water is lost to evaporation.	n+n
Frequency of collocations (zebrafish systems) = 11				
256	However, the detection of subclinical infections may not be a high priority in facilities that rear	zebrafish	embryo because, for example, the adult fish are used as brood stock and not test subjects.	v+n , n+n
675	The Evidence-based Toxicology Collaboration is pioneering this application of SRs in an assessment of the performance of the	Zebrafish	Embryo Test (ZET) in predicting the results of prenatal developmental toxicity tests.	n+n
124	Xenografts of human tumors into	zebrafish	embryos and zebrafish cancer models allowing tumor induction very early in life provide a cost-effective, high-throughput system for drug discovery.	n+n
652	However, this method was not effective for	zebrafish	embryos or early stage larvae.	n+n
1	This effect, called insertional mutagenesis, has been used in forward genetic screens to identify genes involved in	zebrafish	embryonic development.	n+n

ID	Concordance Line			Collocation
332	The lack of information about viral infections in zebrafish can be partially explained by the fact that, in the absence of an epizootic event, chronic morbidity and mortality in a zebrafish colony generally have not prevented researchers from collecting a sufficient number of	zebrafish	embryos to conduct their studies.	n+n
Frequency of collocations (zebrafish embryos) = 6				
456	Breeding is a key element in	zebrafish	husbandry.	n+n
508	Conclusion Breeding is an important part of	zebrafish	husbandry.	n+n
555	There are many areas of	zebrafish	husbandry that still require a great deal of research, and this includes larval rearing.	n+n
556	Although the interest in	zebrafish	husbandry has increased notably in recent years-from animal technologists through managers and directors to veterinarians, legislators, and welfare groups-there are no standardized practices based on sound scientific evidence.	n+n
Frequency of collocations (zebrafish husbandry) = 6				
471		Zebrafish	adults lack strong sexual dimorphism.	n+n
610	The need for embryos has fueled the development of specialized equipment designed to encourage	zebrafish	adults to breed.	v+n , n+n
655	This comparison was also performed by Wilson and colleagues using	zebrafish	adults only.	v+n , n+n
656	For	zebrafish	adults, hypothermal shock or rapid cooling is a more effective euthanasia method than an overdose of MS-222.	n+n
Frequency of collocations (zebrafish adults) = 5				
457	Propagation of	zebrafish	lines, in particular wild-type-derived strains, is closely monitored to ensure that genetic diversity and vigor are maintained.	n+n
458	A robust	zebrafish	line typically carries a large number of polymorphic variations, which may interfere with reproducibility of experiments.	adj+n , n+n

ID	Concordance Line			Collocation
498	Propagation of Wild-Type-Derived Lines A variety of wild-type	zebrafish	lines is available, including the most commonly used laboratory strains: TU, AB, WIK, and Tupfel long fin.	adj+n , n+n
506	Some near-homozygous	zebrafish	lines, such as C32 and SJD, have been made using a combination of inbreeding and heat shock (to generate homozygous gynogenetic diploids) and early pressure (to generate heterozygous gynogenetic diploids), but these lines were difficult to maintain and eventually had to be outcrossed to increase viability.	adj+n , n+n
Frequency of collocations (zebrafish lines) = 5				
9	These are only a few examples and, as research expands, so does the need for development of a comprehensive approach to managing	zebrafish	health.	n+n , n+n
10	With a focus on infectious disease as a primary component of	zebrafish	health, the most commonly encountered diseases are microsporidiosis, caused by Pseudoloma neurophilia, and mycobacteriosis, caused by several species of the genus Mycobacterium.	n+n
301	Identification, characterization, and monitoring of naturally occurring viruses in zebrafish are crucial to the improvement of	zebrafish	health, the reduction of unwanted variability, and the continued development of the zebrafish as a model organism.	n+n
Frequency of collocations (zebrafish health) = 4				
244	For example, it is not difficult to imagine that subclinical infections of the nervous system might have untoward effects on	zebrafish	studies that have behavioral or psychological endpoints, as in some recently published work.	adj+n , n+n
268	Little information exists in the literature on the impacts of asymptomatic mycobacterial infections on the outcomes of	zebrafish	studies.	n+n
275	In one	zebrafish	study, increased atresia (degeneration) of mature eggs was statistically correlated with the occurrence of egg debris.	adj+n , n+n

ID	Concordance Line			Collocation
438	However,	zebrafish	studies of infection and immunity, as well as other types of zebrafish research, are at risk of being confounded by unrecognized naturally occurring viral infections.	n+n
Frequency of collocations (zebrafish studies) = 4				
474	Olfactory cues have been shown to play a role in mediating	zebrafish	reproduction.	n+n
486	These observations suggest that the optimal	zebrafish	reproduction through natural mating occurs when the fish are aged 6 months to 1 year.	adj+n , n+n
489		Zebrafish	reproduction depends strongly on photoperiod.	n+n
509	Also, a better insight into	zebrafish	reproduction and reproductive behavior will come from more comprehensive understanding of molecular mechanisms underlying these processes.	n+n
Frequency of collocations (zebrafish reproduction) = 4				
462	In recent years, understanding of	zebrafish	reproductive biology and behavior has greatly increased.	n+n
481	Also, most of the experiments that have investigated effects of social interactions on	zebrafish	reproductive success were carried out on fish obtained from commercial breeders.	adj+n , n+n
Frequency of collocations (zebrafish reproductive) = 2				
296	Despite the ignorance about naturally occurring	zebrafish	viruses, zebrafish models are rapidly expanding in areas of biomedical research where the confounding effects of unknown infectious agents present a serious concern.	n+n
372	Biosecurity of laboratory zebrafish is affected not only by the current lack of knowledge regarding naturally occurring	zebrafish	viruses but also by the introduction of zebrafish raised by the pet and aquarium trade into laboratory zebrafish facilities.	adj+n , n+n

ID	Concordance Line			Collocation
414	Investigating natural infection with	zebrafish	viruses in addition to experimental infections will improve the usefulness of the zebrafish as a model organism, not only elucidating routes of infection, virulence factors, host defenses, and viral countermeasures to host defenses but also providing information about potential confounding effects, modes of transmission, and necessary biosecurity measures for improved operation of bioaquatics facilities.	n+n
421	The susceptibility of zebrafish to experimental infections with this broad range of viruses suggests not only that naturally occurring viruses occur in zebrafish but also that many viral families may be represented among the as-yet unidentified	zebrafish	viruses.	adj+n , n+n
Frequency of collocations (zebrafish viruses) = 4				
7	Basic research on the growth characteristics, disinfection, and pathogenesis of	zebrafish	mycobacteria is critical moving forward.	n+n
36	Other potentially useful genes that have been used less extensively for	zebrafish	mycobacteria are erp, rpoB, and sod.	n+n
262	When species identifications are made by culture or polymerase chain reaction directly from tissues,	zebrafish	mycobacteria most often turn out to be M. chelonae.	n+n
267		Zebrafish	mycobacteria are not usually identified to the species level because the disease is most often diagnosed by histopathology and the submission of specimens in formalinbased fixatives precludes culturing the organisms.	n+n
Frequency of collocations (zebrafish mycobacteria) = 4				
454		Zebrafish	Breeding in the Laboratory Environment The zebrafish, Danio rerio, has become a major model organism used in biomedical studies.	n+n
466	It is important to emphasize that it may not always be possible to incorporate all of the factors affecting	zebrafish	breeding into breeding protocols performed in the laboratory.	v+n , n+n

ID	Concordance Line			Collocation
467	Detailed step-by-step protocols that describe	zebrafish	breeding are not provided in this article; such protocols can be found elsewhere.	v+n , n+n
510	One of the major concerns related to	zebrafish	breeding and line propagation in the laboratory is preservation of the vigor of a line.	n+n
Frequency of collocations (zebrafish breeding) = 4				
223	The finding of elevated age-specific tumor incidences, remarkably large spontaneous tumors, and hepatocyte megalocytosis in a high percentage of intensive	zebrafish	aquaculture facilities from around the world has profound implications for many institutes at the National Institutes of Health, not just those funding cancer research.	adj+n , n+n
559		Zebrafish	aquaculture housing systems combining design principles from industrial aquaculture, laboratory rodent housing, and research genetics are commercially available from a number of sources, and an increasing number of academic research institutions are constructing large centralized facilities to support their growing zebrafish research programs.	n+n
593	Successful zebrafish rack construction requires attention to the wet, humid, warm environment that results from	zebrafish	aquaculture and the use of materials that will withstand this environment.	n+n
Frequency of collocations (zebrafish aquaculture) = 3				
293	Eggs in turn can be surface-disinfected and hatched in sterile water, which is already a common practice in the	zebrafish	community.	n+n
537	This diversity may reflect the diverse needs within the	zebrafish	community and adaptation of protocols from both hobbyist and aquaculture sources.	n+n
612	With the increased demand for life-support, housing, and husbandry equipment, the	zebrafish	community has seen innovative solutions to its needs and will undoubtedly see improvements in the future.	n+n
Frequency of collocations (zebrafish community) = 3				
81	Conventional polymerase chain reaction (PCR) and real-time PCR-based assays targeting unique portions of the small subunit ribosomal DNA gene are available for testing of	zebrafish	tissues for P. neurophilia.	n+n

ID	Concordance Line			Collocation
144	Certain other infectious agents that often cause profound hyperplasia in	zebrafish	tissues, such as Piscinoodinium pillulare in the gill, have not acted as a tumor promoter in any carcinogen experiments that we have conducted.	n+n
657	Histopathological analysis of adult	zebrafish	tissue showed no evidence of ice crystal formation .	adj+n, n+n
Frequency of collocations (zebrafish tissues) = 3				
224	Now that the	zebrafish	genome is sequenced and knowledge of genetics, genomics, and molecular and cellular development mechanisms has become very sophisticated, interest has grown in use of zebrafish as models for understanding the genetic mechanisms underlying a wide variety of human diseases.	n+n
320	The lack of information about naturally occurring viral infections in zebrafish reflects a lack of investigation in this area rather than an inability of viral pathogens to infect zebrafish, as evidenced by studies documenting the experimental infection of zebrafish with viruses isolated from other fish species and the presence of multiple endogenous retroviruses, retrotransposons, and other retron agents in the	zebrafish	genome.	n+n
Frequency of collocations (zebrafish genome) = 3				
620	The specific agents and methods chosen for	zebrafish	euthanasia will depend upon age and the scientific objectives described in the IACUC animal care and use protocol.	n+n
621	Whatever method is chosen for	zebrafish	euthanasia, "death must be confirmed by examining the animal for cessation of vital signs".	n+n
647	We will discuss three different methods of	zebrafish	euthanasia demonstrating this performance-based approach.	n+n
Frequency of collocations (zebrafish euthanasia) = 3				

ID	Concordance Line			Collocation
305	This article addresses the importance of identifying and characterizing the viral diseases of zebrafish as the scope of zebrafish models expands into new research areas and also briefly addresses	zebrafish	susceptibility to experimental viral infection and the utility of the zebrafish as an infection and immunology model.	v+n , n+n
325	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.	v+n , n+n
433	These experimental infections document	zebrafish	susceptibility to five viral families: Birnaviridae, Rhabdoviridae, Nodaviridae, Iridoviridae, and Herpesviridae.	v+n , n+n
Frequency of collocations (zebrafish susceptibility) = 3				
30	The lengthy incubation time and specialized medium may contribute to the underdiagnosis of this important	zebrafish	pathogen.	adj+n , n+n
313	Improved biosecurity of zebrafish colonies requires a better understanding of natural occurring	zebrafish	pathogens, including viruses, their diagnosis, and how they are transmitted.	adj+n , n+n
315	In fact, few naturally occurring	zebrafish	pathogens of any kind have been well characterized.	adj+n , n+n
Frequency of collocations (zebrafish pathogen) = 3				

APPENDIX H
THE CONCORDANCE LINES OF LEXICAL COLLOCATIONS
(NOUN + VERB)

ID	Concordance Line			Collocation
260		Zebrafish	are now used extensively in tissue transplant studies, in which the immune system is intentionally incapacitated by gamma irradiation, and latent microsporidian infections would likely become clinical in these fish or in other studies that involve immunologically challenged zebrafish.	n+v
422		Zebrafish	have also been used as models of mammalian viral infection, although there are some limitations to their use.	n+v
516	Long established as an aquarium fish, the	zebrafish	(Danio rerio) has been used as a biomedical model since the 1930s.	n+v
658		Zebrafish	and other fish larvae use cutaneous gas exchange for their oxygen demands while their gills are still developing).	n+v
63	Such nonexperimental variation magnifies the importance of evaluating the influences these infections may have when	zebrafish	are used as models for studies on disease, immunology, ecotoxicology, and so on.	n+v
Frequency of collocations (zebrafish-used) = 7				
384	This research not only demonstrates that laboratory zebrafish may still harbor a variety of pathogens from their origin in the pet trade but also provides evidence that	zebrafish	raised in outdoor ponds continue to directly enter laboratory colonies, providing a constant source of new pathogens from other pet and aquarium species.	n+v
373	Biosecurity of laboratory zebrafish is affected not only by the current lack of knowledge regarding naturally occurring zebrafish viruses but also by the introduction of	zebrafish	raised by the pet and aquarium trade into laboratory zebrafish facilities.	n+v

ID	Concordance Line			Collocation
Frequency of collocations (zebrafish-raised) = 4				
430		Zebrafish	treated with the antiviral acyclovir exhibited significantly reduced viral load in all examined regions (head, dorsal midbody, and ventral midbody), whereas zebrafish treated with cyclophosphamide exhibited significantly higher mortality and increased viral loads.	n+v, n+v
361	Florida strain wild type	zebrafish	treated with N-nitroso-N-ethylurea in one study exhibited 100% incidence of cutaneous papillomas; however, no cutaneous papillomas were observed in similar experiments conducted at several other research institutions, suggesting the possibility of an unrecognized oncogenic virus.	n+v
431	Zebrafish treated with the antiviral acyclovir exhibited significantly reduced viral load in all examined regions (head, dorsal midbody, and ventral midbody), whereas	zebrafish	treated with cyclophosphamide exhibited significantly higher mortality and increased viral loads.	n+v, n+v
Frequency of collocations (zebrafish-treated) = 3				
335	Research areas such as aging, cancer, immunity, infection, and toxicology often require that the	zebrafish	be maintained for a much greater portion of their life span and that the histopathologic changes in adult animals be assessed.	n+v
423	For example,	zebrafish	are often maintained at approximately 28°C, whereas some mammalian viruses adapted to replicate at 37°C may not be pathogenic at 28°C.	n+v
329	Historically, the zebrafish has served as a model for vertebrate development and genetics, with virtually all experimentation occurring during the first few days postfertilization; thus, adult	zebrafish	have been traditionally maintained exclusively as breeding stock to produce embryos for experimentation.	adj+n , n+v
Frequency of collocations (zebrafish-maintained) = 3				

ID	Concordance Line			Collocation
113	During the past decade, the	zebrafish	has emerged as a leading model for mechanistic cancer research because of its sophisticated genetic and genomic resources, its tractability for tissue targeting of transgene expression, its efficiency for forward genetic approaches to cancer model development, and its cost effectiveness for enhancer and suppressor screens once a cancer model is established.	n+v
121		Zebrafish	have emerged as a premier vertebrate model system for understanding genes and signaling pathways that control development and mechanisms of disease affecting nearly every organ system.	n+v
Frequency of collocations (zebrafish-emerged) = 2				
85		Zebrafish	frequently spawn spontaneously in aquaria, and hence release of eggs, ovarian fluids, and tissues at spawning provides an important potential route of horizontal transmission.	n+v
487	Sexually mature	zebrafish	can spawn in the laboratory continuously all year at a frequency of two or three times a week.	adj+n , n+v
Frequency of collocations (zebrafish-spawn) = 2				
87	Six-week-old	zebrafish	(AB strain) obtained from the Zebrafish International Resource Center were experimentally exposed to P. neurophilia spores at 10,000 spores per fish.	adj+n , n+v , n+v
105	We also strongly recommend that	zebrafish	be obtained from suppliers who do not maintain zebrafish with other aquarium fish species.	n+v
Frequency of collocations (zebrafish-obtained) = 2				
427	Furthermore, as poikilotherms,	zebrafish	can survive over a range of temperatures, allowing some viral infections to be "halted" by shifting the temperature so that viral replication does not continue, as was recently accomplished with IHNV.	n+v

ID	Concordance Line			Collocation
651	Immediate analysis of the tissue from adult zebrafish that was placed in the garbage disposal showed that none of the adult	zebrafish	survived and that the largest piece of ground up tissue found was about 0.5 in in diameter, with the majority of tissue substantially smaller.	adj+n , n+v
Frequency of collocations (zebrafish-survive) = 2				
353	In fact, the incidence of intestinal neoplasia following treatment with dimethylbenzathracine was significantly higher in	zebrafish	infected by the intestinal nematode Pseudocapillaria tomentosa than in similarly treated uninfected zebrafish.	n+v
52	Given the suggested oral transmission route,	zebrafish	may be infected by the oral route by grazing on the microflora of the biofilm and detritus.	n+v
Frequency of collocations (zebrafish-infected) = 2				

APPENDIX I
THE CONCORDANCE LINES OF LEXICAL COLLOCATIONS
(ADJECTIVE + NOUN)

ID	Concordance Line			Collocation
170	The brain of adult	zebrafish	is histologically quite distinct from that of mammals, with a much greater component of highly cellular areas composed of deeply basophilic embryonal cells surrounding the ventricular system of the forebrain, diencephalon, and myelencephalon.	adj+n , n+v
640	A 2- to 5-?g/ml dosage concentration is recommended to sedate adult	zebrafish	, and 60 to 100 ?g/ml is used for immersion anesthesia.	v+n-n , adj+n
528	Studies reporting zebrafish larval growth in their natural habitats are scarce but suggest that adult	zebrafish	live in streams and come to spawn in flooded, vegetative areas, where the larval forms grow protected from predators before moving into the margins of slow-moving rivers and streams as adults in the rainy season.	adj+n , n+v
653	Because this physical method produces rapid loss of consciousness and subsequent death, thereby minimizing pain, distress, or discomfort to the animal, we believe it to be an acceptable euthanasia technique for adult	zebrafish	.	adj+n
664	Consistent with this idea and the above studies, we observed a brief reaction in adult	zebrafish	in the first 5 seconds when immersed into ice water (2-4°C).	adj+n
541	At midlarval stage, the feeding regimens become very disparate; although Artemia is a well-known foodstuff for both adult	zebrafish	and larval forms with corresponding gape size, usually from 8 to 9 DPF, some facilities do not provide it until 14 to 15 DPF.	adj+n

ID	Concordance Line			Collocation
650	Immediate analysis of the tissue from adult	zebrafish	that was placed in the garbage disposal showed that none of the adult zebrafish survived and that the largest piece of ground up tissue found was about 0.5 in in diameter, with the majority of tissue substantially smaller.	adj+n
Frequency of collocation (adult zebrafish) = 13				
522	Here, protocols have been developed that might owe little to knowledge of the true requirements of larval	zebrafish	and more to the constraints and pressures of fish facilities and laboratories.	adj+n
534	In trials, larval stages able to take Artemia appeared to have higher survival rates than those fed on artificial diets, suggesting that Artemia, although not part of a natural zebrafish diet, has a close approximation of the nutritional needs of larval	zebrafish	.	adj+n
543	Unfortunately, the complete nutritional requirements of both adult and larval	zebrafish	are currently unclear.	adj+n , n+v
539	At this point, some aspects of water quality do not seem as important because larval	zebrafish	seem more tolerant of poorer water quality with respect to free ammonia than adult fish, although temperature will still affect the rate of larval development and thus staging.	adj+n , n+v
259	Larval	zebrafish	are quite susceptible to P. neurophilia infections; hence, the infection could compromise developmental neurotoxicity testing and musculoskeletal development studies.	adj+n , n+v
Frequency of collocation (larval zebrafish) = 5				
45	Tobin and colleagues have identified at least one locus in zebrafish important for M. marinum susceptibility, and Hegedus and colleagues and van der Sar characterized the transcriptome of infected	zebrafish	, suggesting that tools for comparison are available.	adj+n
59	Nevertheless, treatment of infected	zebrafish	may be appropriate when extremely valuable strains or populations are involved.	adj+n
254	Ramsay and colleagues conducted a large, long-term, in vivo study to evaluate the effects of P. neurophilia on growth and fecundity in experimentally infected	zebrafish	, some of which were exposed to stressors.	adj+n

ID	Concordance Line			Collocation
86	Further evidence for the maternal transmission of P. neurophilia was observed in the experiment described by Sanders and Kent, where parasite DNA was detected in the eggs and water from a group spawn of infected	zebrafish	.	adj+n
Frequency of collocation (infected zebrafish) = 4				
356	As with other species, immunocompromised	zebrafish	are likely to be more susceptible to viral infections than wild-type fish, exhibiting higher mortality, higher morbidity, more clinical signs, and more severe histopathologic lesions.	adj+n , n+v
355	In recent years, researchers have developed immunocompromised	zebrafish	to aid in the study of hematopoiesis, tumorigenesis, infection, and immunity.	adj+n
357	Thus, in experimental models where immunocompromised	zebrafish	are required, the use of virus-free fish will play a major role in reducing the variability in data because of confounding factors, such as inflammation and other host responses to infection, and will simultaneously reduce the number of fish required to achieve adequate statistical power.	adj+n , n+v
Frequency of collocation (immunocompromised zebrafish) = 3				
366	The only current effort to generate pathogen-free	zebrafish	is being conducted at the Sinnhuber Aquatic Resources Laboratory at Oregon State University.	adj+n
367	For development of pathogen-free	zebrafish	to be successful, however, naturally occurring viral pathogens affecting zebrafish colonies must be identified and diagnostic assays to detect viral infection must be developed to facilitate the elimination of viral pathogens.	adj+n
Frequency of collocation (pathogen-free zebrafish) = 2				
243	However, investigators may not be fully aware of the potential impact that underlying disease may have on their research when postlarval	zebrafish	are used in medium- or long-term experiments.	adj+n , n+v
257	However, we have seen <0.3% annual mortality in postlarval	zebrafish	at the Oregon State University specific pathogen-free facility that has eliminated this parasite.	adj+n

ID	Concordance Line			Collocation
	Frequency of collocation (postlarval zebrafish) = 2			
375	In contrast to purpose bred research mice, ornamental	zebrafish	are often raised together with other species that harbor numerous pathogens.	adj+n , n+v
299	In addition, many zebrafish research colonies remain linked to the ornamental (pet)	zebrafish	trade, which can contribute to the introduction of new pathogens into research colonies, whereas mice used for research are purpose bred, with no introduction of new mice from the pet industry.	adj+n , n+n
	Frequency of collocation (ornamental zebrafish) = 2			

APPENDIX J
THE CONCORDANCE LINES OF LEXICAL COLLOCATIONS
(VERB + NOUN)

ID	Concordance Line			Collocation
133	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	.	v+n
358	Naturally occurring infections along with oncogenic viruses may also play a role in confounding experiments, including cancer studies using	zebrafish	, because viruses are associated with tumorigenesis in several other fish species.	v+n
364	Oncogenesis due to unrecognized viruses may therefore be a confounding variable in cancer research using	zebrafish	.	v+n
615	Many of the institutions using	zebrafish	(Danio rerio) for research, testing, or teaching are funded by the Public Health Service (PHS) and accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International.	v+n
77	As with other animals used in research, experiments using	zebrafish	with these infections may be subject to nonexperimental variation, potentially confounding results, as has been described in laboratory colonies of rabbits and mice infected with the microsporidian parasite Encephalitozoon cuniculi.	v+n , n+p
Frequency of collocation (using-zebrafish) = 6				

171	These abundant embryonal periventricular cells may predispose	zebrafish	to develop more embryonal neoplasms of the central nervous system, resembling those seen in pediatric cases in humans.	v+n
211	More controlled experiments are needed in order to define the optimal calcium levels in zebrafish diets, but we speculate that, as with bulls showing medullary thyroid neoplasia when fed diets high in calcium designed for lactating dairy cows, elevated dietary calcium may cause hyperplasia of the ultimobranchial gland and predispose	zebrafish	to elevated neoplasm levels.	v+n
Frequency of collocation (predispose-zebrafish) = 2				
318	The lack of information about naturally occurring viral infections in zebrafish reflects a lack of investigation in this area rather than an inability of viral pathogens to infect	zebrafish	, as evidenced by studies documenting the experimental infection of zebrafish with viruses isolated from other fish species and the presence of multiple endogenous retroviruses, retrotransposons, and other retroviral agents in the zebrafish genome.	v+n
443	However, the lack of information regarding naturally occurring viral infections in zebrafish does not reflect an inability of viral pathogens to infect	zebrafish	because zebrafish are easily experimentally infected with a variety of viruses isolated from other fishes.	v+n
Frequency of collocation (infect-zebrafish) = 2				
626	Therefore, the most widely used method to anesthetize or sedate	zebrafish	is immersion, which, in fish, is analogous to inhalant anesthesia in terrestrials.	v+n
640	A 2- to 5-?g/ml dosage concentration is recommended to sedate adult	zebrafish	, and 60 to 100 ?g/ml is used for immersion anesthesia.	v+n-n , adj+n
Frequency of collocation (sedate-zebrafish) = 2				
437	The similarity between the zebrafish and mammalian immune systems, the capacity of the zebrafish model to allow investigation of specific immune system components at different stages of immunologic development, and the molecular, genetic, and imaging tools available for this species make the	zebrafish	particularly useful and attractive as a model for infectious disease and immunity research.	v+n

413	Thus, the similarity between the zebrafish immune system and the mammalian immune system, the genetic and molecular tools available for zebrafish, and the ability to image an entire infected embryo or fish over the course of infection make the	zebrafish	an excellent model to investigate viral pathogenesis and host defenses.	v+n
Frequency of collocation (make-zebrafish) = 2				

APPENDIX K

THE CONCORDANCE LINES OF PURPOSIVE FUNTION

ID	Concordance line
2012(53)2_2(106-113)	The present review focuses specifically on the transmission and control of microsporidia in zebrafish facilities.
2012(53)2_3(114-125)	These studies highlight the need for careful consideration of diet and husbandry in order to ensure valid and reproducible data in research using the zebrafish model.
2012(53)2_5(135-143)	This article addresses the importance of identifying and characterizing the viral diseases of zebrafish as the scope of zebrafish models expands into new research areas and also briefly addresses zebrafish susceptibility to experimental viral infection and the utility of the zebrafish as an infection and immunology model.
2012(53)2_9(179-191)	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.
2012(53)2_10(192-204)	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 .
2013(54)2_10(211-223)	Although model systems such as drosophila, C. elegans and zebrafish have played important roles in biomedicine, such as the discovery of the regulatory role of noncoding RNA in healthy and disease conditions (Ambros, 2003; Ambros, 2008), this review will focus on the mammalian model systems, specifically nonhuman primates (NHPs).

BIOGRAPHY

NAME	Miss Virata Panjanon
DATE OF BIRTH	21 November 1990
PLACE OF BIRTH	Ratchaburi, Thailand
INSTITUTION ATTENDED	Mahidol University, 2009-2012: Bachelor of Arts (English) Mahidol University, 2014-2016: Master of Arts (Applied Linguistics)
HOME ADDRESS	594/537, Khlong Chak Phra Subdistrict Taling Chan District, Bangkok, 10170 Tel. 0870177889 E-mail: virata.panjanon@gmail.com
PRESENTATION	7 th National and International Conference on Humanities and Social Science