



**UNIVERSITI PUTRA MALAYSIA**

***FEATURES OF LEXICAL BUNDLES IN SOCIAL SCIENCE RESEARCH  
ARTICLES OF NATIVE AND NON-NATIVE EXPERT WRITERS OF ENGLISH***

**REZA VASEGHI**

**FPP 2016 22**



**FEATURES OF LEXICAL BUNDLES IN SOCIAL SCIENCE RESEARCH  
ARTICLES OF NATIVE AND NON-NATIVE EXPERT WRITERS OF  
ENGLISH**

By

**REZA VASEGHI**

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,  
in Fulfilment of the Requirements for Degree of Doctor of Philosophy**

**May 2016**

## **COPYRIGHT**

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia



## **DEDICATION**

**Gratefully dedicated to  
My dear wife and parents**



© COPYRIGHT UPM

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment  
of the requirement for the degree of Doctor of Philosophy

**FEATURES OF LEXICAL BUNDLES IN SOCIAL SCIENCE RESEARCH  
ARTICLES OF NATIVE AND NON-NATIVE EXPERT WRITERS OF  
ENGLISH**

By

**REZA VASEGHI**

**May 2016**

**Chairman : Professor Jayakaran A/L A.P. Mukundan, PhD**  
**Faculty : Educational Studies**

Lexical bundles are combinations of three and more words which co-occur most frequently in a given register or genre (Biber et al., 1999). They have been the focus of attention of many researchers in academic written discourse in the last two decades. In spite of the fact that lexical bundles are used much frequently in academic written discourse, it appears that non-native writers have difficulty in the application of these multi-word expressions, particularly in producing writing in their specific scientific field. Having evidence of their significant role in shaping the academic texts, it is worth exploring the usage of lexical bundles in the academic texts, particularly in a specific academic discipline. The present study adopted a corpus-based approach and investigated the usage of lexical bundles in social sciences research articles written by native and non-native Malaysian expert writers in the field. Moreover, the ultimate goal of this study was to develop and offer a useful pedagogical list of lexical bundles for social sciences academic writers.

In the current study, in order to examine the use of lexical bundles by native and non-native Malaysian social scientist writers, discipline-specific corpora, namely, Social Sciences Corpus (SSC) and Malaysian Corpus of Social Sciences (MCSS) were created. The number of tokens for native and non-native Malaysian corpora is 2,007,037 and 325,164 running words, respectively. The corpora were compiled from high-impact leading and prestigious international journals in the field.

Later, AntConc 3.4.3 (Anthony, 2014) software was used in order to extract target bundles from the Social Sciences Corpus (SSC). The concordancer tool was also employed on the data to help the qualitative analysis to examine lexical bundles corresponding contexts and determine the specific functions they perform. Moreover, the two modified versions of analytical frameworks, developed by Biber et al. (1999) and Hyland (2008a), were applied in the current study to classify the bundles according to their structural and functional characteristics. Furthermore, a

series of data refinement criteria were set out by the researcher to create a meaningful, concise and teachable list of lexical bundles and also to limit the number of the lexical bundles for further analysis in terms of their grammatical and pragmatic features. It is worth noting that these criteria were established as a methodological decision to meet the purpose of the current study.

The primary findings of the study indicated that most of the target bundles in the SSC were made up of three-word bundles, and the bundles belonged to a particular disciplinary discourse. In terms of the structural analysis, the results showed that the native social scientist writers show a greater tendency to use more noun and propositional phrases and verb or adjective phrases with 'to' clause fragments in their academic written discourse. The functional analysis revealed that the native social scientists writers used a wide range of research-oriented bundles to describe the content and procedures of their research in detail and also to show that their interpretation is based on empirical results and that the analytical method could be reliably replicated. Overall, the structural and functional analyses led the researcher to conclude that native social scientist writers have their own distinctive characteristics to describe their methodological decisions and argue their ideas based on the existing studies and the results obtained from data.

A comparative analysis between native and non-native corpora also revealed that non-native Malaysian social scientist writers overused and underused target bundles compared to their native counterparts. The results suggest that non-native writers are not able to produce a varied set of expressions. Moreover, the results of structural and functional analyses of target bundles between native and non-native corpora showed few considerable differences as far as the use of lexical bundles were concerned. However, it is interesting to note that non-native Malaysian social scientist writers used much less frequent participate-oriented bundles compared to their native counterparts.

The present study had some pedagogical implications for EAP course instructors and designers to use the useful list of lexical bundles in the class and integrate them into their ESL/EFL curricula. The findings of this study could also be beneficial for writing instructors in providing the teaching materials to raise the novice and non-natives' awareness of appropriate use of lexical bundles to achieve a more native-like style of academic writing.

Abstrak tesis yang dikemukakan kepada Senta Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

**GUGUSAN LEKSIKAL DALAM PENULISAN AKADEMIK PENULIS  
PENUTUR ASAL DAN BUKAN ASAL**

Oleh

**REZA VASEGHI**

**Mei 2016**

**Pengerusi : Profesor Jayakaran A/L A.P. Mukundan, PhD**  
**Fakulti : Pengajian Pendidikan**

Gugusan leksikal (*lexical bundles*) merupakan kombinasi tiga atau lebih perkataan yang wujud secara kerap di dalam sesuatu laras bahasa atau genre (Biber et al.,1999). Gugusan leksikal ini telah menjadi focus pemerhatian ramai penyelidik akademik terutamanya dalam diskursus penulisan bagi tempoh dua dekad kebelakangan ini. Di sebalik hakikat penggunaan gugusan leksikal ini kerap ditemui di dalam diskursus penulisan akademik, ianya juga masih menjadi suatu kesukaran bagi penulis-penulis dari kalangan bukan penutur asal terutamanya dalam penggunaan perkataan bervariasi di dalam penulisan bidang saintifik mereka. Usaha meneroka penggunaan gugusan leksikal dalam penulisan akademik di dalam sesuatu disiplin yang khusus adalah amat berguna memandangkan adanya bukti nyata peranan gugusan leksikal dalam menentukan keberkesanan penulisan akademik. Atas sebab tersebut, kajian ini menggunakan pendekatan korpus dan menyelidik penggunaan gugusan leksikal dalam penulisan artikel kajian yang dihasilkan oleh penulis-penulis dari kalangan penutur asal dan bukan asal (*native and non-native*) Malaysia yang berpengalaman dalam bidang masing-masing. Tambahan lagi, tujuan khusus kajian ini adalah untuk menyediakan dan menawarkan senarai gugusan leksikal yang berdasarkan aspek pedagogi yang boleh digunakan oleh penulis-penulis jurnal akademik dalam bidang sains sosial.

Satu kesatuan kumpulan tulisan-tulisan (*corpora*) berpandukan bidang yang khusus yang dinamakan '*Social Sciences Corpus (SSC) and Malaysian Corpus of Social Sciences (MCSS)*' telah dihasilkan untuk tujuan kajian ini yakni untuk meneliti penggunaan gugusan leksikal oleh penulis-penulis penutur asal dan bukan asal Malaysia dalam bidang sains sosial. Bukti terkumpul dalam korpus penulis-penulis dari kalangan penutur asal dan bukan asal Malaysia ini berjumlah 2,007,037 dan jumlah perkataan adalah sebanyak 325,164. Kesatuan kumpulan tulisan-tulisan ini telah dihimpunkan dari jurnal-jurnal antarabangsa yang berpengaruh tinggi dalam bidang kajian ini.

Seterusnya, perisian AntConc 3.4.3 (Anthony, 2014) telah digunakan untuk mengumpul atau memilih kumpulan gugusan leksikal pilihan kajian dari ‘*Social Sciences Corpus*’ (SSC). Alat penyelarasan khusus juga telah digunakan ke atas data untuk membantu analisa kualitatif mengkaji fungsi khusus gugusan leksikal.

Selanjutnya, dua lagi kerangka analisis yang telah diubahsuai yang telah dirangka oleh Biber et al. (1999) dan Hyland (2008a) telah digunakan dalam kajian ini untuk mengklasifikasi kesatuan kumpulan tulisan-tulisan mengikut struktur dan fungsi masing-masing. Tambahan lagi, satu siri kriteria saringan data telah digunakan oleh penyelidik untuk menghasilkan senarai gugusan leksikal yang bermakna, ringkas dan boleh diajar serta untuk menghadkan bilangan gugusan leksikal untuk analisa seterusnya dalam aspek ciri-ciri tatabahasa dan pragmatik. Adalah penting untuk dinyatakan bahawa ciri-ciri tersebut telah dikenalpasti sebagai satu keputusan metodologi untuk mencapai tujuan kajian ini.

Penemuan-penemuan asas menunjukkan bahawa kebanyakan gugusan sasaran dalam SSC adalah dihasilkan dari gabungan tiga perkataan dan gugusan tersebut mewakili satu laras yang khusus. Dalam aspek analisa struktur, penemuan menunjukkan bahawa penulis-penulis dari kalangan penutur asal dalam bidang sains sosial lebih gemar menggunakan frasa kata-nama, sendi-nama dan kata-nama atau frasa kata-sifat dengan fragmen klausa dalam penulisan akademik mereka. Analisa fungsi pula menunjukkan bahawa penulis-penulis dari kalangan penutur asal menggunakan banyak gugusan leksikal yang berorientasikan kajian dalam menjelaskan kandungan dan prosedur kajian mereka secara terperinci dan juga dalam membuktikan bahawa interpretasi mereka berasaskan keputusan empirikal dan ianya boleh digunapakai untuk menafsirkan kaedah analisis. Secara keseluruhan, analisa struktur dan fungsi telah menerajui penyelidik untuk merangkumkan bahawa penulis-penulis dari kalangan penutur asal dalam bidang sains social memiliki ciri-ciri khusus dalam menjelaskan keputusan metodologi mereka dan dalam berhujah tentang idea-idea mereka yang berasaskan kajian-kajian sedia-ada dan data terkumpul.

Analisa perbandingan di antara kesatuan kumpulan tulisan-tulisan (corpora) penutur asal dan bukan asal Malaysia dalam bidang sains sosial telah membuktikan bahawa penulis-penulis bukan penutur asal menunjukkan penggunaan lebih kerap dan lebih jarang gugusan sasaran berbanding dengan penulis-penulis dari kalangan penutur asal. Hasil kajian mencadangkan bahawa penulis-penulis bukan penutur asal tidak berupaya menghasilkan set ungkapan atau ekspresi yang pelbagai. Tambahan lagi, dapatan kajian berasaskan struktur dan fungsi gugusan sasaran yang diguna oleh penulis-penulis dari kalangan penutur asal dan bukan asal menunjukkan perbezaan ketara dari segi penggunaan gugusan leksikal. Bagaimanapun, penulis-penulis dari kalangan penutur bukan asal Malaysia dalam bidang sains sosial didapati menggunakan gugusan *participate-oriented* yang lebih sedikit berbanding dengan penulis-penulis dari kalangan penutur asal.

Kajian terkini ini memiliki implikasi pedagogi ke atas tenaga pengajar dan pereka kursus EAP dalam menggunakan senarai gugusan leksikal yang berguna dalam pengajaran dan dalam usaha mengintegrasikan gugusan leksikal ke dalam kurikulum ESL/EFL. Hasil kajian ini juga berguna kepada tenaga pengajar bidang penulisan dalam menawarkan bahan pengajaran untuk tujuan melahirkan kesedaran tentang penggunaan gugusan leksikal yang betul di kalangan penulis-penulis baru dan bukan penutur asal agar wujudnya gaya bahasa yang seakan-akan penulis dari kalangan penutur asal dalam semua penulisan akademik mereka.



## ACKNOWLEDGEMENT

The completion of this dissertation owes to all those who contributed me on its way. First and foremost, I should like to extend my deepest appreciation to Professor Dr. Jayakaran A/L A.P. Mukundan, my supervisor, for his time, support, and thoughtful comments throughout the preparation of the dissertation. I would also like to express my grateful thanks to the other members of my supervisory committee, Professor Dr. Chan Swee Heng and Associate Professor Dr. Ain Nadzimah BT Abdullah for their invaluable comments and feedbacks during the study.

I am very indebted to several scholars in the field who helped me with their enlighten guidance in the process of writing of the dissertation: Professor Dr. Douglas Edward Biber, Professor Dr. Hilary Nesi, and Associate Professor Dr. Viviana Cortes. My grateful thanks go to Dr. Masoud Zoghi and Dr. Aliasghar Yousefi Azarfam for their sincere help and support. My heartfelt thanks also in particular go to Mr. Reza Alazmani and Mrs. Seddighe Shakib Kotamjani for their support and encouraging words. They were not only best friends, but a true brother and sister to me.

Last, but not least, I should like to express my sincerest appreciation to my beloved parents as they have always granted me limitless support and inspired in me the desire to continue my education. Innumerable thanks is due to my lovely wife, Hanie, for her support, encouragement, patience, and all the sacrifices she made to achieve this goal. I'm so blessed to have you in my life.

This thesis submitted to the Senate of Universiti Putra Malaysia has been accepted as fulfillment of the requirements for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

**Jayakaran A/L A.P. Mukundan, PhD**

Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Chairman)

**Chan Swee Heng, PhD**

Professor  
Faculty of Modern Languages  
Universiti Putra Malaysia  
(Member)

**Ain Nadzimah BT Abdullah, PhD**

Associate Professor  
Faculty of Modern Languages  
Universiti Putra Malaysia  
(Member)

---

**BUJANG BIN KIM. HUAT, PhD**  
Professor and Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date:

## Declaration by graduate student

I hereby confirm that:

- This thesis is my original work;
- Quotations, illustrations and citations have been duly referenced;
- This thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- Intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- Written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- There is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: \_\_\_\_\_ Date: \_\_\_\_\_

Name and Matric No: Reza Vaseghi, GS31012

## Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: \_\_\_\_\_  
Name of  
Chairman of  
Supervisory  
Committee: Professor Dr. Jayakaran A/L A.P. Mukundan

Signature: \_\_\_\_\_  
Name of  
Member of  
Supervisory  
Committee: Professor Dr.Chan Swee Heng

Signature: \_\_\_\_\_  
Name of  
Member of  
Supervisory  
Committee: Associate Professor Dr. Ain Nadzimah BT Abdullah

## TABLE OF CONTENTS

	<b>Page</b>
<b>ABSTRACT</b>	i
<b>ABSTRAK</b>	iii
<b>ACKNOWLEDGEMENT</b>	vi
<b>APPROVAL</b>	vii
<b>DECLARATION</b>	ix
<b>LIST OF TABLES</b>	xiv
<b>LIST OF FIGURES</b>	xvi
<b>CHAPTER</b>	
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Background of the Study	1
1.2 Lexical Bundles in a Nutshell	3
1.3 Rationale of the Study	5
1.4 Statement of the Problem	8
1.5 Objectives of the Study	10
1.6 Research Questions	10
1.7 Significance of the Study	10
1.8 Scope of the Study	12
1.9 Limitations and Delimitation of the Study	12
1.9.1 Limitations of the Study	13
1.9.2 Delimitations of the Study	13
1.10 Definitions of Key Terms	14
<b>2 LITERATURE REVIEW</b>	<b>16</b>
2.1 Chapter Overview	16
2.2 Introduction	16
2.3 Definition of Corpus	19
2.3.1 Types of Corpora	19
2.4 Theoretical Underpinning of Linguistics	21
2.5 Corpus Linguistics	22
2.6 Corpus Linguistic Approaches	23
2.7 Corpus Processing	24
2.7.1 Output Information from a Corpus	25
2.8 Application of Corpus Linguistics in ELT	25
2.9 An Overview of English for Academic Purposes (EAP)	26
2.10 Corpus Linguistics and Academic Writing	26
2.11 Phraseology	27
2.11.1 Major Approaches to Phraseology	28
2.11.2 Phraseology in Academic English Language	30
2.12 Definition and Description of Lexical Bundle	30
2.12.1 Lexical Bundles Identification	32
2.12.2 Structural and Functional Classifications of Lexical Bundles	34
2.12.3 Lexical bundles' Variability in Registers, Genres and Academic Disciplines	38

2.13	Related Empirical Studies on Lexical bundles in Academic Written Discourse	39
2.13.1	Previous Related Studies: Pros and Cons	49
2.13.2	Final Remark	51
2.14	Past Studies on the Teaching of Lexical Bundles	52
2.15	Problematic Aspect of Defining Native and Non-native Speakers of English	54
2.16	Lexical Bundle: A Negligible Essential in Malaysian Corpus Studies	55
2.17	Towards Developing Malaysian Academic Written (Specialized) Corpus	56
2.18	Theoretical Framework	57
2.18.1	Communicative Competence	58
2.19	Conceptual Framework	63
2.20	Summary	64
<b>3</b>	<b>CORPORA CREATION AND METHODOLOGY</b>	<b>65</b>
3.1	Introduction	65
3.2	Chapter Overview	65
3.3	Research Framework	67
3.4	Corpus Design	69
3.4.1	Purpose, Type and Content of the Corpora	71
3.4.2	Representativeness	71
3.4.3	Corpus Size	73
3.5	Phase 1. Corpora Creation of Native and Non-native Malaysian	74
3.5.1	Compilation Procedure of Social Sciences Corpus (SSC)	74
3.5.2	Compilation Procedure of Malaysian Corpus of Social Sciences (MCSS)	77
3.6	Research Design	78
3.7	Instrumentation	79
3.7.1	AntConc	79
3.7.2	Data Refinement Criteria	81
3.7.3	Structural Classification	83
3.7.4	Functional Classification	84
3.8	Lexical Bundle Identification	86
3.9	Phase 2. Data Extraction and Refinement	87
3.9.1	Lexical Bundles Extraction	87
3.9.2	Data Refinement Criteria	88
3.10	Data Analysis Procedures	93
<b>4</b>	<b>RESULTS AND DISCUSSIONS</b>	<b>96</b>
4.1	Chapter Overview	96
4.2	Research Question No.1	96
4.3	Research Question No.2	102
4.3.1	Structural Classification of SSC Target Bundles	102
4.3.2	Functional Classification of SSC Target Bundles	116
4.3.3	The Relationship between Structural and Functional Characteristics in the SSC	138
4.4	Research Question No. 3 (a)	139

4.5	Research Question No. 3 (b)	144
4.5.1	Structural Classification of MCSS Target Bundles	145
4.5.2	Functional Classification of MCSS Target Bundles	157
4.6	Pedagogical utility of the list of target bundles	166
4.7	Summary	168
<b>5</b>	<b>SUMMARY, CONCLUSION, PEDOGOGICAL IMPLICATIONS, AND RECOMMENDATION FOR FURTHER RESEARCH</b>	170
5.1	Chapter Overview	170
5.2	Summary of Methodology	170
5.3	Summary of Findings and Conclusion	171
5.4	Pedagogical Implications	176
5.5	Contributions of the Study	180
5.6	Recommendations for Further Research	181
	<b>REFERENCES</b>	182
	<b>APPENDICES</b>	199
	<b>BIODATA OF STUDENT</b>	260
	<b>LIST OF PUBLICATIONS</b>	261

## LIST OF TABLES

<b>Table</b>	<b>Page</b>
2.1 Examples of Specialized Corpora	20
2.2 Structural Classification Academic Writing (adapted from Biber et al., 1999, pp. 1015-1024)	35
2.3 Functional Classification of Lexical Bundles (Biber et al., 2004, pp. 384-388)	36
2.4 Functional classification of lexical bundles in academic writing (Hyland, 2008a, pp. 13-14)	37
2.5 Major Pervious Lexical Bundles Studies in Academic Written Discourse	40
3.1 The Detail Contents of the Social Sciences Corpus (SSC)	76
3.2 The detail contents of Malaysian Corpus of Social Sciences (MCSS)	78
3.3 The Modified Version of Structural Classification Academic Writing ( <i>adapted from Biber et al., 1999, pp. 1015-1024</i> )	83
3.4 The Modified Version of Functional classification of lexical bundles in academic writing (adapted from Hyland, 2008a, pp. 13-14)	85
3.5 Detail Description of the Extracted Bundles before Data Refinement	87
3.6 Inclusion and Exclusion criteria	92
3.7 Summary of Data Analysis Procedures	95
4.1 The first 10 most frequent three- to six-word target bundles in the SSC	98
4.2 Comparison of SSC results with Hyland's (2008a) Biology & Business Studies corpus findings	100
4.3 Structural classification of target bundles	103
4.4 Structural comparison of bundles through various disciplines	105
4.5 Noun Phrase Structural Type	107
4.6 Prepositional Phrases Structural Type	109
4.7 Verb-Phrase Structural Type	111
4.8 Adjectival Phrase Structural Type	113
4.9 Clausal Bundles Structural Type	114
4.10 Other expressions	116

4.11	Functional classification of target bundles	117
4.12	Research-oriented's sub-categories	120
4.13	Frequencies of research-oriented sub-categories	121
4.14	Text-oriented's sub-categories	125
4.15	Frequencies of text-oriented sub-categories	126
4.16	Participant-oriented's sub-categories	129
4.17	Frequencies of participant-oriented sub-categories	130
4.18	Frequencies of stance sub-functions	131
4.19	The 10 most frequent three- to six-word target bundles in the MCSS & SSC	141
4.20	Statistically significant overused and underused target bundles in the MCSS	143
4.21	Structural classification of target bundles in the MCSS	145
4.22	Structural classification of target bundles in the MCSS & SSC	147
4.23	The overuse and underuse of noun phrases with post-modifier fragments in the MCSS	149
4.24	Most overused prepositional-phrase fragments in the MCSS	151
4.25	The underuse of first personal pronouns structure in the MCSS	153
4.26	Most overused and underused adverbial clause fragments in the MCSS	157
4.27	Functional classification of target bundles in the MCSS	158
4.28	Functional classification of target bundles in the SSC & MCSS	160

## LIST OF FIGURES

<b>Figure</b>	<b>Page</b>
1.1 The Rational for the Study	7
2.1 A Literature Map of the Research	18
2.2 Distributional Categories (Granger & Paquot, 2008, p. 39)	29
2.3 Schematic representation communicative competence in Celce-Murcia et al. (1995, p. 10)	60
2.4 The visual picture of the focus of the study based on the Celce-Murcia et el's (1995) model	62
2.5 Theoretical Framework of the Study	62
2.6 Conceptual Framework of the Study	63
3.1 Overview of the Procedures of the Present Study	66
3.2 Research Framework of the Study	68
3.3 A Systematic Framework for Corpus Design	70
3.4 AntConc Screenshot Showing the SSC Bundles (Anthony, 2014)	80
3.5 AntConc Screenshot showing the Concordances (Anthony, 2014)	81
3.6 Data Refinement Criteria	82
3.7 AntConc Screenshot Showing the SSC Bundles (Anthony, 2014)	87
4.1 The distribution of three- to six-word target bundles in the SSC	97
4.2 Distribution of structural types of target bundles	104
4.3 Distribution of structural tokens of target bundles	104
4.4 Distributions of functional classification by type	117
4.5 Distributions of functional classification by token	117
4.6 Distributions of research-oriented sub-categories by type	121
4.7 Distributions of research-oriented sub-categories by token	122
4.8 Distributions of text-oriented sub-categories by type	126
4.9 Distributions of text-oriented sub-categories by token	126
4.10 Distributions of participant-oriented sub-categories by type	130
4.11 Distributions of participant-oriented sub-categories by token	130
4.12 Distributions of stance sub-functions by type	132
4.13 Distribution of structural types & functional categories	138

4.14	Distribution of three- to six-word target bundles in the SSC & MCSS	139
4.15	Distribution of structural types of target bundles in the MCSS	146
4.16	Distribution of structural tokens of target bundles in the MCSS	146
4.17	Distribution of structural tokens of target bundles in the SSC & MCSS	148
4.18	Distribution of functional types of target bundles in the MCSS	158
4.19	Distribution of functional tokens of target bundles in the MCSS	159
4.20	Distribution of functional tokens of target bundles in the SSC & MCSS	161
4.21	Example of Complete List of Target bundles	167
5.1	Summary of main findings & pedagogical implications of the study	179

## REFERENCES

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for specific purposes*, 31(2), 81-92.
- Adolphs, S., & Knight, D. (2010). Building a spoken corpus. *The Routledge handbook of corpus linguistics*, 38.
- Akbari, O. (2009). A Corpus Based Study on Malaysian ESL Learners' Use of Phrasal Verbs in Narrative Compositions. (Unpublished doctoral thesis). Serdang: Universiti Putra Malaysia.
- Ali, Z., & Yunus, M. (2004). An ESL writing course: Unravelling students' needs and concerns. *The English Educator Vol. XXXIII*, 2004-1.
- Alipour, M., & Zarea, M. (2013). A Disciplinary Study of Lexical Bundles: The Case of Native versus Non-Native Corpora. *Taiwan International ESP Journal*, 5(2), 1-20.
- Allen, D. (2009). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 1, 105-127.
- Altenberg, B. (1993). Recurrent word combinations in spoken English. In J. D. Arcy (Ed.), *Proceedings of the fifth Nordic Association for English Studies Conference* (pp. 17-27). Reykjavik: University of Iceland.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101-122). Oxford: OUP
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied linguistics*, 22(2), 173-195.
- Amirian, Z., Ketabi, S., & Eshaghi, H. (2013). The Use of Lexical Bundles in Native and Non-native Post-graduate Writing: The Case of Applied Linguistics MA Theses. *Journal of English Language Teaching and Learning*, 5(11), 1-29.
- Anthony, L. (2014). AntConc (Version 3.4. 3). *Computer Software*. Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/software/antconc/> (last accessed May 2015).
- Aronoff, M., & Fudeman, K. (2005). What is Morphology?, *Fundamentals of Linguistics*.

- Arshad, A.S., Mukundan, J., Ghazali, K. Sharifah, Z. Juidah, R., & Edwin, M.V. (2002). *The English of Malaysian School Students (EMAS) corpus*. Serdang: Universiti Putra Malaysia Press.
- Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. (2006). Introduction to Research in Education: Thomson Wadsworth. Belmont, CA.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and linguistic computing*, 7(1), 1-16.
- Atkins, B. S., & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford University Press.
- Atkinson, D. (2004). Contrasting rhetorics/contrasting cultures: Why contrastive rhetoric needs a better conceptualization of culture. *Journal of English for Academic Purposes*, 3(4), 277-289.
- Austin, J. L. (1962). How to do things with words, edited by JO Urmson. New York.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics* (pp. 64-152). Edinburgh: Edinburgh University Press.
- Barlow, M. (2004). Collocate [Computer software]. Houston, TX: Athelstan.
- Basturkmen, H. (2009). Commenting on results in published research articles and masters dissertations in Language Teaching. *Journal of English for Academic Purposes*, 8(4), 241-251.
- Bauer, L., & Renouf, A. (2001). A corpus-based study of compounding in English. *Journal of English Linguistics*, 29(2), 101-123.
- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan: University of Michigan Press.
- Berman, R., & Cheng, L. (2010). English academic language skills: Perceived difficulties by undergraduate and graduate students, and their academic achievement. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquée*, 4(1), 25-40.
- Biber, D. (1990). Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and linguistic computing*, 5(4), 257-269.
- Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243-257.

- Biber, D. (2006). *University language*. Amsterdam: John Benjamins.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275–311.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, 26(3), 263-286.
- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181–190). Amsterdam: Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at .: lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Bin Abdul Kader, M. I., Begi, N., & Vaseghi, R. (2013). A Corpus-Based Study of Malaysian ESL Learners' Use of Modals in Argumentative Compositions. *English Language Teaching*, 6(9), p146.
- Botley, S. P., De Alwis, C., Metom, L., & Izza, I. (2005). CALES: A corpus-based archive of learner English in Sarawak. Final project report, Unit for Research. *Development and Commercialisation, Universiti Teknologi MARA*.
- Braine, G. (1995). Writing in the natural sciences and engineering. *Academic writing in a second language: Essays on research and pedagogy*, 113-134.
- Burrough-Boenisch, J. (2005). NS and NNS scientists' amendments of Dutch scientific English and their impact on hedging. *English for Specific Purposes*, 24(1), 25-39.
- Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5, 31-64.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. *Language and communication*, 1.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied linguistics*, 1(1), 1-47.
- Casanave, C. P., & Hubbard, P. (1992). The writing assignments and writing problems of doctoral students: Faculty perceptions, pedagogical issues, and needed research. *English for Specific Purposes*, 11(1), 33-49.

- Catalán, R. M. J., & Francisco, R. M. (2008). Vocabulary input in EFL textbooks. *Revista española de lingüística aplicada*, (21), 147-166.
- Celce-Murcia, M. (2007). Rethinking the role of communicative competence in language teaching. In *Intercultural language use and language learning* (pp. 41-57). Springer Netherlands.
- Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics*, 6(2), 5-35.
- Chapelle, C. (2001). *Computer applications in second language acquisition*. Cambridge University Press.
- Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30-49.
- Cheng, W., Greaves, C., & Warren, M. (2006). From n-gram to skipgram to conogram. *International Journal of Corpus Linguistics*, 11(4), 411-433.
- Chomsky, N. (1962). *A transformational approach to syntax*. Prentice-Hall.
- Clear, J. (1992). Corpus sampling. *New directions in English language corpora*. Mouton-de-Gruyter, Berlin, 21-31.
- Conrad, S. (1996). Academic discourse in two disciplines: professional writing and student development in biology and history. (Unpublished doctoral dissertation). Northern Arizona University, United States.
- Connor, U., & Mauranen, A. (1999). Linguistic analysis of grant proposals: European Union research grants. *English for specific purposes*, 18(1), 47-62.
- Connor, U., & Upton, T. A. (Eds.). (2004). *Discourse in the professions: Perspectives from corpus linguistics* (Vol. 16). John Benjamins Publishing.
- Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (Vol. 1). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cortes, V. (2002). Lexical bundles in freshman composition. *Using corpora to explore linguistic variation*, 9, 131.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17(4), 391-406.

- Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, 3(1), 43-57.
- Csomas, E. & Cortes, V. (2010). Lexical bundle distribution in university classroom talk. *Language & Computers*, 71, 153-168.
- Cowie, A. P. (Ed.). (1998). *Phraseology: Theory, Analysis, and Applications: Theory, Analysis, and Applications*. Oxford University Press.
- Cowie, A. P. (1999). Phraseology and corpora: some implications for dictionary-making. *International Journal of Lexicography*, 12(4), 307-323.
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 213-238.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, 129-147.
- Crystal, D. (1992). *An encyclopedic dictionary of language and languages*. Oxford: Blackwell Publishers.
- De Cock, S. (2003). Recurrent sequences of words in native speaker and advanced learner spoken and written English. (Unpublished doctoral dissertation). Universit  catholique de Louvain-la-Neuve, Belgium.
- DEMIREL, E. T., & AHMADI, H. S. (2013). Lexical Bundles in Research Article Acknowledgments: A Corpus Comparison. *Hacettepe  niversitesi Eđitim Fak ltesi Dergisi*, 28(28-2).
- De Monnik, I. (2004). Using Corpora and Experimental Data: A Multimethod Approach. In Lejung, Magnus (ed.) *Corpus-based Studies in English. Papers from the 17th International Conference on English Language Research in Computerized Corpora*. Amsterdam: Rodopi, 227-244.
- De Silva, E. (1981). Forms and functions in Malaysian English: the case of modals. *SARE*, 3, 11-23.
- Dong, Y. R. (1998). Non-native graduate students' thesis/dissertation writing in science: Self-reports by students and their advisors from two US institutions. *English for Specific Purposes*, 17(4), 369-390.
- Dontcheva-Navratilova, O. (2012). Lexical bundles in academic texts by non-native speakers. *Brno Studies in English*, 38(2). Retrieved from [http://www.phil.muni.cz/plonedata/wkaa/BSE/BSE\\_2012/BSE\\_2012-38\\_2\\_-XX\\_Dontcheva-Navratilova.pdf](http://www.phil.muni.cz/plonedata/wkaa/BSE/BSE_2012/BSE_2012-38_2_-XX_Dontcheva-Navratilova.pdf)
- Egins, S. (2004). *Introduction to systemic functional linguistics*. A&C Black.
- Ellis, N. C. (1996). Sequencing in SLA. *Studies in second language acquisition*, 18(01), 91-126.

- Ellis, N. C., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics and TESOL. *TESOL Quarterly*, 42(3), 375-396.
- Eriksson, A. (2012). Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In J. Thomas & A. Boulton (eds.), *Input, process and product: Developments in teaching and language corpora*. Brno, Czech Rep.: Masaryk University Press, 195–211.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse*, 20(1), 29-62.
- Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6(1), 3-17.
- Farvardin, M. T., Afghari, A., & Koosha, M. (2012). Analysis of four-word lexical bundles in physics research articles. *Advances in Digital Multimedia*, 1(3), 134-139.
- Ferguson, G. (2001). If you pop over there: a corpus-based study of conditionals in medical discourse. *English for Specific Purposes*, 20(1), 61-82.
- Firth, J. R. (1951). Modes of meaning. *Essays and Studies (The English Association)*, 118–149.
- Flowerdew, J. (1999). Writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8(2), 123-145.
- Flowerdew, J., & Peacock, M. (Eds.). (2001). *Research perspectives on English for academic purposes*. Ernst Klett Sprachen.
- Flowerdew, L. (2013). Corpus-based research and pedagogy in EAP: From lexis to genre. *Language Teaching*, 1-18.
- Flowerdew, L. (2004), 'The argument for using English specialized corpora', in Connor, U. and Upton T. A. (eds), *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins, pp.11–33.
- Fløttum, K., Gedde-Dahl, T., & Kinn, T. (Eds.). (2006). *Academic voices: across languages and disciplines* (Vol. 148). John Benjamins.
- Folse, K. (2004). Vocabulary myths: Applying second language research to classroom teaching. *TESL-EJ*, 14(2).
- Francis, W. N. (1982). *Problems of assembling and computerizing large corpora* (pp. 7-24). na.

- Gläser, R. (1998). The stylistic potential of phraseological units in the light of genre analysis. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 125-143). Oxford: Oxford University Press.
- Grabe, W., & Kaplan, R. B. (1997). On the Writing of Science and the Science of Writing: Hedging in Science Text and Elsewhere. *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts*, 24, 151.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford: Oxford University Press.
- Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective*. John Benjamins Publishing.
- Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. *Phraseology. An interdisciplinary perspective*, 27-50.
- Granger, S., & Rayson, P. (1998). Automatic lexical profiling of learner texts. In S. Granger (Ed.), *Learner English on computer* (pp. 119–131). London: Addison Wesley Longman.
- Gray, J. H., & Densten, I. L. (1998). Integrating quantitative and qualitative analysis: using Latent and manifest variables. *Quality & Quantity*, 32, 419-431.
- Halim, N. B. A., Ahsan, N. I. A., & Munir, F. A. (2011). Writing Final Year Project Report in English: Problems, Motivating and Hindering Factors among UTem Final Year Engineering Students. *Asian Social Science*, 8(1), p27.
- Halliday, M. A. K. (1973). *Explorations in the Functions of Language*.
- Hamp-Lyons, L. (1991). *Assessing second language writing in academic contexts*. Ablex Publishing Corporation, 355 Chestnut St., Norwood, NJ 07648 (clothbound: ISBN-089391-659-5; paperback: ISBN-0-89391-792-3).
- Hamp-Lyons, L. (2011). What Is EAP?. *Handbook of research in second language teaching and learning*, 2, 89.
- Hardy, C., Harley, B., & Phillips, N. (2004). Discourse analysis and content analysis: Two solitudes. *Qualitative Methods*, 2(1), 19-22.
- Harwood, N. (2005). ‘Nowhere has anyone attempted... In this article I aim to do just that’: A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics*, 37(8), 1207-1231.
- Harwood, N., & Petric, B. (2011). English for academic purposes. In Simpson, J. (ed.). *The Routledge Handbook of Applied Linguistics* (pp. 243- 258). London and New York: Routledge.

- Hemmings, B. C., Rushbrook, P., & Smith, E. (2007). Academics' views on publishing refereed works: A content analysis. *Higher Education*, 54(2), 307-332.
- Heng, C. S., Kashiha, H., & Tan, H. (2014). Lexical Bundles: Facilitating University "Talk" in Group Discussions. *English Language Teaching*, 7(4), p1.
- 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(4), 367-383.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Hong, A. L., Rahim, H. A., Hua, T. K., & Salehuddin, K. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis. *3L: Language, Linguistics, Literature®*, 17.
- Howarth, P. (1996). *Phraseology in English academic writing. Some implications for language learning and dictionary making*. Tübingen: Niemeyer.
- Howarth, P. (1998). The phraseology of learners' academic writing. *Phraseology: Theory, analysis, and applications*, 161-186.
- Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English-and Chinese-medium journals. *Journal of pragmatics*, 43(11), 2795-2809.
- Hunston, S. (1995). A corpus study of some English verbs of attribution. *Functions of Language*, 2(2), 133-158.
- Hunston, S. (2002), *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. (2009). The usefulness of corpus-based descriptions of English for learners. *Corpora and language teaching*, 33, 141.
- Hyland, K. (1994). Hedging in academic writing and EAF textbooks. *English for specific purposes*, 13(3), 239-256.
- Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied linguistics*, 17(4), 433-454.
- Hyland, K. (2001a). Bringing in the Reader Addressee Features in Academic Articles. *Written communication*, 18(4), 549-574.
- Hyland, K. (2001b). Humble servants of the discipline? Self-mention in research articles. *English for specific purposes*, 20(3), 207-226.

- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, 7(2), 173-192.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. Routledge.
- Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.
- Hyland, K. (2009). *Academic discourse: English in a global context*. Continuum International Publishing Group.
- Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics*, 32, 150-169.
- Hyland, K., & Hamp-Lyons, L. (2002). EAP: Issues and directions. *Journal of English for academic purposes*, 1(1), 1-12.
- Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183-205.
- Hyland, K. & Tse, P. (2007). Is there an 'academic vocabulary'? *TESOL Quarterly*, 41 (2), 235-254.
- Hymes, D. H. (1972). On communicative competence. In J. B. Pride and J. Holmes (Eds.), *Sociolinguistics: Selected readings*. Harmondsworth: Penguin.
- Imani, A., & Habil, H. (2011). NNS Master's Academic Writing Challenges. *Journal of Edupres*, 1, 42-48.
- Ismail, N., Elias, S., Albakri, I. S. M. A., Perumal, P. D., & Muthusamy, I. (2010). Exploring ESL students' apprehension level and attitude towards academic writing. *The International Journal of Learning*, 17(6), 475-783.
- Jablonkai, R. (2009). "IN THE LIGHT OF": A Corpus-Based Analysis of Lexical Bundles in Two EU-Related Registers. *WoPaLP*, 3, 1-27.
- Jablonkai, R. (2010). English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes*, 29(4), 253-267.
- Jalali, H. (2013). Lexical bundles in applied linguistics: Variations across postgraduate genres. *Journal of Foreign Language Teaching and Translation Studies*, 2(2), 1-29.

- Jalali, H., & Ghayoomi, S. (2010). A comparative qualitative study of lexical bundles in three academic genres of applied linguistics. *Modern Journal of Applied Linguistics*, 2(4), 323-333.
- Jalali, H., Rasekh, A. E., & Rizi, M. T. (2008). Lexical bundles and intradisciplinary variation: The case of applied linguistics. *Iranian journal of language studies*, 2(4), 447-484.
- Jalali, H., Rasekh, A. E., & Rizi, M. T. (2009). Anticipatory 'it' lexical bundles: A comparative study of student and published writing in applied linguistics. *Iranian Journal of Language Studies (IJLS)*, 3(2), 177-194.
- Jenkins, S., Jordan, M. K., & Weiland, P. O. (1993). The role of writing in graduate engineering education: A survey of faculty beliefs and practices. *English for Specific Purposes*, 12(1), 51-67.
- Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.
- Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 269-291). Amsterdam: John Benjamins.
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for specific purposes*, 24(3), 269-292.
- Karabacak, E., & Qin, J. (2013). Comparison of lexical bundles used by Turkish, Chinese, and American university students. *Procedia-Social and Behavioral Sciences*, 70, 622-628.
- Karahan, P. (2013). Self-mention in scientific articles written by Turkish and Non-Turkish authors. *Procedia-Social and Behavioral Sciences*, 70, 305-322.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London and New York: Longman.
- Kim, Y. (2009). Korean lexical bundles in conversation and academic texts. *Corpora*, 4(2), 135-165.
- Koester, A. (2010). Building small specialised corpora. *The Routledge handbook of corpus linguistics*, 66-79.
- Krishnamurthy, R., & Kosem, I. (2007). Issues in creating a corpus for EAP pedagogy and research. *Journal of English for academic purposes*, 6(4), 356-373.
- Labov, W. (1972). *The study of language in its social context*. Penguin Books.
- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. In *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20-34).

- Lee, C. (2012). Using lexical bundle analysis as discovery tool for corpus-based translation research. *Perspectives*, (ahead-of-print), 1-18.
- Leech, G. (1991). The state-of-the-art in corpus linguistics. In B. A. K. Aijmer (Ed.), *English corpus linguistics, Studies in honour of Jan Svartvik* (pp. 8-29). London/New York: Longman.
- Liu, D. (2012). The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes*, 31(1), 25-35.
- Lyons, J. (1977). *Semantics (vols I & II)*. Cambridge CUP.
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing* (Vol. 999). Cambridge: MIT press.
- Mair, C. (Ed.). (1991). *Quantitative or qualitative corpus analysis? Infinitival complement clauses in the Survey of English Usage corpus*. Berlin & New York: Mouton de Gruyter.
- Martin, J. R. (2009). Genre and language learning: A social semiotic perspective. *Linguistics and Education*, 20(1), 10-21.
- McCarthy, M. and O'Dell, F. (2005). *English Collocations in Use*. Cambridge: Cambridge University Press.
- McCormick, D. F. & Whittington, M. S. (2000). Assessing academic challenges for their contribution to cognitive development. *Journal of Agricultural Education*, 41(3), 114-122.
- McCulley, G. A. (1985). Writing quality, coherence, and cohesion. *Research in the Teaching of English*, 269-282.
- McEnery, T., & Wilson, A. (1996). *Corpus linguistics* (Vol. 6). Edinburgh: Edinburgh University Press.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.
- McGrail, M. R., Rickard, C. M., & Jones, R. (2006). Publish or perish: a systematic review of interventions to increase academic publication rates. *Higher Education Research & Development*, 25(1), 19-35.
- Meyer, C. (2002). *English corpus linguistics: An introduction*: Cambridge Univ Pr.
- Milton, J. (2009). *Measuring second language vocabulary acquisition* (Vol. 45). Multilingual Matters.

- Mirzapour, F., & Mahand, M. R. (2012). Hedges and boosters in native and non-native library and information and computer science research articles. *The Southeast Asian Journal of English Language Studies*, 18(2), 119-128.
- Moon, R. (1998). Frequencies and forms of phrasal lexemes in English. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*. (pp. 79-100). Oxford: Clarendon.
- Motta-Roth, D. (1998). Discourse analysis and academic book reviews: a study of text and disciplinary cultures. In I. Fontanet, S. Posteguillo, J. Palmer, & J. Coll (Eds.), *Genre studies in English for academic purposes* (pp. 29-57). Castellon, Spain: Universita Jaume I.
- Mukundan, J., & Aziz, A. (2009). Loading and distribution of the 2000 high frequency words in Malaysian English language textbooks for Form 1 to Form 5. *Pertanika Journal of Social Science and Humanities*, 17(2), 141-152.
- Mukundan, J., bin Saadullah, K. A., & binti Ismail, R. (2013). Malaysian ESL Students' Syntactic Accuracy in the Usage of English Modal Verbs in Argumentative Writing. *English Language Teaching*, 6(12), p98.
- Mukundan, J., Chiew Har, A. L., & Nimehchisalem, V. (2012). Distribution of articles in Malaysian secondary school English language textbooks. *English Language and Literature Studies*, 2(2), 62-70.
- Mukundan, J., & Kalajahi, S. A. R. (2013). *Malaysian Corpus of Students' Argumentative Writing (MCSAW)*. Lulu. com.
- Mukundan, J., & Khojasteh, L. (2011). Modal auxiliary verbs in prescribed Malaysian English textbooks. *English Language Teaching*, 4(1), p79.
- Mukundan, J., & Menon, S. (2007). Lexical similarities and differences in the mathematics, science and English language textbooks. *K@ta*, 9(2), 91-111.
- Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied linguistics*, 10(1), 1-35.
- Nelson, G. (1996). The design of the corpus. *Comparing English Worldwide: The International Corpus of English*, 27-35.
- Nelson, M. (2000). *A corpus-based study of Business English and Business English teaching materials*. Unpublished PhD, University of Manchester, Retrieved November 15, 2005 from <http://www.kielikanava.com/thesis.html>
- Nelson, M. (2010). Building a written corpus. *The routledge handbook of corpus linguistics*, 53.
- Nesi, H., & Basturkmen, H. (2006). Lexical bundles and discourse signaling in academic lectures. *International Journal of Corpus Linguistics*, 11, 283-304.

- Nesi, H., Gardner, S., Forsyth, R., Hindle, D., Wickens, P., Ebeling, S., Leedham, M., Thompson, P. & Heuboeck, A. (2005). 'Towards The Compilation of a Corpus of Assessed Student Writing: An Account of Work in Progress'. Paper presented at Corpus Linguistics Conference, University of Birmingham, UK. Published in the *Proceedings from the Corpus Linguistics Conference Series* 1(1): [www.corpus.bham.ac.uk/PCLC](http://www.corpus.bham.ac.uk/PCLC).
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- Neuendorf, K. A. (2004). *Content Analysis: A Contrast and Complement to Discourse Analysis*. Paper presented at the Qualitative methods: Newsletter for the American Political Science Association Organized Section On Qualitative Methods.
- Norwati, R. (2004). A corpus-based study on an English preposition of place, *at*. In Mukundan, J. et al. (Ed). *ELT Matters 2: Developments in English Language Learning and Teaching*. Serdang: Universiti Putra Malaysia Press.
- Oakey, D. (2002). Formulaic language in English academic writing. In Reppen, R. et al. (eds.). *Using corpora to explore linguistic variation*. Amsterdam: John Benjamins. 111-129.
- O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- Osman, W. H., & Bakar, A. L. A. (2009). Learning to write an academic paper among medical students of Universiti Malaysia Sabah. In *Proceedings of the 2nd International Conference of Teaching and Learning (ICTL 2009) INTI University College, Malaysia* (pp. 1-21).
- Ozturk, I. (2007). The textual organisation of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes*, 26(1), 25-38.
- Palmer, H.E. (1933). *Second interim report on English collocations*. Tokyo: Kaitakusha.
- Partington, A. (2001). Corpus-based description in teaching and learning. *Learning with corpora*, 1000-1022.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards, & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). New York: Longman.
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94.
- Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42, 220-231.

- Quirk, R. (1960). Towards a description of English usage. *Transactions of the philological society*, 59(1), 40-61.
- Rafiee, M. (2013). The Lexical Bundle Approach to Native / Non-native Writing : A Functional Analysis Perspective. *Elixir Literature*. 14326-14330
- Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146-161.
- Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 128-143). London: Longman.
- Reppen, R. (2010). What are the key considerations?. *The Routledge handbook of corpus linguistics*, 31-37.
- Römer, U. (2009). English in academia: Does nativeness matter?. *Anglistik: International Journal of English Studies*, 20(2), 89–100.
- Schmid, H. J. (2003). Collocation: Hard to pin down, but bloody useful. *ZAA*, 51(3), 235-258.
- Salager-Meyer, F. (2008). Scientific publishing in developing countries: Challenges for the future. *Journal of English for Academic Purposes*, 7(2), 121-132.
- Salazar, D. (2010). Lexical bundles in Philippine and British scientific English. *Philippine Journal of Linguistics*, 41, 94-109.
- Salazar, D. (2014). *Lexical Bundles in Native and Non-native Scientific Writing: Applying a corpus-based study to language teaching* (Vol. 65). John Benjamins Publishing Company.
- Saussure, F. De. ([1916] 1959). Course in General Linguistics. Ed. Charles Bally, Albert Sechehaye, and Albert Riedlinger. *Trans. Wade Baskin*. New York: Philosophical Library.
- Schmied, J. (1993). Qualitative and quantitative research approaches to English relative constructions. In Souter, C. and Atwell, E. (eds.) *Corpus Based Computational Linguistics*. Amsterdam: Rodopi.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. *Formulaic sequences: Acquisition, processing and use*, 1-22.
- Scollon, R., & Scollon, S.W. (2001). *Intercultural Communication*. (2nd ed.). Oxford: Blackwell.
- Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education* (Vol. 22). John Benjamins Publishing.

- Seliger, H. W., & Shohamy, E. (1995). *Second language research methods*. Hong Kong: Oxford University Press.
- Shafie, L. A., Maesin, A., Osman, N., Nayan, S., & Mansor, M. (2010). Understanding collaborative academic writing among beginner university writers in Malaysia. *Studies in Literature and Language*, 1(2), 58-69.
- Shahriari Ahmadi, H., Ghonsooly, B., & Hosseini Fatemi, A. (2012). Analyzing research article introductions by Iranian and native English-speaking authors of Applied Linguistics. *International Journal of Research Studies in Language Learning*, 3.
- Silver, M. (2003). The stance of stance: a critical look at ways stance is expressed and modeled in academic discourse. *Journal of English for Academic Purposes*, 2(4), 359-374.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge.
- Sinclair, J. (2005). Corpus and text-basic principles. *Developing Linguistic Corpora: a Guide to Good Practice*. Oxbow Books, Oxford.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214-225.
- Strunkyte, G., & Jurkunaite, E. (2008). Written academic discourse: Lexical bundles in humanities and natural sciences. *Vilnius University, Vilnius*.
- Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, 7(2), 215-244.
- Stubbs, M. (2007a). Quantitative data on multi-word sequences in English: The case of the word world. *Text, discourse and corpora: Theory and analysis*, 163-189.
- Stubbs, M. (2007b). An example of frequent English phraseology: distributions, structures and functions. *Language and Computers*, 62(1), 89-105.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
- Swales, J. M. (2004). *Research Genres: Exploration and Applications*. Cambridge: Cambridge University Press.

- Swales, J. M. (2008). Foreword. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (p. v- viii). Ann Arbor: University of Michigan Press.
- Tang, R., & John, S. (1999). The 'I' in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes, 18*, S23-S39.
- Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied linguistics, 22*(1), 58-78.
- Thornbury, Hyland S. (2002). *How to teach Vocabulary*. Essex: Longman.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*: John Benjamins Publishing Co.
- Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for specific purposes, 20*(1), 83-102.
- Vaughan, E., & Clancy, B. (2013). Small corpora and pragmatics. In *Yearbook of Corpus Linguistics and Pragmatics 2013* (pp. 53-73). Springer Netherlands.
- Vázquez Orta, I. (2010). A contrastive analysis of the use of modal verbs in the expression of epistemic stance in business management research articles in English and Spanish. *Ibérica, (19)*, 77-95.
- Vethamani, M. E., Manaf, U. K. A., & Akbari, O. (2008). Esl Learners Use of English Modals in Narrative Compositions: Syntactic and Semantic Accuracy. *TEFLIN, 19*(2).
- Wei, Y., & Lei, L. (2011). Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC Journal, 42*(2), 155-166.
- Wilkins, D. A. (1976). *Notional syllabuses: A taxonomy and its relevance to foreign language curriculum development*. Oxford Univ Pr.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics, 463-489*.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, New York: Cambridge University Press.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication, 20*(1), 1-28.
- Xiao, R., McEnery, T., & Qian, Y. (2006). Passive constructions in English and Chinese: A corpus-based contrastive study. *Languages in contrast, 6*(1), 109-149.

Zamel, V (1998). Questioning Academic Discourse. In V. Zamel and R. Spack (Eds.), *Negotiating academic literacies: Teaching and learning across languages and cultures* (pp. 187-197). Mahwah, NJ: Erlbaum.

