

An Analysis of Lexical Chunks in Online Business Letters (OBL) And Business Letter Corpus (BLC): A Corpus-Based Study

Hairul Azhar Mohamad¹, Muhammad Luthfi Mohaini¹, Pavithran Ravinthra Nath¹

¹Akademi Pengajian Bahasa, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia
hairazhar@uitm.edu.my

¹Akademi Pengajian Bahasa, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia
luthfimohaini@uitm.edu.my

¹Akademi Pengajian Bahasa, Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia
pavithran@uitm.edu.my

Article history:

Received: 9 June 2020 Accepted: 16 June 2020 Published: 15 July 2020

Abstract

This research investigated into the lexical density and frequencies of five types of lexical chunks located in 300 online business letters. Top 10 websites on business correspondence had been identified in terms of traffic visitors and bounce rate under one million web rankings worldwide. Criterion Sampling method was identified prior to extracting the sample letters from the websites. The data was then run with Antconc Concordance Program (ACP) for lexical density and frequency analysis. Top 15 lexical chunks in online business letters (OBL) were compared against those top 15 in Business Letter Corpus (BLC). Findings revealed that there was a total of 39 916-word tokens and 939 counts of lexical chunks found in this corpus. It was found that more lexical words do not imply more lexical chunks used in based on types of business letters. All 5 types of lexical chunks were identified and ranked in descending order; Sentence Builders (SB) as the most frequent type, followed by Collocations (COL), Deictic locutions (DLs), Polywords (POLs) and Institutionalized Expressions (IUs) as the least frequent type of lexical chunk. Subdivisional analysis indicated that Grammatical Collocations (GCs) were more common than Lexical Collocations (LCs). Majority of lexical chunks were formed more at sentence level than phrasal level. Comparative analysis between top 15 lexical chunks in OBL and BLC discovered that most top lexical chunks in online business letters are representative of those corporate business letters in BLC. Pedagogical implications in terms of the reliability of online business letters for academic reference and future research considerations are also addressed.

Keywords: *Lexical chunks, business letters, corpus studies, corpus linguistics, online business letters (OBL), Business Letter Corpus (BLC)*

Introduction

Business correspondence is part and parcel of a company's communication machinery and business letters form the standard tool the business community uses to deliver important messages to their clients and customers. To be effective, the writing of business letters needs to adopt the business language. This refers to the "unique expressions" of business correspondence which is further identified as business lexical chunks in the linguistic perspective. Lexical chunks (or also known as lexical phrases, lexical bundles, formulaic sequences or expressions, or multi-word expressions) are any combinations of two or more words specifically memorized in chunks and processed conjointly like single words, to be utilised automatically and recurrently in a given context (Lewis, 1993, as cited in Lindstromberg, 2000). It can be highly fixed such as "on behalf of", "as soon as" or relatively flexible with slots and fillers such as "I/We look forward to seeing/meeting/hearing from you". Some combinations have been used so extensively and repeatedly that

they become idioms or idiomatic expressions, for instance “when there is a will, there is a way” or quotations such as “a journey of a thousand miles begins with a first step”. Lexical chunks are pivotal in our business correspondence in verbalisation or composition for various reasons. These include oral communication skills for subconscious regurgitation of lexical chunks as and when needed (Wood, 2004), vocabulary acquisition for memorisation of more than one word or in bundles (Peters, 1983, cited in Schmitt & Carter, 2000) and grammar learning for its route learning of the coordination of phrases according to their rules which is eventually termed by Islam and Timmis (2003) as our grammatically annotated phrasebook.

Problem Statement

The advent of computer-based analysis gave an impetus to the various studies in the area of corpus linguistics on lexical chunks in spoken discourses (Osman and Jusoff; 2009 and Yeganehjoo & Yap, 2012) and academic business texts (Allen, 2009; Durrant & Schmitt, 2009 & Salazar, 2011) in the form of emails (Pastor & Calderón, 2012) and business letters (Xu, 2010; Sinturat, 2010 and Qiong, 2013). They underlined the importance of lexical chunks in order to produce quality business letters in the classroom environment. Sinturat (2010) and Muckian & Woods (2012) were prompted further to study the use of lexical chunks in the writing of business letters sampled from handbooks and textbooks. This aimed to help students to identify the right phrases to use in their business letters. However, Hargittai, Fullerton, Mechen-Trevino and Thomas (2010) and Purdy (2012) pointed out that in the learning of business correspondence, students at universities have the tendency to refer to materials and resources online, for instance by simply referring to business letters available in websites or any online sites as one of their main sources of reference. Unfortunately, some of these business letters were loosely taken from any websites and used as one pleases as their samples for their academic reference or teaching purposes as many studies have shown (Simpson-Vlach & Ellis, 2010; Flowerdew, 2012). This has raised another concern if the online business letters samples appropriately and sufficiently reflect the business letters used in the corporate sector. Hence, studies on lexical chunks of online business letters which are available on various websites are still found to be in need for further investigation.

Objectives of the Study

The aim of the present study is to investigate the representativeness of lexical chunks in online business letters. To do this, frequency analysis of lexical chunks from this domain to find the most frequent types and their contexts of occurrence, and comparative analysis between lexical chunks in online business letters and Business Letter Corpus (BLC) were conducted.

Research Questions of the Study

In order to achieve the above two objectives, two (2) research questions have been correspondingly developed as follows.

- i. What are the frequency counts of five (5) types of lexical chunks in various types of online business letters extracted from top ten (10) business letter online sites and their context of occurrence?
- ii. In what ways the top fifteen (15) lexical chunks in online business letters against the top fifteen (15) lexical chunks in business letters from Business Letter Corpus (BLC)?

Significance of the Study

The present research is significant in the context of academic writing, business correspondence and business letters. The final outcome would be a list of lexical chunks or expressions from business letters taken from one million ranking websites with the highest traffic volume on business correspondence. It can be an additional resource as reference for courses related business English, business writing and letter writing or potentially added as part of the manual or handbook for these courses. Apart from that, the findings would be able to validate if the online business letters taken by students or academic instructors are valid for their reference and if so, what are the requirements needed before they can take those as their samples.

Literature Review

History and Studies of Lexical Chunks

In 1933, Palmer (1933) in Xu (2010) raised his concern that collocational phrases must be learnt by heart or even best learnt as an integral unit or an independent whole instead of piecing the component parts together individually. The idea laid dormant for decades until 1970s when Becker (1975) emerged with his analysis on this subject. He categorically pointed out that we were more likely to find these phrases in a structure of two or three words joint concurrently together for example, 'so far', 'in fact' and 'so as to' in most of human discourses; written or spoken. These analyses led to a series of other textual analyses on lexical chunks further onwards. Eventually, a lexical chunk approach proposed by Lewis (1993) experienced its glory in 1990's which became a new pedagogical approach in the learning and teaching of English language. This method relied heavily on the focus of word combination and the learner's ability to attach its importance and memorise it to be part of their corpus (Willis, 1994).

Zhao (2009) studied the production of lexical chunks through a number of tests which included multiple choice, an interview and writing test and found that those with higher repertoire of lexical chunks obtained showed better performance in their writing test scores. Similarly, a study conducted by Kim (2019) among university students taking Practicum of Practical English program revealed that by emphasising on lexical chunks, the students were able to improve their language skills which led to higher scores in their TOEIC

test. The findings revealed that English learners were more receptive towards the strategy of constructing sentences using prefabricated chunks instead of the traditional word-by-word comprehension. In view of the above, Chao (2016) proposed that lexical approach must be considered as a new method in order to enhance students' writing skills in class since students were found not to be aware of the importance of lexical chunks in writing business letters prior to the lesson given in class. This is concurred by Wright (2019) who found the existence of "a core set of lexical bundles" to be used for academic writing and written academic prose alone.

Corpus Studies in Various Texts (i.e. Research Articles, Textbooks and Business Correspondence)

Corpus Linguistics is a branch of linguistics and according to Bennett (2010), it refers to the study on the language use through corpora. It must contain a huge amount of naturally occurring words, expressions and texts stored electronically for reference and research purposes which helps to answer two basic purposes; firstly, identifying patterns associated with grammatical and lexical features and secondly pattern differences in different varieties and registers. Salazar (2011) conducted corpus-driven study to compare the lexical chunks between native and non-native scientific texts in English in order to generate an inventory of most frequent and common lists of lexical chunks to be used for pedagogical purposes. Two important points in the study were highlighted about non-native against the native style of writing which include lacking variations and producing unnecessary repetitions of phrases. Ang and He (2017) also adopted a corpus-based approach in identifying lexical chunks used in academic essays written by Asian and British native students where these students were found to prefer different types of lexical chunks even though there was no significant difference in the functional use. Corpus analysis of research articles written by Turkish scholars were also investigated in terms of the use of lexical chunks by Betul (2010) and it was revealed based on the search of 4-word lexical chunks that these phrases emerged 20 times in more than one million words and these lexical chunks were analysed structurally and functionally according to the two significant taxonomies of lexical chunks as proposed by Biber, Johansson, Leech, Conrad & Finegan (1999) and Biber, Conrad and Cortes (2004). This general trend of 4-word lexical chunks has been highlighted as the most frequent target structures (Appel & Murray, 2020) and most repeated identified word sequences (Appel & Trofimovich, 2017). It was found that there were strong functional-lexical features and structural patterns which can help the scholars in constructing their research writing and direction of their academic writing. Meanwhile, Ang and Tan's (2018) study of academic writings corpus that looked into journal articles in the field of International Business Management (IBM) revealed the preference towards 3-words lexical chunks which is in agreement with Simpson-Vlach and Ellis' (2010).

Xiao (2011) conducted a corpus study on the representativeness of lexical chunks in 40 sample business letters (i.e. letter of enquiry, complaint, making counter offer and shipment) from some foreign trading

companies and from the letters written by students majoring in International Trade at Ocean University of China based on Nattinger and DeCaricco's (1992) classification model. More lexical chunks were found written by those of the foreign trading companies than the students and there was a significant use of lexical chunks in both groups' English business letters. This is also concurred by another finding on the high frequency use of lexical chunks through a corpus-based study done by Qiong (2013) on the use of lexical chunks in 50 business English letters from higher vocational college business students majoring in China. Similar to those attempts done by Xiao's (2011), the researchers studied the characteristics of lexical chunk and the relationship between these lexical chunks. Sentence builders were discovered to dominate in all three types of lexical chunks, followed by polywords and institutionalized expressions. Students with higher proficiency levels applied more lexical chunks in their business correspondence.

Finally, a marriage between business letters and textbooks were done by Sinturat (2010) in a study to find the top-30 lexical chunks in Business Letter Corpus (BLC) and five Thailand universities' textbooks via Antconc Concordance Program. It was found that the top 30 lexical chunks in BLC were represented by 5 collocational phrases and 25 multi-word-expressions (i.e. Phrasal Constraints, Poly-words and Sentence Builders). As for the textbooks, there were 10 collocations and 20 multi-word expressions in their business letters. The prevalence of top and high frequency lexical chunks in various business documents validated the point on sub-lingual function as proposed by Someya (1999) in the writing of business letters. In addition, a nuanced feature of business sub-language is manifested in the use of phrasal verbs, being one type of grammatical collocations, which were found to be more dominant in the BLC while lexical collocations more dominant in the textbooks. All of the above studies highlighted the importance of lexical chunks in scientific and research articles (Betul, 2010; Salazar, 2011; Farvadin, Afghari and Koosha,2012) numerous schools' and universities' textbooks (Hsu, 2006; Sinturat, 2010; López-Jiménez, 2013; Mukundan & Menon, 2008 & 2012;) and business correspondence from academic and corporate settings (Someya, 1999; Sinturat, 2010; Xiao, 2011; Pastor and Calderón's, 2012; Romova & Hetet, 2012; Hou, 2014;) and the prominent need for its usage to produce effective texts.

Methodology

Quantitative approach and descriptive statistics were adopted to present the research on 300 online business letters extracted from top 10 websites in terms of lexical density, frequencies, lexico-grammatical feature, sentence and non-sentence constructions. Top identified lexical chunks were then compared against those in Business Letter Corpus (BLC).

Research Design

An Analysis of Lexical Chunks in Online Business Letters (OBL) And Business Letter Corpus (BLC): A Corpus-Based Study

The present study adopted quantitative approach with descriptive research design. Under the guide of corpus linguistics, frequency of 5 types of lexical chunks from 300 online business letters were counted and converted into percentages. Top 15 lexical chunks from these 300 online business letters (OBL) were identified for further comparative analyses in terms of rankings with the top 15 lexical chunks from Business Letter Corpus (BLC) which is readily available to be extracted from the BLC official site.

Sampling/ Participants/ Respondents

Since this is a content analysis study, no participants were involved. Three hundred (300) sample business letters were selected from a list of top traffic websites on business letters. Through the assistance of top 20 websites (www.top20sites.com) on various subjects, and the list of websites on business correspondence with most traffic was identified. The traffic of top 20 websites on business letters were given validation and confirmation in terms of their traffic volume and bounce rate via Alexa, the Web Confirmation company (www.alexa.com) by entering their website addresses and the results were identified based on their significant bounce rate (of approximately 50% or more), significant search traffics in terms of percentage (20% or more), both paid and free, that have visited these sites from various search engines for the last three months from the cut-off date set by the present researchers on 31 December 2015 and also high total number of word tokens from the selected business letters (i.e. 3000 or more). It is important to note that standard format and business language at international and corporate level in the early 21st century in line with Globalisation age 3.0 have thus far experienced a major progress with only a nuanced linguistic variation (Evans, 2012; Linn, Sanden & Piekkari 2018). Furthermore, the three criteria were selected based on three important criteria, namely *High Performance Search Traffic Test of the Site*, *Structure of Information* and *Information Focus of the Sites* determined by Fogg and Marable (2003) in their study of websites credibility. Table 1 displays the list of ten (10) websites which had been finalised from twenty websites.

Table 1: List of Ten (10) Websites

	Websites Address	Ranking within 1 million	Bounce Rate*	Search Traffic**	Traf-	Word To-
1.	www.letters.org	51,380	59.90%	24.00%		3211
2.	www.samplelettersfree.org	126,734	63.80%	29.20%		4101
3.	www.samplelettersworld.com	244,927	54.40%	26.60%		3162
4.	www.letters-home.com/	256,306	46.10%	22.00%		3657
5.	ww.officewriting.com	258,827	70.50%	24.80%		4001
6.	www.isampleletter.com	267,545	62.70%	29.40%		3861
7.	www.sampleletters.net/	390,466	72.30%	28.90%		3847
8.	www.sampleletters.org/	665,747	52.70%	23.10%		3859
9.	www.businesslettertemplate.info/	667,273	74.70%	20.70%		3495
10.	www.sampleletters.in/	953,456	49.10%	20.80%		4312

*Bounce Rate is the percentage of visits to the site that consist of a single page view. Significant bounce rate is from 50.00 percent and above.

***Search Traffic is the percentage of traffic, both free and paid, that come to this site from a search engine over the past 3 months, updated daily. The change number shows the difference versus the previous 3-month period. Significant traffic is 20.00 percent and above.*

In terms of choosing the 300 business letters, a Criterion Sampling method was utilised, which is one of the sub-types of Purposeful Sampling (Patton, 2011). Criterion Sampling is done when cases are selected once they meet the predetermined set of criteria (i.e. 6 types of letters, 100-150 words, removal of other irrelevant information such as address, salutations etc).

Materials and Procedure

Afterwards, the data were run with another software Antconc Concordance Program (ACP) which is available for free download online at www.someya-net.com/concordancer. It is a freeware developed by Prof. Laurence Anthony, a researcher and director of the Centre for English Language Education, from Waseda University of Japan which was used by Krajka (2007) on frequency search analysis, Sinturat (2010) on frequency counts of lexical chunks in sample letters from academic textbooks and Hou (2014) on frequency analysis of cover letters.

Then, the top 15 lexical chunks identified from the present research samples were compared with the top 15 lexical chunks of business letters from Business Letter Corpus (BLC) which can be obtained from <http://ysomeya.hp.infoseek.co.jp>. All top 15 lexical chunks or phrases from both sources came in various count of words/ phrases which include 5-gram (5-word), 4-gram (4-word) and 3-gram (3 word) phrases. It was anticipated that many similarities would be discovered between both sources for two types of lexical chunks/ phrases; Collocations (COL) and Sentence Builders (SB).

Data Collection and Analysis

Therefore, for the present research to be completed, it required to go through two phases of data collection and data analysis; identification of lexical chunks in OBL based on adapted classifications from two main models of lexical chunks (Becker, 1975 & Lewis, 1997) and secondly, a comparison between OBL and BL.

1. **Polywords (POL):** Phrases functioning as single words, for example “for good” for “forever”, “by accident” for “accidentally”, ‘in order to” for “with the purpose of”, “so as to” for “so that” and “in the process/ in the midst of” for “currently doing something”.
2. **Collocations (COL):** Phrases in pair or groups of word which occur concurrently at a very high degree of frequency together. There are two types of collocations; a) *Grammatical Collocations* where noun/verb/ adjectives are followed by Prepositions/ grammatical structure (e.g. prefer for, adjacent to, look forward to, put off) and prepositions are followed by possessive determiner and noun (e.g.

An Analysis of Lexical Chunks in Online Business Letters (OBL) And Business Letter Corpus (BLC): A Corpus-Based Study

on/from their/her/his behalf, on your/my own), and b) *Lexical Collocations* where verb is combined with noun (e.g. seek assistance), noun with verb (e.g. storms rage), adverb with adjective (e.g. extremely tired), verb with adverb (e.g. know well), noun with noun (e.g. a world record) and adjective with noun (e.g. a good self or a short strategy).

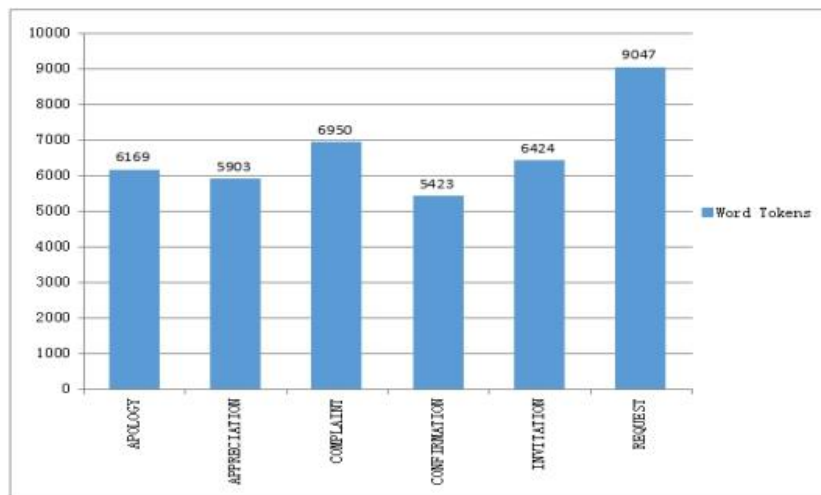
3. **Deictic Locutions (DL):** Phrases which allow nearly low changeability or flexibility which can be clauses such as, “In fact, ...” (which means ‘I would like to add further points), “for that matter, ...” (which means ‘I just thought a better way of making my argument) or or entire sentences so as to direct a person in their written or spoken discourse such as “that-clause”, “We/I + Verb to Be + that” or “I/We + Verb + that ...”.
4. **Sentence Builders (SB):** Sentence-long expressions like a slot and filler which functions as a frame or skeleton for the entire sentences such as “Should you have any enquiries/ questions/ problems...”, “Please/ Kindly contact our person in-charge at...”, “We would like to apologise/ request/ invite...” or “Thank you for memo/ letter/ call/ complaint dated... which usually occur the beginning or end of business letters.
5. **Institutionalised Utterances (IU)/ Fixed Expressions (FE):** Sentence-like clauses or phrases with no variation allowed in written and spoken discourse, common ins social exchanges such as “Thank you so much”, “Thank you in advance”, “Good Morning” and “Have a good day!”.

Findings

The present research introduces the word tokens to represent the symbolic number of words which serve as the tangible representation of the text. Every word token is tantamount to every counted word. The first research question and second research question with the data are addressed thereafter respectively.

Total Word Tokens

Figure 1: Total Word Tokens



As indicated by Figure 1 above, all 6 types of letters give a total of 39 916-word tokens in which each type of letter has an average of 6 652 tokens. Request Letter was found to top the list with 9 047-word tokens. There are not many differences of word tokens for three types of letters; Complaint, Invitation and Apology which range from 6 169 to 6 950 words. Letter of Appreciation and Confirmation have been ranked the least from 5,903 to 5423-word tokens. To briefly conclude, business letters with more word tokens do not indicate more lexical chunks and vice versa.

Research Question No. 1: What are the frequency counts of five (5) types of lexical chunks in various types of online business letters extracted from the top ten (10) business letter online sites and their context of occurrence?

The present researcher identified the five (5) type of lexical chunks from Online business Letters in the form of tables for 5-gram, 4-gram and 3-gram frequency counts according to their types of letters; Apology, Appreciation, Complaint, Confirmation, Invitation and Request. More than one type could be found in one 5-gram, 4-gram and 3-gram tables, thus the data is analysed and presented in the form of frequency counts and types of lexical chunks (LB) for Type 1 (i.e. the first identified type) and Type 2 (i.e. the second identified type) for them to be subsequently tabulated into the overall table for a comprehensive view. There are three parts to the analysis; first part is the frequency counts for all types of LBs, second part is frequency counts for one type of LB that is Collocations (COL), be it Grammatical Collocations (GCs) or Lexical Collocations (LCs) and third part is whether the lexical chunks are formed at sentence or word levels. Below are extracts of 3-, 4- and 5-gram tables for all three-part analyses for Apology letters for illustration.

TABLE 2: Apology Letters (3gram, 4gram, 5gram)

RANK	FREQ.	3-GRAM	LB TYPE 1	LB TYPE 2	TYPES OF COLLOCATIONS	
1	16*	would like to	SB	NA	NA	NA

An Analysis of Lexical Chunks in Online Business Letters (OBL) And Business Letter Corpus (BLC): A Corpus-Based

Study						
2	11*	Thank you for	SB	COL	Grammatical	NA
3	8*	assure you that	DL	NA	NA	NA
4	6	accept our apology	COL	NA	NA	Lexical
5	6	in the future	DL	NA	NA	NA
6	6	sorry for the	COL	NA	Grammatical	NA
7	5	apologies for the	COL	NA	Grammatical	NA
8	5*	I assure you	DL	NA	NA	NA
9	5*	I hope you	DL	NA	NA	NA
10	4	apologize for my	COL	NA	Grammatical	NA
11	4	as soon as	POL	NA	NA	NA
12	4	in order to	POL	NA	NA	NA
13	4	on behalf of	POL	NA	NA	NA
14	4	our apology for	COL	NA	Grammatical	NA
15	4	the delay in	COL	NA	Grammatical	NA
16	4*	This is to	SB	NA	NA	NA
17	3	accept my apology	COL	NA	NA	Lexical
18	38	am writing to	SB	NA	NA	NA
19	3	an excuse for	COL	NA	Grammatical	NA
20	3	and hence I	DL	NA	NA	NA
21	3*	and we hope	DL	NA	NA	NA
22	3	because of the	POL	NA	NA	NA
23	3*	I know that	DL	NA	NA	NA
24	3*	I understand that	DL	NA	NA	NA
25	3*	it would be	SB	NA	NA	NA
26	3	look forward to	COL	SB	Grammatical	NA
27	3*	regret to inform	SB	NA	NA	NA
28	3	so that we	POL	NA	NA	NA
29	3	thank you for	COL	NA	Grammatical	NA
30	3*	understand that you	DL	NA	NA	NA
31	3*	we hope that	DL	NA	NA	NA

RANK	FREQ.	4-GRAM	LB TYPE 1	LP TYPE B	TYPES COLLOCATIONS	OF
1	5*	Please accept our apology	SB	NA	NA	NA
2	4*	am writing this letter	SB	NA	NA	NA
3	4*	I am grateful that	DL	NA	NA	NA
4	3*	I am writing to	SB	NA	NA	NA
5	3*	The thing is that	DL	SB	NA	NA

RANK	FREQ.	5-GRAM	LB TYPE 1	LP TYPE B	TYPES OF COLLOCATIONS	
1	3*	am so sorry to have	SB	NA	NA	NA
2	3*	is not an excuse for	SB	COL	NA	NA
3	3*	was brought to my at-tention	SB	COL	Grammatical	NA
TOTAL					46	9

*Indicators**LC: Lexical chunks**SB: Sentence Builder**COL: Collocation**POL: Polyword**DL: Deictic Locution**IU: Institutional Utterances**Marked with Asterisk: Counts of Sentence-based LBs*

Highlighted Phrases: The second type of lexical chunks

Important Note:

1. "NA" is "Not Applicable" due to its singular type of lexical chunks (LP Type 1 only), hence LP Type 2 is not identified.
2. "NA" under "Types of Collocation" means "Not Applicable" as Collocation is the only type of LB with further sub-categories; either Grammatical OR Lexical.

Table 2 shows the first-part analysis on frequencies of lexical chunks in 5-gram, 4-gram and 3-gram tables for letters of apology in which 189 counts of lexical chunks were found. As to the second-part analysis in terms of the division of collocations (COL) into two; Grammatical Collocation (GC) and Lexical Collocation (LC), forty-six (46) or 70 percent GCs were prevalent compared to only nine (9) or 30 percent LCs. Furthermore, no LCs were found in the generated 4-gram and 5-gram tables. In terms of the sentence- or word-based level analysis which also forms the third-part analysis, 40 of all 49 Deictic Locutions were found to be sentence-based in addition to 67 Sentence Builders which are sentence-based by form.

The same analyses for 3-gram, 4-gram and 5-gram lexical chunks in terms of frequency counts for all 5 LBs (including Type 1 and Type 2) were performed for the remaining five types of letters; appreciation, complaint, confirmation, invitation, request letters. The findings from all 6 types of letters were then tabulated into the overall table of frequency counts for all five (5) types of lexical chunks for Six (6) Types of Letters for. The final overall table of frequency count was presented in Table X as follows.

TABLE 3:
Comprehensive Table of Frequency Counts for Five (5) Types of Lexical Chunks and Six (6) Types of Letters

Types of Lexical Chunks and Types of Business Letters	Sentence Builder	Collocation	Poly word	Deictic Locution	Fixed Ex-pressions/I.U.*	LBs TOTAL COUNT	Lexical chunks Rank	Word Token Ranks
Apology	67	55	18	49	0	189	2	4
Appreciation	72	61	0	3	0	136	4	5
Complaint	74	51	12	23	0	160	3	2
Confirmation	64	39	20	3	0	126	5	6
Invitation	54	42	3	6	0	105	6	3
Request	118	47	31	14	5	215	1	1
TOTAL COUNTS	449	295	84	98	5	939		
RANKS	1	2	4	3	5			

Table 3 shows the comprehensive frequency counts for all types of LBs for 6 types of letters. First, it is important to highlight the total counts of lexical chunks (LBs) and word tokens in the last two columns of the table for the purpose of comparative analysis. It was found that Letters of Request was ranked on the top of list of word tokens, followed by Letters of Complaint ranked as the second, Letters of Invitation as the third, Letters of Apology as the fourth rank and finally Letters of Appreciation as the fifth in terms of the to-

An Analysis of Lexical Chunks in Online Business Letters (OBL) And Business Letter Corpus (BLC): A Corpus-Based Study

tal number of word tokens. Letters of Confirmation generated the least number of word tokens and placed at the sixth rank. Comparative analyses revealed that Letters of Request have the highest LBs, followed by Letters of Apology as the second highest. Letters of Complaint were ranked the third, followed by Appreciation which was ranked the fourth and Confirmation the fifth. Letters of Invitation did not score the lowest word token counts but were found to have the least counts of LBs. From this comparative analysis, it was concluded that total word tokens did not correspond with the total counts of lexical chunks (LBs).

Figure 2:

Total Frequency Counts and Percentage for 6 Types of Lexical Chunks

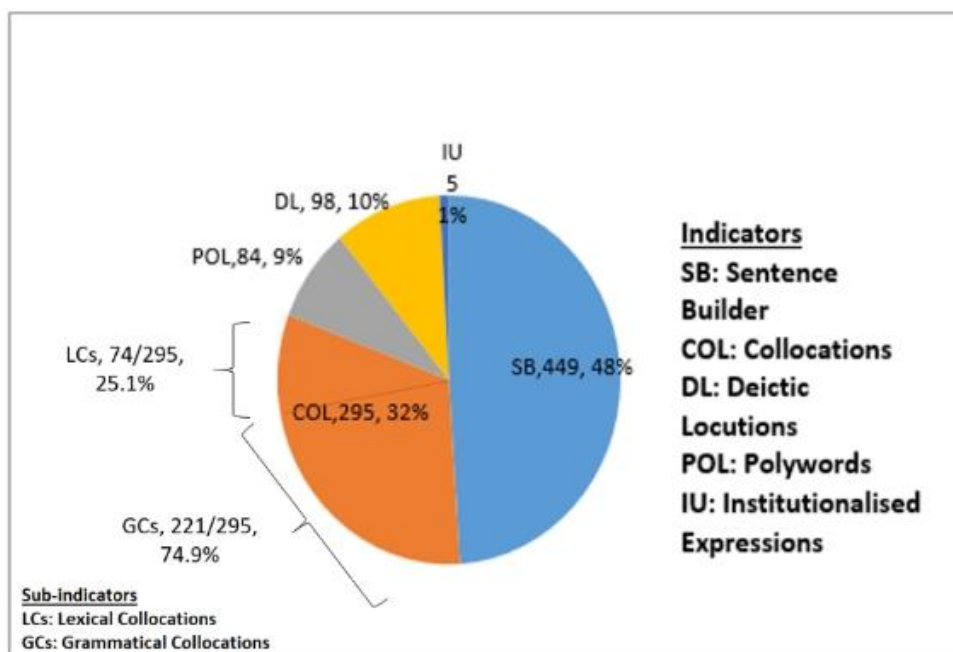


Figure 2 shows the total LB counts in terms of percentages for all six types of LBs. The findings revealed that 449 counts of Sentence Builders (SBs) with the highest counts of LBs accounted for 48 percent as compared to other types of LBs. Next, Collocations (COL) with a total of 295 counts or 32 percent took the sec-

ond biggest proportion of the chart. The third biggest category was represented by Deictic locutions with 10 percent or 98 counts. Finally, Polywords with 9 percent (84 counts) and Institutionalized Expressions with 1 percent (5 counts) formed the last two smallest proportions. Overall, Sentence Builders (SBs) made up the biggest type while Institutionalised Utterances (IUs) formed the smallest type of LBs.

In terms of the analyses of the division into Grammatical (GCs) and Lexical Collocations (LCs), the former represented the most with 221 counts or 74.9 percent, while the latter represented the remainder with only 74 counts or 25.1 percent from the total 295 Collocations found. Based on this, it was concluded that 75 percent or 75 percent were more grammatical than lexical in all six (6) types of letters. As for the analyses of word/phrase against sentence-based level, 79 from 98 Deictic Locution counts or 80.6 percent were found to be dominated by sentence-based LBs. As for overall 939 LBs (i.e. DLs, SBs and IUs all included), 533 or more than 58 percent of them operated at the sentence-based than phrase-based level.

Research Question No. 2: In what ways the top fifteen (15) lexical chunks in online business letters against the top fifteen (15) lexical chunks in business letters from Business Letter Corpus (BLC)?

Table 4:
Comparative Listings of Top 15 Lexical Chunks between Business Letter Corpus (BLC) and Online Business Letters (OBL)

An Analysis of Lexical Chunks in Online Business Letters (OBL) And Business Letter Corpus (BLC): A Corpus-Based Study

No.	BLC Lexical Chunks	Total Count	Types	Online BL Lexical Chunks	Total Counts	Types
1	Thank you for	648	SB	Subject + would like to + Verb	103	SB
2	look forward to	991	COL	Thank you + Adverb + for + Noun	95	SB
3	Subject + look forward to + Verb-ing	750	SB	Sub + hope + that/Subject	85	DL
4	Subject + would like to + Verb	675	SB	look forward to	57	COL
5	If you have + Noun/Article	442	SB	Subject + Verb To Be + writing + to + Verb	48	SB
6	hear from	383	COL	Subject + request + you + to	45	SB
7	as soon as	327	POL	Subject + look forward to + Verb-ing	43	SB
8	I/We + hope	221	DL	Subject + assure you that	30	DL
9	as well as	284	POL	so that	22	POL
10	Please let + Subject + Verb	231	SB	feel free to	20	COL
11	in the future	207	DL	as soon as	19	POL
12	I/ We am/are sure	169	DL	on behalf of	18	POL
13	Please call + Subject	169	SB	This is to + Verb	16	SB
14	Possessive Noun + interest in	164	COL	Apologize for	16	COL
15	If there is + Article/ Noun	161	SB	Please feel free to + Verb	16	SB

Table 4 illustrates a comparative analysis of the lexical chunks between Business Letter Corpus (BLC) and researchers' online business letters. The ranking of lexical chunks phrases identified in these two sources from the highest to the lowest in terms of frequency counts are displayed side by side in the table. It was discovered that all top five (5) Sentence Builders (SB) identified and reported in previously in research question no. 1 for all six (6) types of letters were also found to be listed in the top fifteen (15) lexical chunks in Online Business Letters (OBL).

Several other important findings show that seven out of all 15 lexical chunks generated in both sources generated were Sentence Builders while the remaining eight were the other three types of lexical chunks; Collocations, Deictic Locutions and Polywords. As highlighted in the boxes above, both sources indicated

three (3) similar Sentences Builders. Second important finding is that “look forward to” was the highest Collocation (COL) bundle for both sources and listed in the fourth place in the Online Business Letters (BLC) under examination compared to the second place in the reference source; Business Letter Corpus with only one place difference. Altogether, three (3) Collocation bundles had been generated in both sources which indicated their similar significance level in use. Thirdly, similar Deictic Locution bundle of “Subject + hope + that” was also found in both sources which also marked its paramount importance. In terms of the highest frequency counts, this bundle was likewise found to have been listed as the first DL in both sources. Fourth point to be highlighted was the ranking of Polyword phrase of “as soon as” in which this bundle was represented in both sources as well.

For the final part of the comparative analysis, it was found that both sources showed little difference in the identified types of lexical chunks. There were seven Sentence Builders (SBs) and three Collocations (COLs) found in both BLC and OBL sources with a total of ten (10) out of the top fifteen (15) lexical chunks. As for the remaining five (5) top lexical chunks, it was found that there were two (2) Polywords in Online Business Letters while there existed three (3) Polywords in BLC. Finally, three Deictic Locutions existed in OBL while two (2) in BLC. Likewise, the remaining top 5 lexical chunks for both sources indicated only one minor variance in terms of their types.

It can be comparatively concluded that most of the top 15 lexical chunks were found to be similar in both sources of Business Letter Corpus (BLC) and Online Business Letter (OBL) in terms of type and levels of importance in the form of ranking placement and frequencies. All in all, the overall analyses revealed a very important conclusion that the most types of lexical chunks found in Online Business Letter (OBL) are similarly and correspondingly represented in Business Letter Corpus (BLC) which is the reference source for international standard.

Discussions

Findings from this study revealed several significant discoveries about lexical chunks which are important to be highlighted from the analyses of online business letters (OBL) and Business Letter Corpus (BLC). This study has sampled six (6) types of letters from 10 most visited websites worldwide in which the number of word tokens for these types of letters showed quite a significance difference. Word tokens are referred to as a symbolic count of word to serve as a tangible representation of a fact about text. It was concluded that a greater number of word tokens do not correspond with a greater number of lexical chunks. This may be reflective of the fact that more words used in a business letter do not suggest more lexical chunks to be used. This is due to the nature of lexical chunks themselves which are operated at more-than-two-word structure, instead of a singular individual form. This is in line with the definitions of lexical chunks by several scholars, Becker (1975) and Islam and Timmis (2003) who referred to lexical chunks as more-than-two-word combi-

Universiti Teknologi MARA, Vol. 4, No. 2, 2020

nations. It also verifies the notion of lexical chunks against the notion of individual word expressions. Furthermore, according to Osman and Jusoff (2009) and as supported by Yeganehjoo and Yap (2012), it is the function of lexical chunks which facilitates an efficient textual discourse.

In the first chapter of introduction, the first research question was to find the frequency counts of lexical chunks for 6 types of letters taken from 10 most popular websites with one million rankings worldwide with certain volume of traffic visitors and bounce rate. Sentence Builders (SBs) type emerged to be the most frequent lexical chunks which was then followed by Collocations (COLs). Sentence Builder was defined as phrases up to one sentence long, which operate like a slot and filler and aid as a lexical framework or skeleton in order to express a specific idea (Becker, 1975). Collocations (COLs) were notwithstanding defined as some pairs or groups of co-existing words occur together repeatedly and most commonly associated with verb-noun and adjective-noun pairs as affirmed by Lewis (1997). Someya (1999) upholds this based on his corpus driven study of lexical chunks by highlighting their function as sublanguage in business letters for its specialised use in their own way. It was emphasized in the study on the syntactic forms of “It is ADJ to VERB that” or “It is ADJ that” which is a frequent form of sentence builder found in the business letters which is in parallel with studies conducted by Xu (2010), Xiao, 2011 and Qiong (2013) which pinpointed the salient feature of Sentence Builders in business letters. Grammatical Collocations (GCs) were also found to be more recurrent than Lexical Collocations (LCs) in the online business letters which is consistent with the work of Nhung (2007) who discovered a total of 63.7 percent of Grammatical Collocations (GCs) more than Lexical collocations (LCs) in English enquiry letters from foreign corporations and organizations in Vietnam.

As to Deictic Locutions, this lexical chunks was defined by Becker (1975) as phrases with low variability which exist in the forms of whole utterances or separate clauses and the purpose of these bundles is to direct the discourse which is reflected in the frequent usage of sentence heads or frame tails. Findings of the present study indicated the significant role played by Deictic Locutions in the writing of business letters. These lexical bundles serve as important referential forms or discourse organizers in various settings (Liu & Chen, 2020). Another important conclusion is the important role of most Deictic Locutions to be operated at sentence-based level with more than eighty percent as compared to word-phrase level. This was further validated by Someya (1999) that business letters certainly have its own special language which work at sentence level.

Finally, the research made another salient discovery about the operational level of most lexical chunks found in online business letters. Majority of lexical chunks are sentence-based which dominate more than half percent of all lexical chunks irrespective of types of business letters. This finding was specifically rep-

resented by three types of lexical chunks; Sentence Builders (SBs), sentence-based Deictic Locutions (DLs) and Institutionalized Expressions (IUs) which also imply that these three constitute more than half of the lexical chunks found in the 300 online business letters. This recurrently underpins the notion of sub-language of business letters (Someya, 1999). The research sheds some light on the language teaching researchers who are also commonly teachers on the importance of phrase and sentence-based approach in the teaching of business letters to the students.

From the analysis of the second research question, it was concluded that lexical chunks in Online Business Letters (OBL) are more dominant and reflected in those Business Letter Corpus particularly in the use of sentence builders, collocations and deictic locutions. Furthermore, the representativeness of Grammatical Collocations (GCs) in OBL is nearly equivalent to those in BLC. The same is the case found for the top 15 lexical chunks in online business letters as the identified top lexical chunks showed a high similarity index from those top lexical chunks in Business Letter Corpus (BLC). Numerous studies have applied Business Letter Corpus (BLC) as their source of standard reference (Someya, 1999, Someya, 2000 & Sinturat, 2010) as well as a pedagogical resource to teach the writing of business letters (Flowerdew, 2012). This thereby demonstrates the BLC's usefulness and reliability for use as reference and teaching purposes. This also implies that online business letters which are found to be mostly representative of Business Letter Corpus (BLC) in terms of lexical chunks can be picked up as reference in the teaching and the writing of business letters.

Pedagogical Implications And Considerations For Future Research

First and foremost, in refining the students' writing of effective business letters, it is of paramount importance for the academic teachers of EOP or EAP to teach the correctly and commonly structured sentences used in business letters, in lieu of the traditional way of over-emphasis on the number of words or short phrases. Secondly, the application of sentence-based forms in business letters is more dominant compared to individually parsed ones. In fact, semi-fixed phrases are frequently found in academic or corporate business letters which is a worthy note for them to also take into consideration in their lesson plans of business letter courses. For future research considerations, in view of the present researchers' lack of opportunity to analyse the causative impact of lexical chunks in online business letters on the writing performance, hence this research paves the way for other researchers to adopt the identified lexical chunks and experiment the impact, especially those sentence-based ones on the ability of the students in writing their academic business letters.

References

An Analysis of Lexical Chunks in Online Business Letters (OBL) And Business Letter Corpus (BLC): A Corpus-Based Study

- Allen, D. (2009). Lexical chunks in leaner writing: an analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 1 (1), 105-127.
- Ang, L.H. & Tan, K.H. (2018). Specificity in English for Academic Purposes (EAP): A corpus analysis of lexical bundles in academic writing. *3L: The Southeast Asian Journal of English Language Studies*, 24(2), 82-94. <https://doi.org/10.5281/zenodo.3819303>
- Appel, R., & Murray, L. (2020). L1 differences in L2 English academic writing: A lexical bundles analysis. *Journal of English for Academic Purposes*, 46 (2020), 100873. <https://doi.org/10.1016/j.jeap.2020.100873>
- Appel, R., & Trofimovich, P. (2017). Transitional probability predicts native and non - native use of formulaic sequences. *International Journal of Applied Linguistics*, 27(1), 24-43. <https://doi.org/10.1111/ijal.12100>
- Becker, J. D. (1975). The phrasal lexicon. In B. Nash-Webber & R. Schank (Eds.), *Theoretical issues in natural language processing*, 1 (pp. 70-73). Cambridge, MA: Bolt, Beranek, & Newman.
- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan: University of Michigan Press.
- Betul, B. (2010). Analysis of four-word lexical chunks in published research articles written by Turkish scholars. *Applied Linguistics and English as a Second Language Theses*. Retrieved on September 10, 2015 from www.scholarworks.gsu.edu.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English (Vol. 2)*. MIT Press.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10 (3), 245-261.
- Chao, L. E. I. (2016). A Lexical-Chunk Based Study of Business English Correspondence Writing. *Sino-US English Teaching*, 13(4), 277-283.
- Čolović-Marković, J. (2012). The effects of explicit instruction of formulaic sequences on second-language writers (Doctoral dissertation). The University of Utah.
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 213-238.
- Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics (IRAL) in Language Teaching*; 47 (2) 157-177.
- Evans, S. (2012). Designing email tasks for the Business English classroom: Implications from a study of Hong Kong's key industries. *English for Specific Purposes*, 31(3), 202-212.
- Farvardin, M. T., Afghari, A., & Koosha, M. (2012). Analysis of four-word lexical chunks in physics research articles. *Advances in Digital Multimedia*, 1 (3), 134-139.

- Flowerdew, L. (2012). Exploiting a corpus of business letters from a phraseological, functional perspective. *ReCALL*, 24 (02), 152-168.
- Fogg, B. J., & Marable, L. (2003). How Do Users Evaluate the Credibility of Web Sites? A Study with Over 2,500 Participants. Retrieved on April 10, 2016 from <http://htlab.psy.unipd.it/>
- Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication*, 4, 27.
- Ang, L. H., & He, M. (2017). Multi-word Sequences in Learner Corpora: A Corpus Analysis of Lexical Bundles. *Creative Practices in Language Learning and Teaching (CPLT)*, 5(1), 19-33.
- Hou, H. I. (2014). "Please consider my request for an interview": A Cross-cultural Genre Analysis of Cover Letters Written by Canadian and Taiwanese College Students. *TESL Canada Journal*, 30(7), 45-62.
- Hsu, J. Y. (2006). An Analysis of the multiword lexical units in Contemporary ELT Textbooks. In *International Conference on English Teaching and Learning in the Republic of China*. Retrieved on September 12, 2014 from www.files.eric.ed.gov
- Islam, C. & Timmis, I. (2003). Lexical Approach 1- What Does the Lexical Approach Look Like? Retrieved on May 10, 2016 from <http://www.teachingenglish.org.uk>.
- Kim, N.B. (2019). Effects of the use of lexical chunks on practical English proficiency. *English Language & Literature Teaching*, 25(1), 59-79.
- Krajka, J. (2007). Corpora and language teachers: from ready-made to teacher-made collections. *Computer resources for language learning*, 3(1), 26-55.
- Lewis, M. (1993). *The lexical approach: the state of ELT and the way forward*. Hove, England: Language Teaching Publication.
- Lewis, M. (1997). *Implementing the lexical approach*. Hove, England: Language Teaching Publications.
- Li, Q. (2014). An empirical study on the application of lexical chunk to college English writing. *Journal of Language Teaching and Research*, 5 (3) 682-688.
- Linn, A., Sanden, G. R., & Piekkari, R. (2018). Language standardization in sociolinguistics and international business: Theory and practice across the table. *English in business and commerce: Interactions and policies*, 19-45.
- Lindstromberg, S. (2000). My Good-bye to the Lexical Approach. Retrieved on Jan 20, 2014 from <http://www.hltmag.co.uk/mar03/martmar031.rtf>.
- Liu, C.-Y., & Chen, H.-J. H. (2020). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes*, 58, 122–137.
- López-Jiménez, M. D. (2013). Multi-Word Lexical Units in L2 Textbooks. *Revista Española de Lingüística Aplicada (RESLA)*, 6, 333-348.

- Menon, S., & Mukundan, J. (2012). Collocations of High Frequency Noun Keywords in Prescribed Science Textbooks. *International Education Studies*, 5(6), 149-160.
- Muckian, M., & Woods, J. (2012). *Business letter handbook*. Massachusetts; Adams Media.
- Mukundan, J., & Menon, S. (2008). Nouns and their extended units of meaning: A corpus analysis of nouns used in the Science and English Language textbooks. *Jurnal Sastra Inggris*, 8(2), 90-111.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Toronto: Oxford University Press.
- Osman, N. B., & Jusoff, K. (2009). Routinizing Lexical Phrases on Spoken Discourse. *International Education Studies*, 2 (2), 188.
- Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: Kaitakusha.
- Pastor, M. L. C., & Calderón, R. M. (2012). Lexical variations in business e-mails written by non-native speakers of English. *LSP Journal*, 3 (2), 4-13.
- Patton, MQ. (2001). *Qualitative Research and Evaluation Methods (2nd Edition)*. Thousand oaks, CA: Sage Publications.
- Peters, A. (1983). The Units of language acquisition. In Schmitt, N. & Carter, R. (2000). *Lexical phrases in language learning*. *The Language Teacher*, 28 (4), 6-10.
- Purdy, J. P. (2012). Why first-year college students select online materials as their favourite. *First Monday*. 5 (2), 212-252.
- Qiong, G. (2013). *A study of lexical chunks in business English correspondence writings by higher vocational college students*. (Master's Thesis). Anhui University.
- Romova, Z., & Hetet, J. (2012). Letters to The Editor: Results of corpus analysis. *New Zealand Studies in Applied Linguistics*, 18 (2), 45-63.
- Salazar, L. (2011). *Lexical chunks in scientific English: A corpus-based study of native and non-native writing*. (Doctoral Thesis). Universitat de Barcelona.
- Schmitt, N. & Carter, R. (2000). *Lexical Phrases in language learning*. *The Language Teacher*, 28 (4), 6-10.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sinturat, T. (2010). *An analysis of lexical phrases in business letters: an online business letter corpus and textbooks* (Doctoral dissertation). Srinakharinwirot University.
- Someya, Y. (1999). *A corpus-based study of lexical and grammatical features of written business English* (Doctoral dissertation). University of Tokyo.
- Someya, Y. (2000, August). *Online business letter corpus KWIC concordancer and an experiment in data-driven learning/writing*. In 3rd Association for Business Communication International Conference, Doshisha University, Kyoto, Japan. Retrieved February 24, 2015 from www.someya-net.com

- Someya, Y. (2007). Online business letter corpus KWIC concordance. 1.2 million Words. 2007 – Present.
- Willis, D. (1994). *The lexical Approach*. 3rd ed. London: HarperCollins Publisher.
- Wood, D. (2004). An empirical investigation into the facilitating role of automatized lexical phrases in second language fluency development. *Journal of language and learning*, 2 (1), 27-50.
- Wright, H. R. (2019). Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions. *English for Specific Purposes*, 54, 1–14.
- Wu, S., Witten, I. H., & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL*, 22(1), 83-102.
- Xiao, Z. (2011). *Lexical chunks and English business correspondence* (Master's thesis). Ocean University of China.
- Xu, L. (2010). *A correlation study on lexical chunks and business correspondence writing in EFL in China* (Doctoral dissertation). Kristianstad University.
- Yeganehjoo, M., & Yap, N. T. (2012). Lexical access in production of idioms by proficient L2 learners. *3L; Language, Linguistics and Literature, the Southeast Asian Journal of English Language Studies*, 18(1), 87-104.
- Zhao, Y. (2009). *An empirical study of L2 learners' use of lexical chunks and language production* (Doctoral dissertation). Kristianstad University.