

## **A Corpus-Based Lexical Evaluation of L1 Arabic Learners’ English Literary Essays**

DOI: <https://doi.org/10.33806/ijaes.v23i2.470>

Nahla Nola Bacha and Victor Khachan  
*Lebanese American University, Lebanon*

Received on 28.01.2023      Accepted on 03.06.2023      Published on 20.6.2023

**Abstract:** Evaluating learners’ writing quality has been quite challenging. One important indicator of writing quality is the use of lexis in texts. However, more efficient evaluative guidelines should be explored. Although corpus-based lexical studies have provided analyses of various text genres for word frequencies and keywords indicative of lexical density, and thus writing quality for teaching/learning purposes, learner literary texts remain under researched. This study explores the word frequencies in a corpus of N=206 L1 Arabic learners’ literary essays written in English in one literature course at an English medium university in Lebanon. Lextutor platform was used to analyze the word frequency profiles which indicate the lexical density level and Voyant Tools platform to analyze the content keyword profiles, which preview thematic representations and character features. Main findings indicated a dichotomy between literary knowledge and language proficiency. The content keywords previewed themes and character features adequately which showed the learners’ knowledge of the literary text. The word frequency profiles, however, indicated a low lexical density and, thus, a low language proficiency level. Implications for pedagogy and recommendations are made for further researching this “controversial dichotomy” in learners’ literary essays in developing the literary edge for well-rounded learners versus improving their language proficiency level.

**Keywords:** corpus linguistics; Lebanon; lexical richness; literary corpora; text analysis

### **1. Introduction**

English as a foreign/second language (EFL/ESL) learners have found it difficult to produce literary essays at the university level. Often with little support, their essays earn low evaluations ascribed to their limited knowledge of one significant writing feature, the vocabulary or lexis, required to write essays based on literature (Kroll 1991; Mukattash 2003; Ullah, Uzair and Mahmood 2019; Bacha 2020e; Shakeel and Khan 2021).

Although many current analytic/holistic evaluation guidelines have proved adequate in evaluating learners’ written texts, there is a need for more detailed information of learners’ writing proficiency levels (Schmitt and Rodgers 2020; Bacha 2001a). Other evaluation methods have also provided teachers and learners with an understanding of learners’ writing. These include portfolio assessment (e.g. Bahous 2008), lexical feedback strategies (e.g. Diab and Awada 2022), peer and teacher corrective feedback (e.g. Diab 2006) and computer assisted language learning (CALL) (Hamadeh, Bahous, Diab, Nabhani (2020) among others.

Most studies view vocabulary as the building blocks of discourse and a significant indicator of lexical level (density) and sophistication, qualities of good writing (Scardamalia and Bereiter 1987; Kroll 1991; Engber 1995; Johnson, Acevedo and Mercado 2016; Babanoğlu and Sütçü 2017; Brynildssen 2000; Grant and Ginther 2000; Yu 2010; Vögelin and Stefan 2021; Viera 2022; Zhang 2022). In addition, genre based studies have shown that different text types reveal various lexical features specific to disciplines (Johns and Dudley-Evans 1997; Behzadi and Behmardi 2009; Itani and Bahous 2019; Abu Ramman and Jihad Hamdan (2022), and thus the added challenge for both teachers and learners. This challenge has recently been addressed through the use of computer corpus based studies whereby texts have been analyzed and evaluated. However, these studies are met with criticism from some researchers who maintain that the context of texts is often not taken into consideration (Stubbs 2001; Widdowson 1991; 2000a,b 2008; Reppen, 2010; Boulton 2016).

Nevertheless, corpus based methods provide indirect and direct applications (Römer 2009; Graesser, Dowell and Moldovan 2011) in identifying the level and type of vocabulary used in texts (Hoey, Mahlberg, Stubbs and Wolfgang 2007; Yoon 2011; Boulton 2016; Egbert, Burch and Biber 2020). Practitioners and researchers have produced relevant materials, and teachers and learners have extensively explored texts, referred to as data-driven learning (Römer, 2009; Barabadi and Khajavi 2017; Laosrirattanachai and Ruangjaroon 2021). Ridwan (2011) experimented with a graduating class, and although it was challenging using corpus based techniques, the students were interested and engaged. Other studies also indicated positive findings (Stubbs 2005; O'Keeffe, McCarthy and Carter 2007; Römer 2006; Mahlberg 2012; Bychkovsk and Lee 2017).

In addition, Sinclair (2007) argues that corpora studies have practical and pedagogical value (Flowerdew, 2009) that have led to much understanding of texts and 'intellectual exploration'. Biber, Johansson, Leech, Conrad and Finegan's (1999) investigation of the grammar in speech and writing, note that corpus based lexical analysis research has indeed contributed to many studies in understanding and evaluating texts (Altenberg and Tapper 1998; Bolton, Nelson and Hung 2002; Smith and Kelly 2002; Biber 2011; McEnery and Xiao 2011; McIntyre 2013; López-Couso, Méndez-Naya, Núñez-Pertejo and Palacios-Martínez 2016; Viera 2022).

However, the studies have mainly focused on compiling learner academic corpora (Hunston 2002; Bacha 2005b; Granger 2008; Learning corpus bibliography 2023) and corpus based analyses on learners' academic texts such as research papers, academic essays and disciplinary texts (Hunston 2002; 2009; Benavides 2015; Khan, Bychkovska and Lee 2017; Khan and Mohammad 2018; Lin and Lin 2019). One major extensive learner computerized corpus of argumentative essays, in many languages from Europe, the Middle East and Asia, has further contributed to the understanding of learners' writing (Granger, Dagneaux, Meunier, and Paquot 2009). Schmitt and Rodgers (2020) observe the importance of learner corpora, and maintain that the International Corpus of

Learner English (ICLE) is a good example of both academic and non-academic language.

Although there is a great deal of corpus analysis studies on academic and non-academic language of various genres including literature, especially the novel (Mahlberg 2010; Höglund and Syrjänen 2016), there are very few, if any, available studies on learners' literary texts. The novelty and contribution of the current study is that it is an interdisciplinary literary-linguistic one in proposing an alternative evaluation method of learners' vocabulary proficiency level using two corpus-based computer platforms.

## 2. Review of literature

### 2.1 Literary corpora

Literary corpora-based studies have focused on professional text genres to explore differences and/or similarities in discourse and vocabulary profiles. To illustrate, the *Corpus of Early American Literature* (CEAL) includes texts from 1690–1920 (Höglund and Syrjänen, 2016). One other is the *Corpora of 18<sup>th</sup> century Prose* (C18P), a ten million word literary corpus taken from British novels between 1700 and 1830 and which can be of use to literary critics. This work has mainly been done on Dickens' and Shakespeare's literary texts (Gemeinbock 2016). Additionally, Alsuweed (2015), in his doctorate research, compiled a corpus of Charles Dickens' texts based on the study of the latter's semantic and lexical profile in order to help learners understand Dickens' style and vocabulary. Corpus analysis of literary texts, therefore, is a relatively 'young' field (Granger 2008).

Two corpus-based computer platforms are Lextutor and Voyant Tools. Lextutor averages the content words (referred to as type) over the number of words in the text (tokens) to give an index and/or percentage of the vocabulary level of the text (lexical density). Voyant Tools (Sinclair and Rockwell 2021) analyzes key content vocabulary frequencies. Cirrus Tools in Voyant tools, analyzes keyword frequencies that predict the themes and character features in texts. This platform then can provide learners with a preview of possible themes and characters before they read the entire literary work. Often than not, learners need guidance in reading and understanding professional literary works as the culture, and thus the vocabulary, can be very challenging for the learner (Römer 2006). The use of Cirrus Tools could be an effective method in facing the challenge. Utilizing Lextutor and Voyant Tools offer the teaching/learning context an efficient way to read, write and evaluate literary texts.

Recently, there have been a few corpus linguistic studies carried out on novels using web-based computer tools to compile literary terms. Asif, Zaidi and Yasmeen's (2021) work on Jane Austen's novel, *Pride and Prejudice*, utilized Cirrus Tools in Voyant Tools (Sinclair and Rockwell 2021). Another study was done by Ajmal and Shoukat (2020) using AntConc 3.5.8 computer-based platform (Anthony 2020) on James Joyce's novel, *A Portrait of an Artist as a Young Man*. In compiling the highest to lowest word frequencies, and visualized as word clouds, major themes could be identified and interpreted. Also, characters'

features, through keywords and high frequency words, could be predicted and interpreted.

The foregoing and similar work in analyzing vocabulary frequencies and identifying keywords in 19<sup>th</sup> and 20<sup>th</sup> century novels, such as *Harry Potter*, *Sherlock Holmes*, and Shakespeare's works, have contributed to the teaching/learning of literature (Asif et al. 2021; Ullah et al. 2019). Specifically, these studies have helped learners to obtain previews of the novels' themes and characters through vocabulary frequencies before reading them in their entirety which otherwise might be a great ordeal for learners (Ullah et al. 2019; Wahid 2011). Research has shown that the objective results that are obtained from the corpus-based literary analyses can often confirm the critics' literary interpretations and/or provide alternatives (Mahlberg, Wiegand, Stockwell and Hennessey (2019). These studies, however, did not include learners' literary texts.

It was, perhaps, with Shaw (2009) that learner literary texts began to draw attention. Shaw (2009) compared the lexical functional features in learners' literary texts with those in professional literary research papers. He examined the lexical density and placement (initial, middle or end of a sentence) of linking adverbials (e.g. however, thus) in learners' texts to those in professional literary research articles. Findings revealed that learners' texts appear to have a higher density of adverbial links, mostly in initial position, and within shorter distances between these adverbial links than those in the professional texts. In this way, learners' weaknesses could be identified. Shaw (2009:217-219) emphasizes, though, that '...genre and discipline influence density [use] and profile [level]' and learner literary texts '...remain largely uncharted at the linguistic level'.

## 2.2 Vocabulary profiling and lexical density

Vocabulary profiling involves studying the number, variation and sophistication of content and grammatical words. Content words give meaning to the text such as nouns, adjectives, verbs, adverbs; for example, *Catherine, came, lovely, often*, while grammatical or function words act as links within and between sentences; for example, conjunctions, prepositions and articles such *but, on, the*. Lexical density refers to the ratio between content words (types) and the total words (tokens) in a text. The proportion of content and function words, and thus the lexical density level in any text, depends upon genre type and shows the lexical profile and type in the text in question (Biber and Conrad 2001; Hussein and Abdul-kadhim 2020). In fact, Hussein and Abdul-kadhim (2020:2) state that descriptive genres show higher lexical densities than those in narrative texts. They mention that

'The concept of *lexical density* (henceforth LD) refers to the description of content words to the total number of the text words (in Singla 2012:35)'. .... It is the 'degree of richness of a text in terms of meanings, ideas and information (in Al-Wahy 2016:5)'.

Hussein and Abdul-kadhim (2020:3) conclude by accepting their hypothesis that

‘First, corpus-based techniques are quite helpful in producing empirical and quantitative stylistic descriptions of literary texts in so far as lexical density is concerned, second, lexical density profile is an efficient measure for lexicality of literary texts as well as the lexical development of the authorial style, and third, the high percentages of lexical density are due to the author’s rich lexical resources that lead to a denser lexical profile’.

The compilation of academic language corpora has led to insightful studies in profiling different aspects of learners’ language and benchmarking the vocabulary profiles in learners’ texts. First, the iconic Coxhead’s (2000) Academic Word List (AWL) has laid the foundation for researchers, linguists and teachers to benchmark students’ and learners’ academic vocabulary profile to begin tertiary education (Bacha and Khachan 2012c). To our knowledge, no such compilation is available for learners’ literary language except as part of Coxhead’s AWL (2000).

Second, according to Nation (1990), there are two types of vocabulary counts of university texts: 1) a basic vocabulary or the most frequent vocabulary; for example, the General Service List (GSL), that is found in most texts and entail 80% (K1 – first 1,000 words and K2 second 1,000 words) of the text such as *day* and *clock*; 2) low frequent vocabulary such as the Academic Word List (AWL), 10% (K3, 570 families) of academic texts selected from arts, science, law, and commerce texts, and 10% Technical Terms/Off List (OFL) some of which are of Greek, Latin and French origin and appear in certain disciplines (Coxhead, 2000).

Nation (2001) maintains that the use of vocabulary in an academic text should be no higher than 70% for K1, 10% or higher for K2, 10% or higher for AWL and 10% or higher for OFL words (discipline specific words). Shakeel and Khan (2020) compiled an Academic Literary Word List (ALWL) from professional literary research articles which included both academic and literary words. Along with Nation’s (2001) lists, ALWL could help in future research once its validity is examined.

Hussein and Abdul-Kadhim (2020) provide a comprehensive and rationale account how researchers, using different measures, have investigated ‘lexical density as a marker of stylistics’ in texts, define and quantify it. From among these measures, they select Ure’s (1971) measure as it can deal with large quantities of corpora and is a valid statistical measure of lexical density (LD). Ure (1971) calculates lexical density as the number of lexical items (content words to the exclusion of functional words) divided by the number of tokens (number of words in the text) multiplied by 100%. Ure (1971) interprets high lexical density as showing frequent lexical (content words) references.

As an illustration of Ure’s (1971) evaluative method, Hussein and Abdul-Kadhim (2020) analyzed three novels by Ernest Hemingway, utilizing Lextutor (Cobb, 2021). They found that Hemingway’s last novel, *A Moveable Feast*, had a higher lexical density (54%) than his first, *The Sun Also Rises* (48.6%) and second novel, *A Farewell to Arms* (49%). These measures are used as benchmarks in the current study. Hussein and Abdul-Kadhim (2020) conclude by citing Neumann’s (2014:16) comment that Hemingway’s style became progressively more lexically

sophisticated and that these LD's are averages for written fictional genres. Hussein and Abdul-kadhim (2020:16) further remark that these '... measured LD Profiles are considered *model* LD averages for written fictional genres.' and '.... it is clear that LD can be used as a stylistic marker in literary texts.' Additionally, 'These decisions are reflected in high lexical density percentages and denser lexical profiles where factual information is used to convey and capture events in words'. In the current study, lexical density will be based on the type/token ratio as explained in section 2.2.

Although there is extensive research on professional literary and academic texts (Biber et al. 1999; Biber 2011; Bacha and Khachan 2012c; Pretila, Doro and Pipalova 2015; Wulff 2017 Picoral 2018), few have been done on learner's literary texts. The current study is an attempt to fill this gap.

### **3. Significance of the current study**

Since the vocabulary in learners' literary essays are under studied, the contribution of corpus analysis in investigating the lexical density and vocabulary profile of learners' literary essays becomes important for pedagogic purposes (Granger 2002; Mahlerg 2007; McEnery and Xiao 2011; Bacha and Khachan 2012c; Boulton and Thomas 2012). Schmitt and Rodgers (2020) claim that vocabulary is the major contributor to meaning-making and the organization of language. Thus, vocabulary becomes significant to study along with appropriate use of corpus-based lexical tools and analytical methods (Moon 2008).

It is important that the lexis in the learners' literary writing be diagnosed so that learners know how to develop their writing. The debate on whether literature can actually help learners' lexical development is viewed with skepticism by some researchers (Toolan 2004; Römer 2009; Castello 2008). On the other hand, there are those who argue that literature does help (Ihejirika 2014; Armstrong 2017; Bacha 2020e). They claim that literature offers a context for language learning, widens learners' lexical repertoire, focuses on authentic language, provides a motivational learning environment and develops learners' critical thinking skills which are necessary for scholarly communication (Butler 2006; Paran 2006; Römer 2009; Hoey 2013; Ihejirika 2014; Bacha 2016d; Nation and Meara 2020; Mart 2021).

In addition to this, learner corpus literary analysis is significant as it contributes to the identification of errors, comparisons with professional texts, production of materials, and evaluation of texts (Bolton et al. 2002; Sinclair 2004). Further studies have recommended that a lexical-grammatical syllabus be provided in which corpora is used in the teaching/learning context (Lewis 2000; Hunston 2002). Literary learner corpus-based studies are further significant in offering a wide interdisciplinary field of study in linguistics and literature which would enlighten educators' understanding of vocabulary in learners' literary texts. This would give an alternative vocabulary evaluation of the learners' literary essays and contribute to the teaching/learning of essays on literature in similar EFL/ESL contexts.

Most significantly, there is a controversy in the academy between the language and discipline teachers whether or not language (sentence structure, grammar, vocabulary, mechanics and so on) should also be taught and learned in the disciplines such as literature courses. Language teachers argue that discipline teachers also have a role to teach the language as they, the language teachers, do not have the knowledge of the disciplines' content. This has led to the formation of writing across the curriculum programs (WAC) where language and disciplinary teachers sometimes work together (Johns and Dudley-Evans 1997; Johns, Bawarshi, Coe, Hyland, Paltridge, Reiff and Tardy 2006; Defazio, Jones, Tennant and Hook 2010). The current study touches on this debate in the two research questions. Since studies in learners' literary corpus analysis are limited and in order to explore the lexical density for alternative evaluation measurements, this study addresses the two research questions below.

Research Question 1: Are learners' lexical profiles indicative of literary language required proficiency levels?

Research Question 2: To what extent do learners' literary essays mirror thematic and character knowledge based on Hemingway's novel, *A Farewell to Arms*?

#### 4. Method

The research design is a word frequency approach using quantitative corpus analysis techniques (Sinclair 1991; Hunston and Francis, 2000; Lewis 2000; Hussein and Abdul-Kadhim 2020). Word frequency analyses are used to study the lexical density in the 9 sub-corpora with a total of N=206 learner essays in the corpus. Each of the 9 sub-corpora are analyzed separately, and then the results are averaged. The focus of the study is on the total average of the 9 sub-corpora. The steps involved in the procedure are 1) selecting the learners' literary essays, 2) selecting corpus platforms and 3) analyzing the data.

##### 4.1 Selecting the Learners' literary essays

Critics have described Earnest Hemingway's use of vocabulary as 'simple,' with relatively few adverbials and adjectives. This style is often referred to as the 'iceberg theory', as it allows readers to ponder below the surface to interpret deeper meanings and produce appropriate vocabulary (Daoshan and Shuo 2014). The novel, *A Farewell to Arms*, was thus selected for this style which could help the learners develop their critical thinking skills in the reading and writing process. The learners could also relate to the events of the struggles in love and war. It was also the first assigned reading in the course. The steps in selecting the essay corpus are given below:

1. A total N=206 learner literary essays on *A Farewell to Arms*, or the 9 sub-corpora, were uploaded from the corpus of the 20<sup>th</sup> Century American Course. This corpus is part of a larger corpus bank of learners' literary essays from different literature courses at the university in the current study. The selected corpus had been written in English by L1 Arabic students.

2. Essay topics for the 9 sub-corpora focused on the novel's characters and themes, and the essays were written in four paragraphs: an introductory paragraph, two body paragraphs and a concluding paragraph.
3. The essay writing had been carried out in the classroom for an hour as part of the first regular course tests and after the first four weeks of the semester. During the four weeks prior to the test, learners had read and discussed the elements of the novel, mainly the themes and characters.

## 4.2 Selecting corpus platforms

Among the many computer corpus tools (Scott 2004) to analyze literary data, Lextutor and Voyant Tools were found adequate for the purposes of the current study.

Lextutor computer platform was selected as it offers a web-based program which students, teachers and researchers can use in investigating a variety of word uses in texts and can be used for learning, teaching and research. It is suitable as both academic and specific discipline word frequencies are examined and, therefore, can help in the evaluation of learners' essays. Although there are other corpus tools, the authors found Lextutor to be rigorous in analyzing word frequencies based on a classified list of words, (GSL and AWL) necessary for university academic and discipline studies.

### 4.2.1 Vocabulary profiling

Lextutor platform allows for a number of lexical analyses one of which is vocabulary profiling. According to the web source:

'VocabProfile [(VP) (<https://www.lextutor.ca/vp/comp/>)] is a computer program that performs lexical text analysis. It takes any text and divides its words into four categories by frequency: (1) the most frequent 1000 words of English, (2) the second most frequent thousand words of English, i.e. 1001 to 2000, (3) the academic words of English (the AWL, 550 word families that are frequent in academic texts across subjects), and (4) the remainder which are not found on the other lists'.

This was the main corpus tool in Lextutor platform in analyzing the word frequencies of the N=206 corpus to obtain the vocabulary profile (Refer to section 2.2 in the current study for a fuller account).

### 4.2.2 Frequency analysis

Thus, VP measures the proportions of low and high frequency vocabulary used by a native speaker or language learner in a written text. A typical native student's result is 70-10-10-10, or 70% from the first 1000 words, 10% from the second thousand, 10% from the AWL, and 10% from the fewer frequent words that do not appear in the other lists (mainly related to the various disciplines). This relatively simple tool has been useful in understanding the lexical acquisition and performance of second language learners.' ([Research Uses of VocabProfile \(lextutor.ca\)](https://www.lextutor.ca)).



Results of the frequency analysis are often reported as lexical indices. However, Ure (1971) goes a step further in measuring lexical density and calculates it as a percentage by multiplying it by 100%. Hussein and Abdul-kadhim (2020:5) report that ‘Many corpus-based studies use Ure's method, for it can be applied to large amounts of corpora’. Laufer and Nation (1995) researched and validated Vocabprofile (VP) in Lextutor and found that it has a wide measure of learners’ language proficiency ([Research Uses of VocabProfile \(lextutor.ca\)](#)). In relation to this wide measure, Hussein and Abdul-kadhim (2020:2) state that

‘Lexical analysis has been approached from different perspectives and throughout a variety of studies. Such perspectives range from *lexical diversity*, *lexical richness*, *lexical variation* to *lexical density* and are widely discussed under the same single heading of lexical analysis. Lexical density, in particular, has been dealt with in different fields of study fulfilling different analytic purposes’.

Ure’s (1971) measurements of lexical density is based on K1, K2, AWL, OFL in benchmarking the data in the current study. Nur (2015:20), in interpreting the data, states:

‘To value the vocabulary level of the students’ essay writing scores in academic perspectives, the use of K1 should be at a maximum or less than 70% of all words produced; K2 should be at a minimum or more than 10% of all words; AWL should be at a minimum or more than 10% of all words; and the same is true for the OLW [off list/discipline words] that should be a minimum or more than 10% of all words produced.’

In other words, a good level of vocabulary in a literary text would show a vocabulary profile of less than 70% of the total words from K1, more than 10% from K2, more than 10% from AWL and more than 10% from OLW.

### 4.2.3 Voyant corpus platform

The Voyant corpus platform through Cirrus Tools visualizes lexis in word clouds, links, graphs, trends, tree diagrams, or bubbles. It is effective in providing learners a thematic and character visual preview of the most frequent words in a literary text. These words are positioned centrally in the word clouds and are sized the largest (refer to Figures 1-5 in the results section of the current study). Thus, an objective insight into the writer’s literal messages is possible in relation to the real world (McNaught and Lam 2014; Hussein and Abdul-kadhim 2020; Masood, Shafi and Darwesh 2020).

Learners, teachers and researchers could make use of this objective insight in offering/comparing their own interpretations with the wider professional literary criticism arena (e.g. Sadaka 2022; Sadaka and Panossian 2022).

The authors of the current study claim that although statistical results are important in the evaluation of the vocabulary level in learners’ written texts, it is important that they also understood the topic and used relevant content in the essays. Cirrus Tools in [Voyant Tools \(voyant-tools.org\)](#) were utilized to identify

the frequent content words in the essays and their possible relationships to the novel's main themes and characters. Cirrus Tools gives a statistical count of the highest and lowest word frequencies in texts which preview possible main themes and characters. If used by the learners, they could preview the literary content of a text and better understand the literary work before reading it. This enables a valuable preliminary reading exercise. Asif et al. (2021) report that utilizing corpora in analyzing literature motivated students and that Voyant Tools is important and beneficial in the literature classroom.

### 4.3 Analyzing the data

Thus, both corpus platforms, Lextutor and Voyant tools, are suitable to analyze the learners' literary lexis in texts. Although the lexical analysis measurements (Ure 1971; Coxhead 2000; Nation 2001) are averages of lexical profiling in academic articles/chapters, and not in students'/learners' writings, due to their rigorous analysis, validity and reliability, they have been used to benchmark the lexical levels of the learners' literary essays. These corpus platforms are programmed with the measurements and automatically analyze the given data according to selected techniques and tools (Asif et al. (2021; [Research Uses of VocabProfile \(lextutor.ca\) 2023](#)). Once the 9 sub-corpora were digitalized as plain text to suit the platforms, uploaded on the relevant websites and the various tools selected, the analysis would provide statistical and visual results. These results are then formatted on excel sheets in tables and figures as shown in section 5 of the current study. Hussein and Abdul-kadhim (2020:7) report that

Ure's [1971] formula states that the LD is measured by means of dividing the number of lexical items [content words] by the total number of tokens [total words] multiplied by (100). For example, the sentence (The outer space is the expanse which exists beyond the earth and between celestial objects) contains (15) tokens as a total and only (7) lexical items (outer, space, expanse, exists, earth, celestial and objects). Thence, the LD of this sentence reads as ( $\frac{7}{15} \times 100 = 46\%$ ).

The LD of 46% can also be interpreted as an index of 0.46.

## 5. Results and discussion

The findings of the study are discussed below according to the two research questions.

### 5.1 Research question 1: Are learners' lexical profiles indicative of literary language required proficiency levels.

Table 1 indicates the results of the N=206 learners' essays as analyzed for lexical density according to GSL and AWL showing an overall relatively low lexical density 0.49.5 index or 49.5%. The second, third, fourth, fifth and sixth columns mention the vocabulary percentages of Nation's (2001) K1% and K2%, Coxhead's (2000) Academic Word List (AWL%) and Off List Words (OFL%) and the total number of essays and words in each class, respectively, in each of the nine sub-corpora. The seventh column notes the lexical density index/level. Although

the average number of words varied in the nine sub-corpora, the results remain valid since the aim of the current study is to report the lexical density and lexical frequencies in general terms.

**Table 1: Lexical Profiles of Learners' Literary Essays**

Literary Corpora: 9 Sub Corpora	K1%	K2%	AWL%	OLW%	Words/Essays	Lexical Density Index
1	82.54	3.74	3.72	10	8,128 words 22 essays	0.48
2	80.1	4.16	3.94	11.9	19,330 words 28 essays	0.49
3	88.7	3.6	3.1	4.56	13,194 words 23 essays	0.49
4	83.58	3.69	3.24	9.48	11,505 words 24 essays	0.49
5	80.77	4.33	4.38	10.52	8,196 words 25 essays	0.51
6	80.48	3.54	4.43	11.56	12,043 words 20 essays	0.51
7	82.42	3.51	4.68	9.39	8,949 words 21 essays	0.49
8	88.3	3.7	3.8	4.19	20,000 words 25 essays	0.51
9	89.3	3.6	3.4	3.64	13,542 words 18 essays	0.49
<b>Average.</b>	<b>84.02</b>	<b>3.76</b>	<b>3.85</b>	<b>8.36</b>	<b>114,345 total words 206 essays</b>	<b>0.4955</b>
<b>Note:</b>						
The lexical density index (0 to 1) can also be calculated as a percentage (Ure 1971)						
For example, 0.48 index or 48%.						

Table 1 shows the lexical analysis results of K1, K2, AWL and the OLW of the 9 sub-corpora with averages of 84.02%, 3.76%, 3.85% and 8.36% respectively. These results are below the required lexical level according to Nation's (2001) calculations. Although rows in Table 1 show some differences in the Off List words (OLW) (discipline related) over the 9 sub-corporal, the average is still below the minimum required level of 10%. One would have expected higher percentages in a few of the classes since the learners were attending a literature course. Nevertheless, the OLW total average of 8.6% is much better than those of K1, K2 and K3 (Hussein and Abdul-kadhim 2020).

The results in Table 1 further show the lexical density index column ranging between 0.48 and 0.51 with an average of 0.49.5. It appears that learners' texts, irrespective of the total number of words or essays in each class, have a relatively low academic lexical profile with most of the vocabulary from the basic K1 1,000, ranging between 80.2%, in row 2, to 89.3% in row 9, and an average of 84.02%. This average is above 70% by 14.02% which indicates that the learners' vocabulary is at a basic level.

To address the first research question, linguistic corpus-based lexical profiling gives insight into the lexis of learners' literary texts. This confirms many of the studies that learners' language proficiency level can be evaluated through lexical density profiles (Bacha 2005b; Shaw 2009; Ullah, et al. 2019; Hussein and Abdul-Kadhim; 2020; Asif et al., 2021). The results concerning the learners' literary essays in addressing research question 1 indicate that learners used fewer words from the K2 and/or AWL and, to a certain degree, OLW vocabulary (discipline related). This implies a need for teaching/learning methods to focus on developing the learners' vocabulary Profile.

Thus, to answer research question 1, learners' lexical profiles are indicative of low literary language levels.

**5.2 Research question 2:** To what extent do learners' literary essays mirror thematic and character knowledge based on Hemingway's novel, *A Farewell to Arms*?

The second research question explores content vocabulary that predicts possible themes and characters. The findings in Figures 1-4, in general terms over the 9 sub-corpora, showed that the learners' literary essays indicate the main characters: *Henry* and *Catherine*, and that *war*, one of the high frequency words, is predictive of a main theme which links *Rinaldi to Henry and Catherine*. Words such as *hospital, change, time, difficult* and so forth in the cirrus cloud (Figure 6) intensify the theme of *war* and suggest other themes such as *illness, suffering and death*. Words such as *love, farewell, escape, success, good, world* may envision another side to war and death, that of love and peace. In fact, in the novel, Henry and Catherine wish to run away from the 'real' world where only disillusion and loneliness exist. This drive for a better peaceful world is central but unattainable. Henry and Catherine, in the novel, run away from the war only to find death when Catherine and the baby die in childbirth.

Further, the interrelationships among the characters and themes are visualized in the link cloud where *Henry* is central as represented by the high frequency (Figure 1) and by the large size of his name (Figure 6). There are also direct links between *Henry, Catherine* and *love* due to high frequencies predicting a character-thematic relationship (Figures 1 and 2). The findings in the extensive word cloud, as visualized in Figure 6, indicate students' engagement with the novel. Sample visuals and interpretations according to themes and characters are illustrated below in Figures 2-5 in sections 5.2.1. and 5.2.2.

**5.2.1 Thematic and character knowledge**

Voyant Tools describes Cirrus Tools as giving a word count of the highest word frequencies in the corpus. Results in Figure 1 below indicate the seven highest content word frequencies over 500 which range from 508 to 1,787 frequencies in the 9 sub-corpora. As noted, the main characters, *Henry* and *Catherine* show the highest frequencies with *war, love* and *life* as possible predictions of the novel's themes and related to another character, *Rinaldi*, the doctor.

	Term	Count
1	henry	1787
2	catherine	1316
3	war	1285
4	love	1221
5	novel	595
6	life	536
7	rinaldi	508

Figure 1. Highest word frequencies in the 9 sub-corpora

Figures 2-4 below show these frequencies in word clouds which give a convenient overview of the content features and relationships. Furthermore, findings in Figures 2-4, indicate the more frequent words are in blue which link with the words in orange, the latter being in the closest proximity to the blue words in the text. As discussed in section 5.1., the character *Henry* (Figure 2) shows thematic and character links to keywords of *love*, *war* and *Catherine*. Similarly, the words in blue in Figures 3-5 predict characters and themes. Figure 6 shows a wider perspective of content word frequencies (the 275 highest word frequencies) which indicate the learners' extensive knowledge of the novel. Examples are *psychology*, *mental*, *relationship*, *death*, *crazy*, *protagonist* that showcase the characters' features and/or personalities. Catherine's mental instability, for example, portrayed in the novel, is one side of her insecure personality as she continually hesitates to the end to believe Henry's love is true. Through themes such as *love* and *death* perhaps a comparison can be made between the heroine Catherine with others such as Jane in the novel *Jane Eyre* as the two women, although almost a century apart, struggle in a patriarchal society for 'freedom' (Sadaka and Panossian 2022) and for psychological peace of mind. Corpus analysis can help in objectively visualizing relationships through words among characters over time.

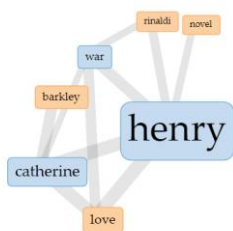


Figure 2. Vocabulary links viewed from the character perspective of *Henry*

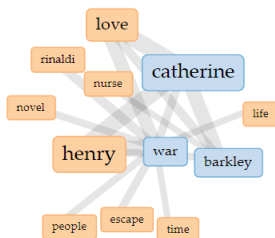


Figure 3. Vocabulary links viewed from the character perspective of *Catherine*

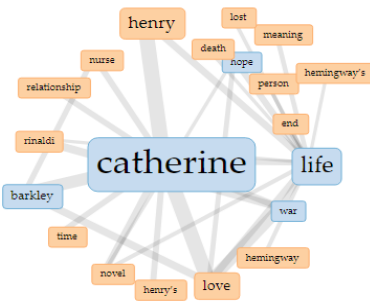


Figure 4. Vocabulary links viewed from a wider character perspective of *Catherine*

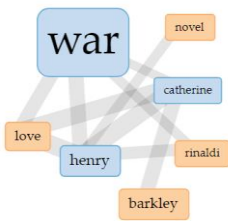


Figure 5. Vocabulary links viewed from the theme perspective of *war*



### 5.2.2 Vocabulary trends

Cirrus Tools provide line graphs to view the relative high word frequencies over the text. Results in Figures 7-10 visualize some of these words graphically over the 9 sub-corpora. These results are interesting as they showcase the actions and events in the novel. For example, *war* does not seem to be a concern in some essays, while in others the *war* action develops to its highest as in the 7<sup>th</sup> sub-corpora. *Catherine, Henry and love* are more closely related throughout the 9 sub-corpora. It is interesting that relatively high word frequencies such as *hope, life and death* when separately plotted in Figures 7-9 are related when compared. For example, the highly plotted lines of *hope, life and death* in Figures 7-9 reflect a great deal of *hope and life and death (in war)* in the beginning of the novel, but then *death (Catherine and her baby), hopelessness and life* are visualized the least towards the end of the graph. This is showcased in the novel when life was hopeful for Henry and Catherine at the beginning of their relationship, but it turned out to be a tragedy at the end with Catherine dying in childbirth. This is also a reflection of Hemingway's own life in searching for hope and meaning in life in the beginning but ending in his own suicide. A century ahead, and a link to the theme of disillusion, Sadaka (2022) sadly remarks that

‘Three Ps cause and aggravate my illness: the Protests, Pandemic, and Pandæmonium. I name my illness 3P and I seek a rhyming therapy in writing—a 3P-Therapy—to combat my illness of illusion and disillusionment’ ... ‘...it is hard to determine whether I am alive or dead, having hopeful fears or fearful hopes, being a citizen of the world or a monstrosity of stoic survival.

These words mirror those in Hemingway's novel and his life. These words are also mirrored in the below visual clouds. Could themes and characters be linked across literary works through words over time? Could the meaning of life through words be viewed across literary works over time? Sadaka (2022) through words is linked to Henry in the aphoria of time. Perhaps, corpus-based lexical analysis can contribute in answering these questions.

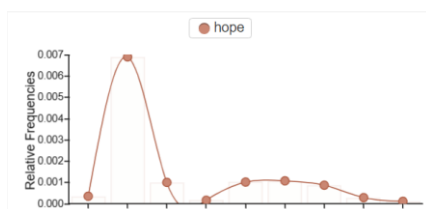


Figure 7. Trend of the theme *hope* over the learners' essays in the 9 sub-corpora



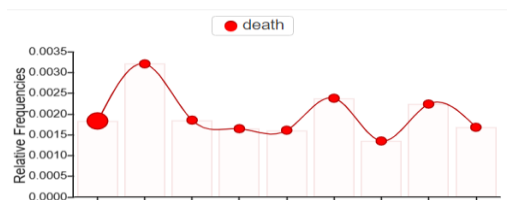


Figure 8. Trend of the theme *death* over the learners' essays in the 9 sub-corpora

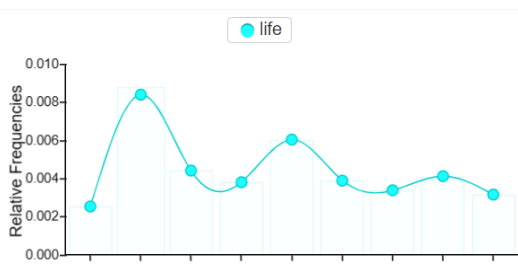


Figure 9. Trend of the theme *life* over the learners' essays in the 9 sub-corpora

Although the lexical profiles of the learners' essays were lower than the expected statistical benchmark of vocabulary levels, the frequency of words in revealing and visualizing the highest and lowest word frequencies showed two important points for the current study. First, the most frequent words were directly related to the essay topics showing relevance of content and also to the links of high and low word frequencies revealed the relationships of characters among each other as well as hinting at the themes as a reflection of those in the novel.

It can be assumed that the students quite understood the content of the novel as shown in the links and word and trend clouds in the Cirrus Tools. The 9 sub-corpora revealed *Henry*, *war* and *love* to be consistently central. Although the essay prompts did not include these words, it is apparent that students showcased their knowledge from class discussions and, perhaps, interest in the events. The cirrus clouds in the 9 sub-corpora visualized the high frequent words which related to main characters and themes. These results confirm the studies on professional novels (Ullah et al. 2019; Hussein and Abdul Kadhim 2020; Asif et al. 2021). They also confirm Mahlberg's et al. (2019) research that corpus based methods can confirm literary interpretations done by critics.

Thus, to answer research question 2, learners' literary essays mirror extensively thematic and character knowledge based on Hemingway's novel, *A Farewell to Arms*?

## 6. Conclusion

The aim of the study was to identify the lexical density and lexical frequencies in a writing sample of EFL L1 Arabic learners which contribute to the research on evaluating learners' literary vocabulary through corpus-based methods. The corpus of the learners' literary essays in the 20<sup>th</sup> century American novel literature course at the university in this study indicated limited vocabulary in the target language, English, which confirms studies that learners often find it difficult to produce the necessary required vocabulary (Kroll 1991; Mukattash 2003; Ullah, et al. 2019; Bacha 2020e). In evaluating the learners' literary essays, analytic and holistic scoring methods have sometimes been found inadequate as far as the vocabulary is concerned (Bacha 2001a). Instead, this study used a corpus-based lexical analysis to explore the learners' literary essays for lexical density. Through the use of corpus linguistic tools, the main finding indicated that according to the benchmarks used in the current study, the learners' vocabulary proficiency is not up to the quality required for university studies. The study notes other evaluative methods used (e.g. Bahous 2008; Diab and Awada 2022) and contributes to a corpus-based method that instructors can use in identifying lexical density and evaluating learners' writing.

It seems the learners in the study are on top of the knowledge of the novel as they were able to signal and pinpoint important themes despite their relatively low language proficiency. It remains a controversial dichotomy for educators at the university and higher institutions in the Arab world as to what we want to develop in literary studies: knowledge of the text or the language. Ideally, we want to do both. If we want to develop them as well rounded citizens, content should take priority, but if we need to further their language, more work needs to be done. We can say, nevertheless, that at this point, the learners showcased adequate literary knowledge although they did not have the required language proficiency level. The answer, however, is left for future research.

Implications of the present study for researchers and teachers are to develop learners' vocabulary (Nation, 2021) which have been found to be weak (Gilroy and Parkinson 1997; Mukattash 2003; Paran 2006; Aijmer 2009; Shaaban 2017; Bacha 2020e). Although there have been debates on the value of corpus-based lexical studies (Sinclair 1991; Nelson 2000; de Beaugrande 2001; Widdowson 1991; 2000a, b 2008), the current study has indicated that corpus based analysis can make a positive contribution to the literary teaching/learning context especially in understanding and evaluating in objective terms learners' use of vocabulary (Reppen 2010; Boulton 2016).

The limitations of the current study are mainly two. The sample was small for generalization to a wider population. Larger data samples are needed from a wider learner literary data base and from other universities, using experimental and control groups so that any findings of learners' language development can be generalized. Second, the sample essays were from one literature course. Data needs to be taken from various literature courses for more valid and reliable results.

It is recommended that future research explore and compare learners' lexis in other disciplinary texts as well as the development of their lexical level using

corpus based lexical linguistic methods. There is also a need to build large learner literary corpora of quality language and knowledge levels to benchmark learners' literary corpora. More lexical corpus-based studies could also compare lexical density and thematic/character keyword frequencies with those in learners' literary writing and novels across time.

As a final word, educators and researchers, need to take more note that in our technological age, learners at a very young age are increasingly using computers, platforms and social media over pen and paper. Corpus-based analysis offers these learners a wide spectrum of inspiring teaching/learning techniques that can bring a smile on both learners' and teachers' faces as they struggle with the challenging task of producing quality literary writing and effective evaluation at universities.

**Acknowledgements:** The authors thank the editors and reviewers who gave informative feedback and dedicated valuable time to reviewing our manuscript.

Nahla Nola Bacha, (PhD) – Corresponding Author

Professor of Applied Linguistics

Lebanese American University, Lebanon

ORCID ID: 0000-0002-1083-2818

Email: [nbacha@lau.edu.lb](mailto:nbacha@lau.edu.lb)

Victor Khachan, (PhD)

Associate Professor of Applied Linguistics

Lebanese American University, Lebanon

ORCID ID: 0000-0003-3636-8459

Email: [vkhachan@lau.edu.lb](mailto:vkhachan@lau.edu.lb)

## References

- Abu Ramman, Ronza and Jihad Hamdan.** (2022). ‘Conceptual metaphors in political discourse: Evidence from the speeches of King Abdullah II of Jordan’. *Dirasat: Human and Social Sciences*, 49(5):189-205.  
DOI: 10.35516/hum.v49i5.3468.
- Aijmer, Karin.** (ed.). (2009). *Corpora and Language Teaching*. Amsterdam: John Benjamins.
- Ajmal, Muhammad and Aisha Shoukat.** (2020). ‘A corpus based study of James Joyce’s *A Portrait of the Artist as a Young Man*’. *Corporum: Journal of Corpus Linguistics*, 30(1):57–68.
- Alsweed, Muhamad.** (2015). Accessing Dicken’s style as an EFL learner: A corpus stylistic approach to lexical style. PhD Thesis. University of Huddersfield.
- Altenberg, Bengt. and Marie Tapper.** (1998). ‘The use of adverbial connectors in advanced Swedish learners’ written English’. In S. Granger (ed.), *Learner English on Computer*, 80-93. London: Longman.
- Anthony, Lawrence.** (2020). ‘Lawrence Anthony's Website’. Retrieved from Lawrence Anthony’s Website:  
<https://www.laurenceanthony.net/software/antconc/>.
- Armstrong, Matthew I.** (2017). ‘Literature and language teaching: Bridging EGAP and ESAP’. In G. Brooks (ed.), *The 2016 PanSIG Journal Innovations in Education*, 1-9.
- Asif, Muhammad, Shafqat Zaidi and Fariha Yasmeeen.** (2021). ‘A corpus-based study of Austen’s novel *Pride and Prejudice* using Cirrus Tool’. *Corporum: Journal of Corpus Linguistics*, 4(1):60-72.
- Babanoğlu, Pinar M. and Kerem Sütçü.** (2017). ‘A corpus-based study on the use of contractions by EFL learners in argumentative essays’. *International Journal of Applied Linguistics and English Literature*, 6(2):56-62.
- Bacha, Nahla Nola.** (2001a). ‘Writing evaluation: What can analytic versus holistic essay scoring tell us?’ *System*, 29(3):371-383.

- Bacha, Nahla Nola.** (2005b). 'Academic vocabulary: A corpus linguistic approach'. *International Journal of Arabic-English Studies*, 6(1):123-146.
- Bacha, Nahla Nola and Victor Khachan.** (2012c). 'A lexical corpus based analysis of L2 academic vocabulary: A case study'. *The Asian ESP Journal*, 8(1):53-74.
- Bacha, Nahla Nola.** (2016d). 'Teaching the novel in a university English as a Foreign Language (EFL) context: An exploratory study in Lebanon'. *International Journal of Learning, Teaching and Educational Research*, 15(12):155-173.
- Bacha, Nahla Nola.** (2020e). A 'marriage' between language and literature: Writing an argumentative literary essay in EFL contexts'. *International Journal of Language and Literature*, 8(2):31-39.
- Bahous, Rima.** (2008). 'The self-assessed portfolio: A case study'. *Assessment and Evaluation in Higher Education*, 33(4):381-393.  
DOI: 10.1080/02602930701562866.
- Barabadi Elyas and Yasser Khajavi.** (2017). 'The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary'. *Cogent Education*. 4(1):2-13.  
DOI: 10.1080/2331186X.2017.1283876.
- Behzadi, Lale and Vahid Behmardi.** (2009). *The Weaving of Words: Approaches to Classical Arabic Prose*. Beirut Texte und Studien (Serial No. 112), Orient-Institut Beirut. ISBN 978-3-89913-678-4.
- Benavides, Carlos.** (2015). 'Using a corpus in a 300 level Spanish grammar course'. *Foreign Language Annals*. 48(2):218-235.
- Biber, Douglas.** (2011). 'Corpus linguistics and the study of literature: Back to the future'. *Scientific Study of Literature*. 1(1):15-23.
- Deborah Schiffrin, Deborah Tannen, and Heidi Hamilton (eds.),** *The Handbook of Discourse Analysis*, 175-196. Massachusetts: Blackwell.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan** (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Bolton, Kingsley, Gerald Nelson and Joseph Hung.** (2002). 'A corpus- based study of connectors in student writing'. *International Journal of Corpus Linguistics*, 7(2):165-182.
- Boulton, Alex** (2016). 'Integrating corpus tools and techniques in ESP courses'. *ASp La Revue DuGERAS* 69. |DOI-10.4000/asp.4826.
- Boulton, Alex and Thomas J** (2012). 'Corpus language input, corpus processes in learning, learner corpus Product'. In James Thomas and Alex Boulton (Eds.), *Introduction to input, Process and Product: Developments in Teaching and Language Corpora*, 7-34, Masaryk University Press.
- Brynildssen, Shawna.** (2000). 'Vocabulary's influence on successful writing'. *ERIC Digest D157*. ED446339 2000-11-00.

- Butler, Ian.** (2006). 'A brighter future? Integrating language and literature for first-year university students'. In A. Paran (ed.), *Literature in Language Teaching and Learning*, 11-25. TESOL Inc.
- Bychkovska, Tetyana and Joseph J. Lee.** (2017). 'At the same time: Lexical bundles in L1 and L2 university student argumentative writing'. *Journal of English for Academic Purposes*, 30:38-52.
- Castello, Erik.** (2008). *Text Complexity and Reading Comprehension Tests*. Peter Lang.
- Cobb, Tom.** (2021). Web Compleat lexical tutor/vocabulary profile. [www.lextutor.ca](http://www.lextutor.ca). (Retrieved on April 10, 2023).
- Coxhead, Averil.** (2000). 'A new academic word list'. *TESOL Quarterly*, 34(2):213-238.
- Daoshan, MA. and Zhang Shuo.** (2014). 'A discourse study of the ice-burg principle in *A Farewell to Arms*'. *Studies in Literature and Language*, 8(1):80-84.
- de Beaugrand, Robert.** (2001). 'Interpreting the discourse of H. G. Widdowson: A corpus-based critical discourse analysis'. *Applied Linguistics*, 22(1): 104-121. DOI:10.1093/applin/22.1.104.
- Defazio, Joseph, Josette Jones, Felisa Tennant and Sara Anne Hook.** (2010). 'Academic literacy: The importance and impact of writing across the curriculum – a case study'. *Journal of the Scholarship of Teaching and Learning*, 10(2): 34-47.
- Diab, Nuwar and Ghada M. Awada.** (2022). 'Effect of language learning strategies and teacher versus peer feedback on reducing lexical errors of University Learners'. *International Journal of Arabic-English Studies*, 22(1):101-124. <https://doi.org/10.33806/ijaes2000.22.1.6>
- Diab, Rula.** (2006). 'Error correction and feedback in the EFL writing classroom comparing instructor and student preferences'. *English Teaching Forum*, 44(3):2-13. <https://eric.ed.gov/?id=EJ1107984>.
- Egbert, Jesse, Brent Burch and Douglas Biber.** (2020). 'Lexical dispersion and corpus design'. *International Journal of Corpus Linguistics*, 25:89-115. DOI: 10.1075/Ijcl.18010.Egb.
- Engber, Cheryl A.** (1995). 'The relationship of lexical proficiency to the quality of ESL compositions'. *Second Language Writing*, 4(2):139-155.
- Flowerdew, Lynne.** (2009). 'Applying corpus linguistics to pedagogy: A critical evaluation'. *International Journal of Corpus Linguistics*, 14(3):393-417. DOI: 10.1075/ijcl.14.3.05flo.
- Gemeinbock, Iris.** (2016). 'Representativeness in corpora of literary texts: Introducing the C18P project'. University of Vienna.
- Gilroy, Marie and Brian Parkinson.** (1997). 'Teaching literature in a foreign language'. *Language Teaching*, 29:213-25.
- Graesser, Arthur C., N. Dowell, and C Moldovan.** (2011). 'A computer's understanding of literature'. *Scientific Study of Literature*, 1(1):24-33. The University of Memphis.

- Grant, Lesie, and April.** (2000). 'Using computer-tagged linguistic features to describe L2 writing differences'. *Journal of Second Language Writing*. 9:123-145. [https://doi.org/10.1016/S1060-3743\(00\)00019-9](https://doi.org/10.1016/S1060-3743(00)00019-9).
- Granger, Sylviane.** (2002). 'A bird's-eye view of computer learner corpus research'. In Susan Granger, J. Hung, S. Petch-Tyson and J. Hulstijn (eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*, 3-33. Amsterdam: John Benjamins.
- Granger, Sylviane.** (2008). 'Learner corpora'. In A. Ludeling, and M. Kyto (eds.). *Corpus linguistics. An International Handbook*. 1-14. Berlin: Walter de Gruyter Publisher.
- Granger, Sylviane, Estelle Dagneaux, Fanny Meunier and Magali Paquot.** (2009). *International Corpus of Learner English*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Hamadeh, Wissam, Rima Bahous, Rula Diab, and Mona Nabhani.** (2020). 'Using social media to enhance second language learning'. *Computer-Assisted Language Learning Electronic Journal*, 21(2):132-149 .
- Hemingway, Ernest.** (2004). *A Farewell to Arms*. Penguin.
- Hoey, Michael, Michaela Mahlberg, Michael Stubbs and Teubert Wolfgang.** (eds.), (2007). *Text, Discourse and Theory and Analysis*. Continuum.
- Hoey, Michael.** (2013). *Lexical Priming*. Taylor and Francis. Retrieved from <https://www.perlego.com/book/1617387/lexical-priming-a-new-theory-of-words-and-language-pdf>. (Original work published 2012).
- Höglund, Mikko and Kaj Syrjänen.** (2016). 'Corpus of early American literature'. *ICAME Journal*, 40(1):16-38.
- Hunston, Susan and Gill Francis.** (2000). *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamin.
- Hunston, Susan.** (2002). *Corpora in Applied Linguistics*. Cambridge: University Press.
- Hunston, Susan.** (2009). 'The usefulness of corpus-based descriptions of English for learners: The case of relative frequency'. In Karin Aijmer (ed.), *Corpora and Language Teaching*, 141-156. Amsterdam: John Benjamins.
- Hussein, Khalid Shakir and Ina'am Abdul-Jebbar Abdul-kadhim.** (2020). 'A corpus-based stylistic identification of lexical density profile of three novels by Ernest Hemingway: *The Sun also Rises, A Farewell to Arms, and A Moveable Feast*'. (Retrieved on 15, December 2022). <https://www.bing.com/search?q=Khalid+Shakir+Hussein>
- Ihejirika, Richard C.** (2014). 'Literature and English language teaching and learning: A symbiotic relationship'. *English Language Teaching*, 7(3):84-90.
- Itani, Layla Youssef and Rima Bahous.** (2019). 'Media discourse analysis of news reports in Lebanese newspapers', *Global Media Journal*, 17(32):1-10.
- Johns, Ann and Tony Dudley-Evans.** (1997). 'English for specific purposes: International in scope, specific in purpose'. *TESOL Quarterly*, 25(2): 297-314.

- Johns, Ann M. Anis Bawarshi, Richard M. Coe, Ken Hyland, Brian Paltridge and Mary Jo Reiff Christine Tardy.** (2006). 'Crossing the boundaries of genre studies: Commentaries by experts'. *Journal of Second Language Writing*, 15:234-249.
- Johnson, Mark D., Anthony Acevedo and Leonardo Mercado.** (2016). 'Vocabulary knowledge and vocabulary use in second language writing'. *TESOL Journal*, 7:700-715. <https://doi.org/10.1002/tesj.238>.
- Khan, Mohammad Shamsur rabb Khan and Ali Mohammad.** (2018). 'Literary texts in the EFL classroom: Applications, benefits and approaches'. *International Journal of Applied Linguistics and English Literature*, 7(5): 167-179.
- Kroll, Barbara** (ed.). (1991). *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press.
- Laufer, Batia and Paul Nation.** (1995). 'Vocabulary size and use: Lexical richness in L2 written productions'. *Applied Linguistics*, 16(3):307-322.
- Laosrirattanachai, Paul and Sugunya Ruangjaroon.** (2021). 'Implementation of a data-driven hospitality lexis learning programme. *3L Language, Linguistics, Literature*, 27(1): 1-21.
- Learning Corpus Bibliography.** [Learner Corpus Association](https://www.learnercorpusassociation.org) (Retrieved from <https://www.learnercorpusassociation.org> on March 24, 2023).
- Lewis, Michael.** (Ed.) (2000). *Teaching Collocation: Further Developments in the Lexical Approach*. Language Teaching Publications.
- Lin, Ching-His and Yen-Liang Lin.** (2019). 'Grammatical and lexical patterning of *make* in Asian learner writing: A corpus-based study of *ICNALE*', *3L.The Southeast Asian Journal of English Language Studies*. [Yen-liang Lin | Semantic Scholar](#).
- López-Couso, María José, Belén Méndez-Naya, Paloma Núñez-Pertejo and Ignacio M. Palacios-Martínez** (2016). *Corpus Linguistics on the Move: Exploring and Understanding English through Corpora*. Leiden: Brill.
- Mahlberg, Michaela.** (2007). 'A corpus stylistic perspective on Dickens' *Great Expectation's*. In M. Lambrou and Stockwell (eds.), *Contemporary Stylistics*, 19-31. London: Continuum.
- Mahlberg, Michaela.** (2010). 'Corpus linguistics and the study of nineteenth-century fiction'. *Journal of Victorian Culture*, 15(2):292-298. DOI:10.1080/13555502.2010.491667.
- Mahlberg, Michaela.** (2012). 'The corpus stylistic analysis of fiction – or the fiction of corpus stylistics'? In J. Mukherjee, and M. Huber (eds.), *Corpus Linguistics and Variation in English*, 77-95. Leiden: Brill.
- Mahlberg, Michaela, Voila Wiegand, Peter Stockwell, and Anthony Hennessey.** (2019). 'Speech-bundles in the 19th-century English novel'. *Language and Literature*, 28(4):326–353.
- Mart, Cagri Tugru** (2021). 'Literature-based instruction: A worthwhile approach for the mastery of a second language'. *3L Language, Linguistics, Literature*, 27(2):49-61.
- Masood, Muhammad Hamzah, Shahzeb Shafi and Maqsood Darwesh.** (2020).



‘Study of Shakespearean themes and characters through corpus based voyant tools approach’. *Conference: National Conference on Education (NCE)*.

- McEnery, Tony and Richard Xiao.** (2011). ‘What corpora can offer in language teaching and learning’. In E. Hinkel (ed.), *Handbook of Research in Second Language Teaching and Research*, 361-381. London: Routledge.
- McIntyre, Dan.** (2013). ‘Corpora and literature’. *Encyclopedia of Applied Linguistics*, University of Huddersfeld. 1-14.
- McNaught, Carmel and Paul Lam.** (2014). Using Wordle as a supplementary research tool. <https://nsuworks.nova.edu/tqr/vol15/iss3/8>
- Moon, Rosamund.** (2008). ‘Sinclair, phraseology, and lexicography’. *International Journal of Lexicography*. 21(3):243-254.  
<https://doi.org/10.1093/ijl/ecn027>
- Mukattash, Lewis.** (2003). ‘Towards a new methodology for teaching English to Arab learners’. *International Journal Arabic-English Studies*, 4:211-234.
- Nation, Paul.** (1990). *Teaching and Learning Vocabulary*. London: Newbury House.
- Nation, Paul.** (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, Paul.** (2021). ‘Is it worth teaching vocabulary?’ *TESOL Journal*, 12(4).  
<https://doi.org/10.1002/tesj.564>
- Nation, Paul and Paul Meara.** (2020). ‘Vocabulary’. In N. Schmitt and M. Rodgers (3<sup>rd</sup> ed.) (eds.), *An Introduction to Applied Linguistics*, 35-54. London: Routledge.
- Nelson, Gerald.** (2000). ‘An introduction to corpus linguistics’. *Journal of English Linguistic*, 28(2):193-196.
- Neumann, Stella.** (2014). *Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German*. Berlin: De Gruyter, Mouton.
- Nur, Aliyah.** (2015). ‘Analyzing vocabulary level of the students’ essay writing and academic textbook utilizing lextutor’. *Eternal Journal for Corpus*, 1(1):15-27.
- O’Keeffe, Anne, Michael McCarthy, and Ronald Carter.** (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Paran, Amos.** (2006). ‘The stories of literature and language teaching’. In A. Paran (Ed.), *Literature in Language Teaching and Learning*, 1-10. TESOL Inc.
- Picoral, Adriana.** (2018). *Corpus Linguistics on the Move: Understanding English through Corpora*. Leiden: Brill.
- Pretila, Päivi, Katalin Doró and Renata Pípalová.** (eds.), (2015). *Lexical Issues in L2 Writing*. Cambridge: Cambridge scholars Publishing.
- Reppen, Randi.** (2010). *Using Corpora in the Classroom*. Cambridge: Cambridge University Press.
- Research Uses of VocabProfile (lextutor.ca).**

(Retrieved from [https://www.lextutor.ca/vp/eng/vp\\_research/html](https://www.lextutor.ca/vp/eng/vp_research/html) on March 23, 2023

- Ridwan, Wahid.** (2011). 'The use of corpus-based techniques in literary analysis: Exploring learners' perceptions', *Asiatic*. 5(1):124-128.
- Römer, Ute.** (2006). 'Where the computer meets language, literature, and pedagogy: Corpus analysis in English studies.' In A. Gerbig and A. Müller-Wood (eds.), *How Globalization affects the Teaching of English: Studying Culture through Texts*, 81-109. E. Mellen Press.
- Römer, Ute.** (2009). 'Corpus research and practice: What help do teachers need and what can we offer?' In Karin Aijmer (ed.), *Corpora and Language Teaching*, 83–98. John Benjamins.
- Sadaka, George.** (2022). 'Illnesses of illusion and disillusionment: From euphoria to aporia', *Life Writing*, DOI:10.1080/14484528.2022.2127627.
- Sadaka, George and Vicky Panossian.** (2022). 'Policing Victorian women's desire: Retracing mirrored patriarchy in Jane Eyre and Villette', *Brontë Studies*, 47(2):128-140, DOI: 10.1080/14748932.2022.2043675
- Sampson, Geoffrey and Diana McCarthy** (eds.), (2005). *Corpus Linguistics in a Widening Discipline*. London: Continuum.
- Scardamalia, Marlene and Carl Bereiter.** (1987). 'Knowledge telling and knowledge transforming in written composition'. In S. Rosenberg (ed.), *Advances in Applied Psycholinguistics Reading, Writing and Language Learning*, 2:142-175. Cambridge: Cambridge University Press.
- Schmitt, Norbert, and Michael Rodgers.** (eds.), (2020). *An Introduction to Applied Linguistics*. London: Routledge.
- Scott, R.** (2004). 'Institutional theory: Contributing to a theoretical research Program'. In K. G. Smith, and M. A. Hitt, (eds.), *Great Minds in Management: The Process of Theory Development*, 460-484, Oxford: Oxford University Press.
- Shaaban, Kassem.** (2017). 'Challenges of teaching English in tertiary education in the Arab world. The ongoing rivalry between English and French in Lebanon: Current practices and future directions'. In A. Gebril (ed.), *Applied linguistics in the Middle East and North Africa. Current Practices and Future Directions*, 162-182. Amsterdam: John Benjamins.
- Shakeel, Rana K. and Muhammad K. Khan.** (2021). 'Compiling literary academic word list (LAWL): A corpus analysis of English papers'. *Corporum: Journal of Corpus Linguistics*, 3(2):92-115.
- Shaw, Philip.** (2009). 'Linking adverbials in student and professional writing in literary studies: What makes writing mature'. In M. Charles, D. Picorari, and Susan Hunston (eds.), *Exploring the Interface between Corpus Linguistics and Discourse*, 215-235. London: Continuum.
- Sinclair, John** (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John.** (2004). *Trust the Text*. London: Routledge.

- Sinclair, John.** (2007). 'Introduction'. In Michael Hoey, Michaela Mahlberg, Michael Stubbs and Wolfgang Teubert (eds.), *Text, Discourse and Corpora: Theory and Analysis*, 1–7. London: Continuum.
- Sinclair, Stefan and Geoffrey Rockwell.** (2021). Voyant tools version 2.4 Web-based text analysis suite. *Language Learning and Technology*, 25(3):43–50. <https://voyant-tools.org>.
- Smith, J., and Colleen Kelly.** (2002). 'Stylistic constancy and change across literary corpora: Using measures of lexical richness to date works'. *Computers and the Humanities*, 36(4):411-430. [www.jstor.org/stable/30204686](http://www.jstor.org/stable/30204686).
- Stubbs, Michael.** (2001). 'Texts, corpora and problems of interpretation: A response to Widdowson'. *Applied Linguistics*, 22(2):149-72.
- Stubbs, Michael.** (2005). 'Conrad in the computer: Examples of quantitative stylistic techniques'. *Language and Literature*. 14(1):5-24.
- Toolan, Michael.** (2004). 'Values are descriptions: Or, from literature to linguistics and back again by way of keywords. *Belgian Journal of English Language and Literatures* (BELL New Series 2), 11-30.
- Ullah, Zafar, Muhammad Uzair and ArshadMahmood.** (2019). 'Extraction of key motifs as a preview from 2017 Nobel prize winning novel, 'Never Let Me Go': An interactive word cloud study'. *Journal of Research in Social Sciences*, 7(2):83-98.
- Ure, J.** (1971). 'Lexical density and register differentiation'. *Contemporary Educational Psychology*, 5:96-104.
- Viera, Rodrigo Tovar- .** (2022). 'Lexical richness of abstracts in scientific papers in Anglophone and non-Anglophone Journals'. *3L: Language, Linguistics, Literature*, 28(2). <http://doi.org/10.17576/3L>
- Voegelin, Cristina K. and Stefan Keller.** (2021). 'Linguistic features of formative feedback on ESL argumentative writing: Comparing pre-service and experienced teachers'. *Issues in Applied Linguistics* 22(0). 1-22. DOI10.5070/L422049267.
- Wahid, Ridwan.** (2011). 'The use of corpus-based techniques in literary analysis: Exploring learners' perceptions'. *Asiatic*, 5(1):104-127.
- Widdowson, Henry G.** (1991). 'The description and prescription of language'. In J. Alatis (ed.), *Georgetown University Round Table on Languages and Linguistics*. 11–24. Washington, DC: Georgetown University Press.
- Widdowson, Henry G.** (2008). 'The novel features of text: Corpus analysis and stylistics'. In A. Gerbig and O. Mason (eds.), *Language, People, Numbers*, 93-304. Amsterdam: Rodopi.
- Widdowson, Henry G.** (2000a). 'Critical practices: On representations and the interpretation of text'. In S. Sarangi and M. Coulthard (eds.), *Discourse and Social Life*, 155-169. London: Longman.
- Widdowson, Henry G.** (2000b). 'The limitations of linguistics applied'. *Applied Linguistics*, 21(1):3-25.

- Wulff, Stefanie.** (2017). 'What learner corpus research can contribute to multilingual research'. *International Journal of Bilingualism*, 21(6):734-753.
- Yoon, Choongi.** (2011). 'Concordancing in L2 writing class: An overview of research and issues'. *Journal of English for Academic Purposes*, 10:130-139.
- Yu, Guoxing.** 2010). 'Lexical diversity in writing and speaking task performances'. *Applied Linguistics*, 31(2):236-259.
- Zhang, Xiaofei.** (2022). 'The relationship between lexical use and L2 writing quality: A case of two genres'. *International Journal of Applied Linguistics and English Literature*.  
<https://onlinelibrary.wiley.com/doi/10.1111/ijal.12420?af=R>.